Spatio-Temporal Fuzzy Clustering of Functional Magnetic Resonance Imaging Data

by Mark Douglas Alexiuk

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy in Electrical Engineering

October 2006

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

Spatio-Temporal Fuzzy Clustering of Functional Magnetic Resonance Imaging Data

BY

Mark Douglas Alexiuk

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

Of

DOCTOR OF PHILOSOPHY

Abstract

Magnetic resonance imaging (MRI) is a preferred imaging modality due to its high resolution images of *in vivo* tissue. Functional MRI (fMRI) infers organ function using blood flow intensities. However, multiple response models for hemodynamics and, more specifically, neural activation, contend for widespread adoption. Development of models, imaging techniques and various types of noise compound problems in analysis and motivate the use of exploratory data analysis to elicit intrinsic data structure. This work demonstrates the utility and efficacy of a novel exploratory data analysis technique derived from a robust, unsupervised learning method, fuzzy C-means (FCM). The algorithm, designated FCM with feature partitions (FCMP), integrates feature relationships in the clustering process. One feature relation not widely exploited in fMRI analysis is the high probability that temporally similar time courses are also spatially proximal. FCMP has exploited this relation to generate both novel and robust data inferences. Both synthetic and *in vivo* fMRI data are examined. FCMP is compared to benchmarks from industry and academia, including FCM, cluster merging, CHAMELEON and EvIdent®. Ten distinct experiments examine aspects of FCMP with respect to fMRI analysis, in particular, means to integrate distinct feature subsets and feature relationships, sample membership in regions of interest, use of validation indices for fMRI, and data-driven global thresholding. Efficacy of FCMP for fMRI analysis is shown in terms of noise reduction, statistical specificity, and discovery of novel spatial relations between time courses in regions of interest.

To Tressa, Eve, Erin and Lily.

Acknowledgments

This thesis could not have been completed without the encouragement and patient support that many people shared.

First, I would like to thank my family for lending joy to the mundane.

A debt of gratitude is owed to my thesis advisor, Dr. N. Pizzi. Research under Dr. Pizzi has provided a panoramic vista of pattern recognition problems, hands on software development, and a philosophical foundation for questioning methodological limits.
Much of the research was conducted at the Institute for Biodiagnostics in the Pattern Recognition Systems Group, which is part of Biomedical Informatics. Interaction with this creative group of technical experts has significantly improved source code quality and sped the completion of this thesis. Special thanks to A. Demko and R. Vivanco. Other researchers in Biomedical Informatics offered countless helpful comments and include R. Summers, Dr. M. Jarmasz, Dr. M. Alexander, Dr. R. Baumgartner, and Dr. R. Somorjai.

Faculty at the University of Manitoba deserve credit for inspiration while staff have been very helpful. This is true for both the engineering and mathematics departments. Professors who made particularly deep impressions were E. Schwedyk and M. Pawlak, and R.W. Quackenbush.

Parallel to my studies, I have benefited from the mentoring of two professional engineers, R. Bernhardt and B. Malenko. A valued collaborator not mentioned above is G. Wiebe. K. Stupak has been a constant encourager. Finally, the camaraderie between graduate students in the midst of their studies is a special kind of solace. Thank you all.

The thesis committee members deserve thanks for their valued comments and criticism: Profs. J.F. Peters III, K. Ferens, B. Li, W. Pedrycz, and A. Bargiela.

Finally, financial assistance provided through NSERC grants and the National Research Council have underwritten this work and were gratefully received.

There is no panacea in data analysis.
P. Huber

One can show the following: given any rule, however fundamental or necessary for science, there are always circumstances when it is advisable not only to ignore the rule, but to adopt its opposite.
P. K. Feyerabend

Spatio-Temporal Fuzzy Clustering of Functional Magnetic Resonance Imaging Data

by Mark Douglas Alexiuk

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba

October 2006

Table of Contents

Figure Index

Index of Tables

1. Overview

[Knowledge is] a rich storehouse for the glory of the Creator and the relief of man's estate.

Francis Bacon

Introduction

Magnetic resonance imaging (MRI) [Kupe] is a non-invasive medical imaging technology that provides data concerning *in vivo* organ state and function. Commercial development of MRI has found many applications supplementing and replacing traditional radiological modalities. MRI studies are, however, notorious for the voluminous quantities of data produced. The challenge for MRI analysts is the detection, extraction, and transformation of information from a large set of complex data. MRI data consists of radio-frequency (rf) intensities at volume elements (*voxels)* having unique spatial coordinates, thus providing organ state information. Functional MRI (fMRI) data consists of a sequence of MRI acquisitions over alternating states of subject activity and rest. Blood flow changes between these states, or the hemodynamic response, is the basis of inferring organ function. The time series of intensity values recorded at a single voxel is called a time course (TC). When the sequence of rest and active states is known, it is called the activation or stimulus paradigm. Correlation [Edwa] between the paradigm and TCs defines regions of interest (ROI) in the active organ. Since fMRI datasets contain unanticipated, yet desirable, information, exploratory data analysis (EDA) [Tuke]

techniques are used. These unsupervised pattern recognition techniques generate hypotheses about the data that are ultimately validated by comparison to a *gold standard,* often an expert in the field. The process of exploring the intrinsic structure of a dataset often leads to novel investigation.

Data Acquisition

In vivo data acquisition inevitably contains undesirable characteristics attributable to the subject, whether human or animal, or to the diagnostic equipment [Brow]. For example, subject movement introduces ghosting artifacts while instrument instability can cause voxel drift. Many methods are used to mitigate the effects of noise on analysis. For example, image registration techniques re-align shifted and perturbed fMRI images. A large variety of signal and image processing techniques [Gonz] are applied to recover and to enhance fMRI datasets. However, amelioration can be limited by time, computational costs, the small number of patient/disease cases, and uncertainty in the hemodynamic model. Other efforts to reduce noise in the data acquisition process have limited application, such as contrast enhancement through suppression, weighting and absorption methods. Attempts to acquire noise-free images include use of specialized rf-coils and differentiated rf-echo sequences. Techniques to address noise continue to be develop and

can offer tissue or organ-specific solutions.

Neural Activation Studies

Neural activation studies represent one of the most challenging types of fMRI analysis. As in other fMRI studies, a stimulus is applied to the subject. Neuron signal intensity levels are then recorded for one or more coronal slices of the subject's brain. These studies are challenging due to the limited body of knowledge of neural function and interdependencies, the incompleteness of generalized hemodynamic response models [Duann], and the presence of noise in the dataset. Synthesized fMRI time series and *in vivo* neural activation data are examined. Synthetic data were generated to examine specific hypothetical fMRI analysis scenarios. *In vivo* data, with visual and tactile stimuli, were acquired at the Institute for Biodiagnostics (IBD) (www.ibd.nrc-cnrc.gc.ca), a research institute of the National Research Council (NRC) of Canada (www.nrc-cnrc.gc.ca).

Motivation

This research is motivated by the robust fMRI analysis of neural activation studies deemed possible by an EDA algorithm that, while enunciating intrinsic data structure that includes so-called *unanticipated* TCs, minimizes the impact of noise and incorporates

expert knowledge, especially in terms of feature relations, into the algorithm. Modifications to fuzzy C-means (FCM) clustering are proposed due to the broadly based, resilient performance of its objective function based optimization. A novel variant, FCM with feature partitions (FCMP), is developed and applied to synthetic and *in vivo* fMRI datasets. An additional motivation is the clarity that a fuzzy clustering algorithm, one that considered relationships between features, would bring to the field as it would encompass some mechanical aspects observable in many extent FCM adaptations.

Original Contributions and Benefits

Several original contributions are collected in this work. First, a mathematical formulation of a generalized fuzzy clustering algorithm is derived which suggests a novel, encompassing cluster algorithm taxonomy. One benefit of this formulation is the succinct manner in which cluster analysis can now be expressed. Also, the generalized algorithm is readily adapted to exploit unique situations not related to fMRI analysis, and defines a structured manner in which adaptation occurs. Secondly, the application of the generalized algorithm to spatio-temporal fuzzy cluster analysis is examined with respect to various synthetic and *in vivo* fMRI datasets. Discovery of regions of interest is examined and compared to benchmarks. Algorithm robustness is tested over a range of

noise levels. A method to define data-driven global thresholds for fMRI datasets using bridge voxels is presented. Finally, visual cluster validity indices and spatial cluster-assignment maps are demonstrated as visualization methods for use with FCMP and fMRI datasets.

1.1 Preamble

In order to convey an overall thesis context, short comments on EDA, FCM, and fMRI are presented with respect to the novelty and utility of FCMP.

Exploratory Data Analysis

Scientific discovery involves a process of concept formation, measurement, and analysis. Hypotheses are framed; instruments are devised and built; experiments are recorded and analyzed. This process repeats itself with variations directed by the judgment of the scientist. Many components of this process are challenging, for instance, determining significant features for measurement beforehand, or deriving simple equations after the fact. Therefore, algorithms have been developed that operate with agnostic attitudes towards data organization and contain a minimal mathematical model. EDA algorithms exhibit an exploratory character open to opportune solutions and elicit inherent data structure. Parameters are typically *data-driven,* being based on dataset properties and not on *a priori* knowledge.

FCM as an EDA Technique

FCM [Bezd1] has proved to be both a highly successful and widely applicable EDA technique. As it partitions the samples, FCM uses a data model that assumes only that C spherical clusters exist in the dataset. However, FCM is also well suited to exploit specialized data structures that limit accurate representation by spherical clusters. Variations include hyper-ellipsoid clusters [Gust], modifying distance calculations based on partial knowledge of sample labels [Pedr5], and the use of robust metrics [Bobr]. Since FCM is often modified to characterize such idiosyncrasies, the unmodified algorithm, using C spherical clusters, is often referred to as *vanilla* FCM. When the value of C is in question, validation indices [Höpp] [Wei] quantify the fitness of FCM parameters to dataset structure. However, it is an open question as to which adaptation of FCM is preferable for any given dataset. The problem becomes more tractable when datasets are restricted to a single problem domain. This thesis modifies FCM to develop a fuzzy clustering algorithm specific to fMRI datasets.

In the development of a specific FCM modification, it became apparent that a general cluster analysis formulation would be beneficial in order to organize the many FCM variants into a cluster analysis taxonomy. Fuzzy clustering literature largely consists in

adapting FCM to dataset structural idiosyncrasies by modifying the objective function, changing metrics, or *ad hoc* heuristics. A succinct characterization of such variations would make comparisons between variations more coherent and focus future variations. One adaptation absent from the literature concerns the explicit integration of distinct feature subsets where prominence is given to inter-feature relationships. Implicit integration, where specialized metrics replace the Euclidean metric, is awkward as a general mechanism. Grouping features into subsets that distinctly contribute to the convergence of the algorithm has numerous justifications: features may be measured through different modalities, contain different types and levels of noise, exhibit specialized patterns, or prompt clarifying heuristics. This thesis describes a novel modification to FCM for use with fMRI datasets, develops a general cluster formulation useful for taxonomy, and provides a mechanism to integrate distinct feature subsets.

FCMP

FCMP defines feature relationships using a formal structure, called a feature partition, consisting of a triple:

a) a set of feature indices denoting the membership of a feature in the partition.
b) a metric used to calculate distance between the features in the partition.
c) a weight calibrating the relevance of the feature partition to the objective function.

FCMP generalizes FCM; validation methods and visualizations for FCM can be retained or extended for FCMP. The advantage of FCMP is that feature partitions can fuse contributions (to sample-centroid distances) from sample features and calibrate the objective function. Each such fusion incrementally constrains algorithm convergence between that of FCM on individual feature subsets. In FCMP, feature-specific distances are combined, or fused, at each algorithm iteration. Judicious selection of feature subsets, metrics, and weights, allow FCMP to be tuned to elicit dataset structure within a given set of feature relationships. Discrimination of the contributions of individual features to the clustering process is a valuable tool when noise contaminates only some features. Finally, feature partitions provide a means to normalize features with disproportionate magnitudes, variance, or to maintain feature-metric associations.

FCMP and fMRI

Sample features in a fMRI study decompose into two categories:

1. Spatial features that denote the position of the neural activity or the voxel location. These features are represented as a triple, {x,y,z}; comparisons between locations use the Euclidean distance.
2. Temporal features that record signal intensities at a specific location for each of n time instances. Intensity feature is represented by the n-tuple $\{t_1, t_2, ..., t_n\}$; comparisons between time series commonly use Pearson correlation.

Most current fMRI analysis considers only the temporal intensities. However, an EDA

perspective suggests that all features be included. Using only temporal intensities excludes known feature relations in fMRI. When applying FCMP to fMRI, we adopt the feature cleavage into spatial and temporal domains and form a feature partition for each domain. Each partition retains the metric commonly associated with its features. The remaining variables, partition weights, are manipulated to exhibit known feature relations, for example, temporal similarity of samples suggests spatial proximity of the samples. This approach achieves the desirable result that increasing the spatial partition weight causes fMRI ROIs to exhibit enhanced spatial connectivity, a measure of robustness.

1.2 Scope

This thesis examines FCMP [Alex3], a generalization of FCM and its applications to problems in fMRI analysis. FCM is a widely used data analysis technique that elicits inherent structural information from a dataset. Problems in fMRI analysis are varied and include: grouping of similar time courses, discriminating between novel and noisy time courses, and defining spatial regions of interest. It is shown that FCMP exhibits critical features of an EDA technique, since it:

- Maintains implicit inter-sample relations through membership matrices;
- Detects novel structure;

- Represents groups of objects without masking the identity of components.

Since FCMP is presented as a generalized cluster algorithm, several adaptations (or specializations) of FCMP are applied to real-world datasets. An adaptation of the generalized algorithm to a specific problem is termed a *specialization*. One specialization examined in depth incorporates spatio-temporal information into the FCMP objective function. In particular, the application and extension of FCM to the following problems is examined:

- Incorporating spatial context into fuzzy clustering of fMRI time series;
- Discovering novel time courses in an fMRI neural activation study;
- Making robust statistical statements about fMRI ROIs.

Feature partition benefits are anticipated to be specific to datasets and objective functions.

Thesis Structure

The remainder of this thesis conforms to the following structure. Chapter 2 discusses concepts in EDA, fundamentals of pattern recognition and classification. Chapter 3 provides details on cluster analysis, including formula comparisons between several types of clustering algorithms. Chapter 4 introduces FCMP as a novel generic clustering formula as well as its specialization to fMRI problems. Chapter 5 outlines the basic theory behind MRI and neural activation studies. Chapter 6 introduces benchmark algorithms for the efficacy of FCMP. Chapter 7 lists the experiments executed and

discusses the results. Chapter 8 draws conclusions from this investigation and points to areas of future research. Appendices provide additional information on fundamental concepts, formula derivations, glossaries, acronyms, and symbols used in the thesis.

2. Exploratory Data Analysis

Far better an approximate answer to the right question, which is often vague,
than an exact answer to the wrong question, which can always be made precise.
John W. Tukey

Exploratory data analysis (EDA) [Tuke] differs from model based analysis in that it skeptically imposes only the sparsest mathematical models on the data. Since skewed data models misrepresent data structure, the sparsest models hide the intrinsic data properties the least. Models used by EDA tend to be *data-driven* where model parameters are determined solely by dataset statistics. EDA principles place high importance on visualization of the data, examination of sample subsets, and sample/feature inclusion. Robust statistical methods are commonly used in order that even outliers or noisy samples can contribute to some degree [Burn] [Cove] [Hube] [Krza] [Mart] [Rous]. Cluster analysis is a popular example of EDA [Baum3] [Demk] [Fuji] [Mose].

The main benefit of EDA techniques is the characterization of intrinsic structure which includes the possibility of useful, yet unanticipated, results. Data-driven algorithms are ostensibly objective means to discover data structure since external models are deprecated. In fact, the large field of unsupervised learning has as its purpose the elicitation of intrinsic data organization and includes: self-organizing maps [Koho], fuzzy

clustering [Pizzi1], visualization methods [Kirb] [Van], Fuzzy Group Method of Data Handling(GMDH) [Farl], and Higher Order Statistical Analysis (HOSA) [Nikias]. EDA is a critical tool in problem domains such as: biomedicine, cryptography, drug interaction studies, and financial time series. This chapter discusses fundamentals of EDA, classification, preprocessing, and metrics.

Justification for EDA

Pattern recognition systems must be able to operate when samples are contaminated by noise or under other non-ideal conditions. In order to capture patterns that are generalizable, EDA methods are often preferred over model-intensive approaches, especially under the following conditions:

Model Paucity or Sparsity: No single mathematical model is accepted by the research community for the data. The phenomenon underlying the data is poorly understood.

Arbitrary Choice of Models: Lack of *a priori* information about the data makes the choice of any one mathematical model unjustified.

Model Limitations: Linear or low order models are insufficient for the final application. It is difficult to combine local, simple classifiers or to achieve an acceptable generalization error.

Noisy Samples: Samples are contaminated by noise.

Tarnished Class Labels: Classifiers map samples to class labels, which are the gold standard for classification problems. However, labels are *tarnished* when they less than

100% correct. Burnishing is the process of re-labeling the samples to improve the label accuracy. The detection of tarnished labels and the selection of burnishing methods are generally open problems.

Absence of Class Labels: Class labels do not exist at all for the dataset.

The possibility of these conditions should be evaluated before a model-intensive algorithm is harnessed to the data. Re-evaluation should occur as additional samples are acquired.

Model Benefits and Risks

Data models provide the opportunity to fit the data to the model, thereby determining values which characterize the data. However, data can also be fit to the model through the rejection of *outliers* or *questionable* samples. When the data is fit to a model, residuals of where the model and the dataset are at variance can simplify any remaining characterization tasks. A first order approximation may be defined and residuals successively removed from the model by augmenting the model with higher order approximations. However, a model places definite conceptual limits on the parameter-space. These limits may preclude the definition and execution of significant experiments. This completes a brief summary of the benefits and risks of data models.

2.1 Principles of EDA

It has been previously noted that one possible undesirable result of using mathematical models to describe data is a loss of access to novel information. For example, summarizing a distribution as the mean and variance of a Gaussian distribution is a sufficient characterization if and only if the distribution is in fact Gaussian. If it is not Gaussian, higher order statistics or other descriptors must be used to differentiate the distribution from a Gaussian one. The residual error between the model and the dataset must be accounted for as its significance to problem solution has not yet been determined. On the other hand, for purposes of pragmatism, datasets require a high degree of summarization in order for the analyst to understand the structure. Thus, insufficient summarization is one aspect of constraining the communication of data characterization. Data characterization which observes the following principles reduce the probability of mis-characterizations while summarizing.

- **Internal structure must determine the external model:** parameters are data-driven.
- **Samples must be included rather than excluded:** Use all the data. Severe noise may necessitate the use of robust measures. In this case, the weight of noisy samples is decreased, but not nullified.
- **Display structural information:** Use visualizations of the data and its associated statistics. Augment traditional data displays to facilitate information

communication at an intuitive level. Use images to augment statistical summaries. Images allow latent information (otherwise observable only in higher order statistics or by heuristics) to be expressed or at least adumbrated, rather than silenced in a truncated scalar.

Model Free Analysis

Model free analysis is synonymous with EDA. Model-light analysis refers to the use of minimal, or general, models. Approaches to model-light analysis are often based on different definitions of novelty: factor analysis and principal component analysis (PCA) use (co-)variances [Li]; cluster analysis examines localized structure; independent component analysis (ICA) [Hyva] examines statistical independence [Duann]. All of the model-light analysis approaches listed above have been used for fMRI. Discussions of model based analysis methods and definitions of both novelty and noise follow.

Model Based Analysis

A common model based approach to fMRI analysis is the generalized linear model (GLM) [Kryza] which decomposes the measured TC y_i into the sum of a scaled true time course x_i and a residual error ϵ_i

$$(1) \qquad y_i = \beta_i x_i + \epsilon_i$$

GLM is ultimately unsatisfying as it rarely leads to the discovery of novel region of interest (ROI) or TCs. However, it may be used to detect linear trends in the data and

allow for trend correction. Trendy samples can also be removed from the dataset before further analysis occurs.

Novelty and Noise

Definitions of novelty and noise are context dependent. Noise at high levels will entirely mask novel patterns. Determining whether any single sample is noise or novel may pragmatically be determined by statistical tests on the total population. This statistical context determines whether an individual sample is, for example, an outlier and depends upon hypotheses about the sample pool. This approach may also be followed for sub-groups of samples. The set of all hypotheses about the novelty of a sample with respect to the sample pool is the *sample novelty context.* Similarly, all hypotheses about the novelty of a collection of samples is the *sample collection novelty context.* These novelty contexts act as a null hypothesis to be rejected or accepted based on empirical distribution probabilities.

The importance, or cost, of false negatives and false positives must also be weighed in analyzing a classification or decision system. What is the cost of mis-classification? A loss, or risk, matrix details penalties associated with different types of mis-classification errors. Consider an example of determining novelty. Given a uni-modal Gaussian

distribution, how can one determine which samples, if any, are novel? Measures of novelty inevitably compare a single sample or a small group of samples against global statistics. A novelty context must be constructed and its probability determined with respect to the data distribution. Figure 1 shows a single cluster with possible outliers.

Group Representations

Clustering, examined in Chapter 3, involves the representation of a collection of samples by a single archetype sample, the centroid. The centroid defines the centre of a cluster with geometric characteristics derived from the samples in the cluster. Alternately, the term cluster refers to a collection of similar samples, while centroid refers to a feature vector representative of the cluster. While many clustering algorithms define the centroid to be the feature-wise mean of the cluster, alternate centroid definitions are beneficial. For example, a weighted mean reduces the influence of outliers on the centroid value and the cluster shape. A group representation problem requires that each sample in the cluster be faithfully represented by the centroid and that the centroid minimizes an error function in terms of the samples in the cluster. Note that error is accumulated over each cluster and often considers residual errors between samples and their associated centroids.

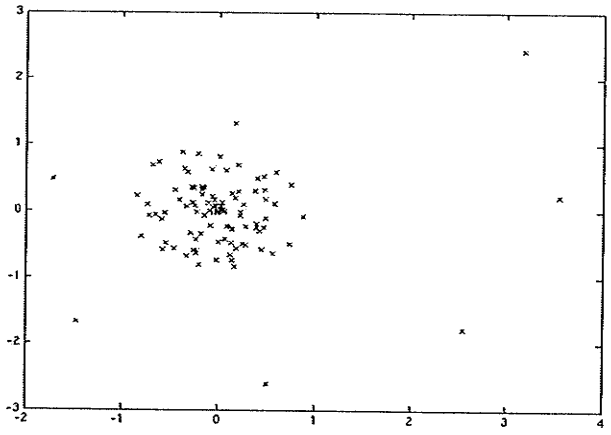


Figure 1. Structure ambiguity.

A single cluster with outliers or a spiral extending from a cluster?

The error function defines the centroids by its minimization. Common group representatives, or centroid definitions, include: mean, median, heuristic based representation where rules determine sample inclusion / exclusion.

All of the above group representations may also be weighted to modify sample contributions. Common error functions are: mean square error (MSE), weighted mean square error, distance metrics based on correlation, existence of the centroid in the original dataset. It may also be significant that the centroid lie in a part of the feature space satisfying external criteria. Given that samples may contain noise, robust statistics are one way to increase confidence in centroid definition.

Robust Statistics

Robust statistics have a significant tolerance to the contamination of the samples to which they refer. That is, the addition of noise to the dataset will change the value of a robust statistic (e.g. median) much less than that of a non-robust statistic (e.g. mean). Statistics may be characterized by their breakdown point. A breakdown point for a statistic is the number of worst case samples required to generate an arbitrary value for the statistic [Huber], see Table 1.

Heuristics for Group Representatives

Consider the following example of a heuristic based centroid definition. Let the centroid be the sample that, being present in the original dataset, also best characterizes the neighbourhood around a sample. Let the neighbourhood around sample x be defined using distance and similarity thresholds (T_d and T_s). The neighbourhood of x is the set of all samples which exceed the similarity threshold and are within the distance threshold.

Table 1. Statistical breakdown points.

Statistic	*Breakdown point*	*Equation*	*Explanation*
Mean	$\frac{1}{N}$	$\bar{X}=\frac{1}{N}\sum_{i=1}^{N} x_i$	Not robust. A single large outlier can cause the mean of a set to take on an arbitrary value.
Median	$\frac{N}{2}$	$X_{median}=Y_{\frac{N+1}{2}}$ if N is odd $X_{median}=\frac{Y_{\frac{N}{2}}+Y_{\frac{N+1}{2}}}{2}$ if N is even	Robust. Half of the samples in a set would need to take on large outlier values before the median of the set would become an arbitrary value. Note that Y=sort(X).

For x,y∈X, consider a neighbourhood G_x around x where y∈G_x if and only if y is similar to x and y is close to x. A representative y_x for the neighbourhood G may be defined as the median of the samples in G_x

$$G_x=\{y|S(x,y)>T_S \wedge d(x,y)<T_d\} \quad (2)$$

$$y_x=\text{median}(G_x) \quad (3)$$

An extension of this for sulci may require that the representative be spatially located in a region type shared by most of the samples in the neighbourhood (e.g. white or gray matter). Since the sulci have irregular shape, small deviations from the centroid may result in significant change in voxel characteristics (e.g. moving from gray matter to white matter).

Relative Sample Importance

The detection of outliers in a dataset is important and samples near class boundaries may have more significance for classifier design than those around the class centroid or mode. EDA principles suggest that all samples should normally be considered. Rejecting any single datum should be temporary and generally used for the investigation of dynamic properties of the membership values. This would occur in solution stability testing where changes to the centroid are measured when one or more samples are excluded from consideration.

2.2 Pattern Recognition

Pattern recognition is the identification of recurring structure in a dataset [Alex2] dataset. Pattern recognition is a challenging problem because sequences do not always contain exact replicas of individual forms or pattern atoms. This is true whether the patterns are simple or complex. Relations between individual samples are discernible by displaying a table which quantifies the similarity of each pattern to every other. Intra-sample distances, using one of a variety of metrics, may be substituted for a similarity measure. Comparisons are *relative*, when only the samples are considered, or *absolute*, when the comparison includes reference to an external standard. The respective merits of common metrics and similarity measures will be discussed later in this chapter. It is common to

augment datasets with a single additional feature for each sample, called a class label, which denotes an authoritative categorization of the sample. For example, samples may be categorized as healthy or diseased tissue samples.

A common goal of pattern recognition is the design of an automated classification system. Classification requires that the relationship between discriminatory features and given class labels be made explicit, either in the form of an equation, or given implicitly, in a heuristic decision process. It is often desired to select the minimum number of features possible that allow class-based discrimination of the samples. This feature selection process determines the discriminatory ability of different feature subsets with respect to the class labels. It is advantageous when a simple classifier design, with few parameters, is being used or when over-fitting of the data to the model is to be avoided. In the process of feature selection, a mapping between sample features and class labels is optimized. Many techniques exist both to discover this optimal relationship and to exploit it once it is discovered. Classifiers have many incarnations, such as a rule base, an algebraic equation to discriminate between objects, or a neural network. Classifier implementations are generally selected to fit end use criteria such as:

1. Adaptation to additional, as yet unknown, samples.
2. Computational time of classification.
3. Memory use and access times.

4. Hardware restrictions, e.g. weight, power, specialized transducers.
5. Environmental conditions, e.g. extreme temperatures.

Classifiers may also be combined into consensus groups [Hoth] to improve overall specificity or sensitivity. This section discusses only supervised classification where class labels have been provided by a reliable authority. The identification of structure can be aided by examining subsets of the data using domain-specific metrics, statistical properties or heuristics. These subsets define local classification rules. For example, a sample grouping having two different class labels may be separable or discriminated using a linear decision rule. Such a rule would, most likely, be valid only for that specific location. Many local, linear classifiers may be aggregated to define a global and non-linear decision rule. Partitioning the sample space, or considering local sample groupings, along with a partition-induced reduction in samples, where a sample group is replaced by a single representative, increases the feasibility of problem solution. Such an approach also allows the designer to enhance or refine the classifier in a piece-meal manner. When only representatives of the original samples are considered, the problem is said to be a redacted, simplified or reduced problem. Reduced problems are generally easier to solve and often may be mechanically extended to the original problem. A solution to the reduced problem can be extended to the original problem by re-adding

samples and refining any local classifiers affected. Reduced problems are also more likely to be solved by simpler algorithms. One method of reducing a classification problem considers discriminating between subsets of samples. Replacing classes of samples by cluster centroids results in a trivial separation problem that can be mechanically extended. The solution to the reduced problem is a first order approximation to the solution of the original problem. Any reduced solution can be subsequently supplemented by higher order terms and refinements to the discriminant function affect a decreasing number of samples[1].

Figures 2-4 demonstrate various two class problems in classification. In Fig. 2, a linear classifier discriminates between the dark (o) and light (x) samples. Figure 3 shows that a classifier composed of the line segments is sufficient to discriminate between the two classes. Finally, Fig. 4 shows a distribution in which no simple linear nor non-linear classifier separates the classes. One approach to discriminating the samples in Fig. 4 is to tessellate the feature-space until a number of simple classifiers prove sufficient.

1 Refining a classification system by modifying or adding local classifiers ensures that the number of samples affected is monotonic non-increasing.

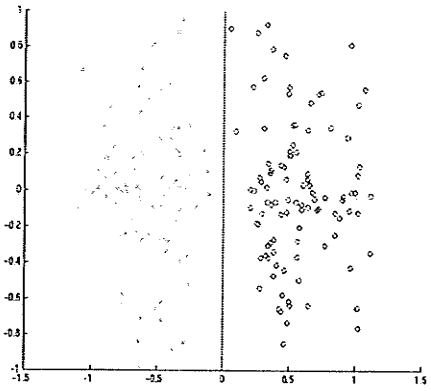


Figure 2. Linearly separable data.

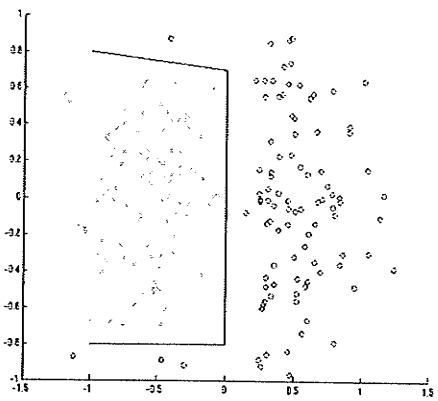


Figure 3. Non-linearly separable classes.

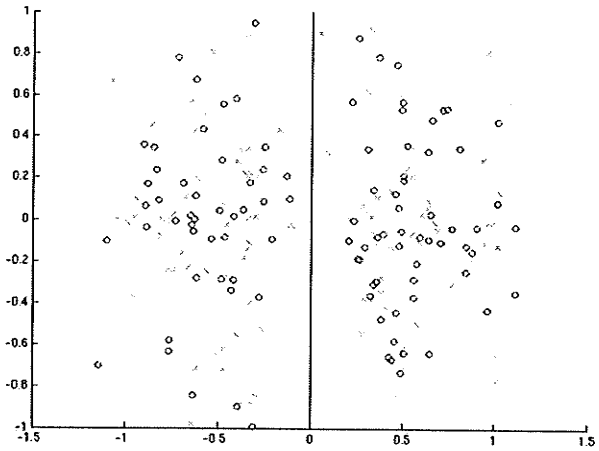


Figure 4. Classes inseparable by a single, global, linear classifier.

Alternately, transformations which map the samples into a more tractable feature space may be sought. One transform method uses cross-products of feature pairs as additional features. New features are non-linear combinations of original features and may enhance

classification quality. Another transformation defines a mapping to a higher dimensional space, for example in methods on differential manifolds, where discriminatory hyper-planes may be found by routine methods, for example linear discriminant analysis (LDA).

2.3 Classification Fundamentals

Höppner describes classification as the determination of a mapping from a feature space to a classification space labels. Many methodologies may be used to determine the map, or classifier, to implement sample decision rules. These rules assign class labels to each sample or reject the sample from all classes whatsoever. A brief taxonomy of classifiers follows with examples.

Classifier Taxonomy

A taxonomy of classifiers [Jain], see Fig. 5, may be based on the level and use of a priori knowledge. Given complete prior information, the Bayes classifier optimally discriminates between samples. When one has incomplete statistical information, supervised methods and unsupervised methods are considered. Supervised methods are further subdivided into parametric classifiers, which determine optimal values for the parameters based on the data distribution, and non-parametric techniques such as

distribution estimation, for example maximum entropy, and artificial neural networks (ANN). ANNs use a subset of the data as a training set to learn a discriminant function. This discriminant function is then validated on the remaining samples. Unsupervised methods attempt to resolve signals from mixed sources, as in blind source separation [Amari]. In the absence of labels, cluster analysis is a common choice. This exhausts a review of the first taxonomy.

Another classifier taxonomy may be constructed by considering the end result of classification, see Fig. 6. The taxonomy is induced by considering the intermediate and final organizations of the dataset, that is, how the samples are mapped to class labels. Mappings of the samples to different classes may be hard, or exclusive, when a sample belongs to only one subset. Mappings are soft, or non-exclusive, when a sample belongs to one or more subsets. These in turn may be subdivided into extrinsic-goal-oriented (supervised) methods and intrinsic-goal-oriented (unsupervised) methods. Finally, the dyad hierarchical versus partitional is used to characterize the process of forming the clusters. A hierarchical method imposes a top-down or bottom-up order while partitioning is more general and may involve several clusters of comparable size. Other possibilities for taxonomies exist, including those based on:

- Discriminant functions:

- Linear, e.g. linear discriminant analysis (LDA) [Pizzi4].
- Non-linear, e.g. quadratic discriminant analysis (QDA) [Croux].

- Mode of classification:
  - Hierarchical, e.g. decision trees [Quin].
  - Non-hierarchical, e.g. multilayer perceptron [Bish].
- Class label use:
  - Supervised, e.g. Bayes classifier.
  - Unsupervised, e.g. FCM.
- Classifier training/optimization:
  - Iterative, e.g. self-organizing maps [Koho].
  - Non-iterative, which is often analytic, e.g. matrix inversion.

Classification performance measures, such as kappa scores or class label entropy, are used to select specific algorithms for a dataset.

A final taxonomy for classifiers can be considered using the Vapnik-Chervonenkis (VC) dimension [Schol]. The VC dimension considers the ability of classifiers to categorize arbitrary class labels. All sample to label mappings for the dataset are considered. The VC dimension for a set of classifiers F is the maximum number of vectors h that can be divided into 2 classes in all 2^h possible ways using any classifier $f \in F$.

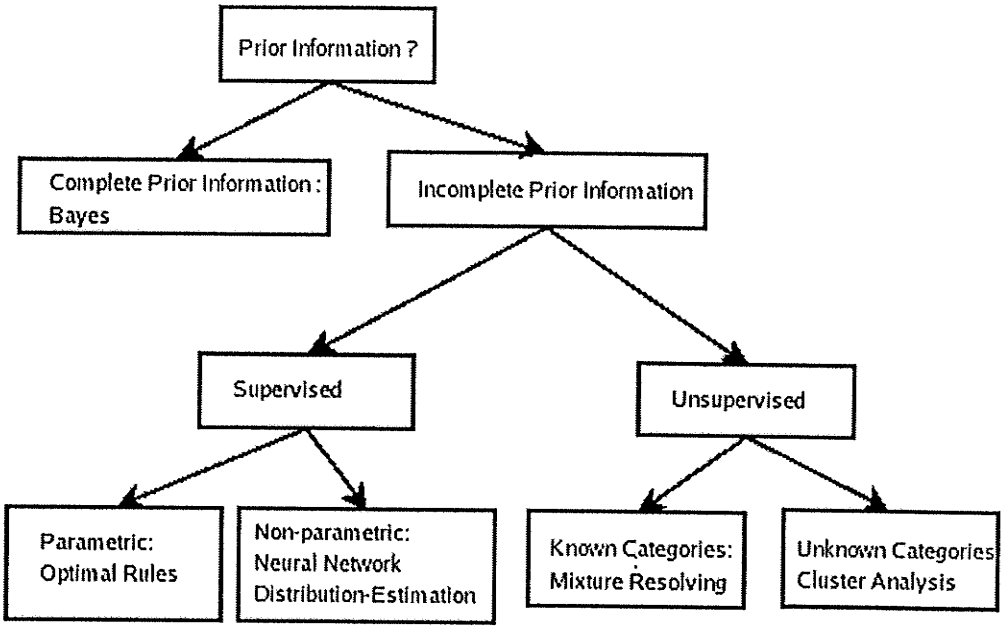


Figure 5. Classifier taxonomy based on a priori knowledge.

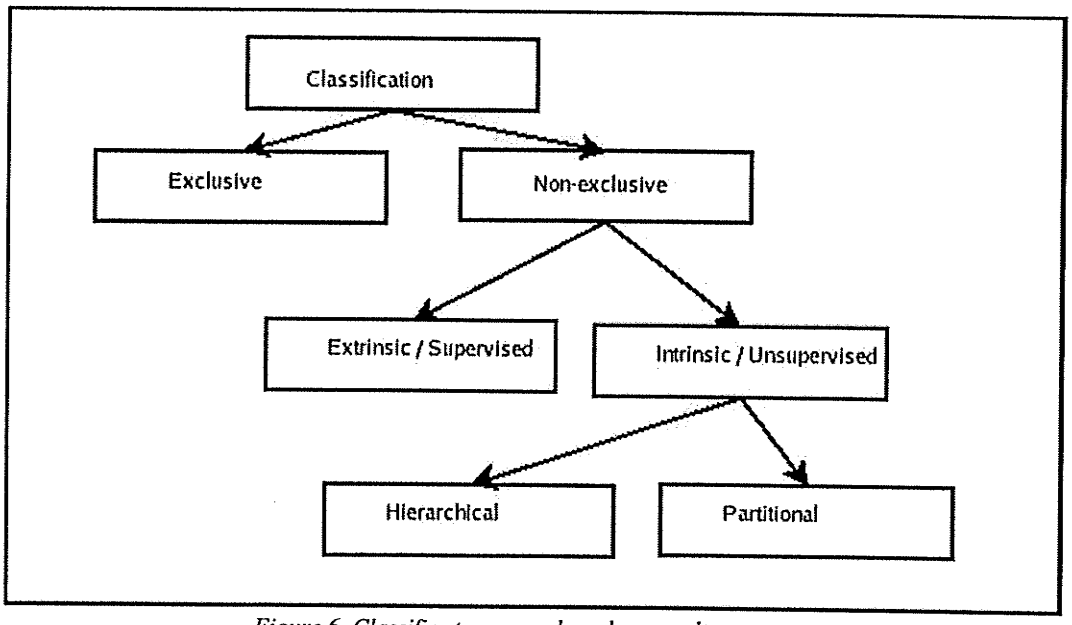


Figure 6. Classifier taxonomy based on result structure.

After choosing a classifier, one must consider implementation details. Computer

languages and libraries must be decided upon as well as whether a parallelization of the algorithm [Ben] [Kita] [Pela] is required. Parallelization is paramount when algorithm execution time is critical and execution on a multiprocessor system, such as a Beowulf cluster, is possible.

2.4 Preprocessing

The crux of analysis is the decomposition of an object into its primary components or in terms of first principles. Analysis is the process by which a complex agglomerate is reduced to a series of fundamental descriptors. A set of descriptors is the qualitative and quantitative values associated with an object. A discriminating descriptor is a descriptor that can be used to distinguish samples from two or more classes. In this reductive process, qualitative and quantitative descriptors characterize a complex object. For example, consider a simple analysis of a sequence of ordered numbers. Essentially, statistics are computed and the sequence compared to archetypes in the domain of interest. Goals of preprocessing include removing noise from the data and transforming the data into a tractable feature space for the classifier. Transformations generate discriminating descriptors for subsequent classifier design. Specific preprocessing steps include: reducing the number of features [Bryl], transforming the features to aid human expert intuition [Alex1], or high-lighting correlated variables [Cox1]. Feature selection

and preprocessing operations are guided by classification rules of thumb, e.g. the minimum number of samples per feature per class required to design a robust classifier is 10. Perceptions on possibly contaminated data also figure prominently in the application of preprocessing steps. The literature shows that genetic algorithms (GA) [Arei] [Band] are commonly used to select features. Standard preprocessing operations accumulate as heuristic rules of thumb in distinct application fields.

Signal Processing

The main task of signal processing is the transformation of the signal into a more desirable space [Elli] [Fran] [Wang1]. Here, desirable may take on a variety of meanings. If the signal is to be stored, the transformation may be compression. If two signals are compared, perhaps only portions of the signals are relevant; a window may be applied or an average value computed. Signal analysis is the reduction of the signal into a small set of descriptors. This may be accomplished through a characterization of a given signal in terms of another time series, or as a decomposition of the signal into projections. The analyst modifies the signal processing operations used based on the utility of the generated descriptors, such as residual errors between approximations. Thresholding and filtering [Just] [Tyan] are common operations used to enhance signal components, enable computation of more robust descriptors, or limit the total possible

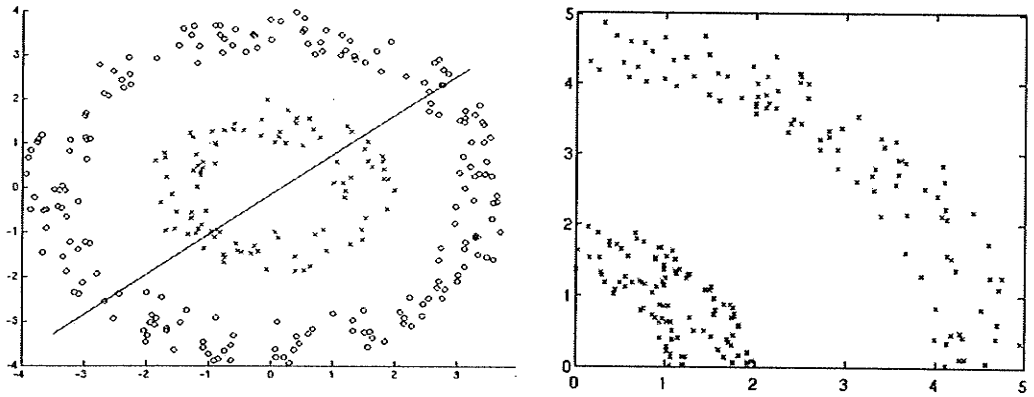
number of descriptors able to be generated. A common signal decomposition is its Fourier transform or wavelet transform [Alex6]; a popular current descriptor for lung sound signals is fractal dimension and Renyi entropy [Gnit]. Common signal processing operations include detecting periodic components and reducing noise through filtering.

Image Processing

The purpose of image processing is similar to signal processing. It is the transformation of a matrix of *pixel* values into a discriminatory descriptor, a summarizing feature or statistic, such as texture. Some image features are: number of regions, region size, shape, texture, and object placement. Common operations to facilitate these computations are: image (2d) filtering [Gonz], segmentation [Hara] [Wool], thresholding [Seul], and mathematical morphology [Serr].

Preprocessing for Classification

Data preprocessing is a common precursor to classification. As an example, consider concentric annuli, see Fig. 7, where samples in each annulus belong to different classes.



(a) Data with original features where a linear discriminant function fails to discriminate.

(b) A transformation of the original dataset that permits linear discrimination.

Figure 7. A problematic dataset for a linear classifier solved by preprocessing.

No global, linear classifier exists for such a distribution: the best case may be that a classifier is wrong[2] fully half of the time. However, using the transformation $f(x,y)=x^2+y^2$ as a preprocessing step remaps the samples to a space where a linear classifier exists. The samples in the inner annulus are transformed to the first quadrant close to the origin. Samples in the outer annulus are also in the first quadrant but are further from the origin than the other samples. The gap between the two distributions in the transformed space is $r_g = r_{2i} - r_{1o}$, where r_{1o} is the outer radius of the samples on the inner annulus and r_{2i} is the inner radius of the samples on the outer annulus. The existence of a linear classifier now depends only on the gap r_g.

2 The assumption that the classifier will be wrong half the time follows under additional assumptions: the samples for both annuli must be uniformly distributed with respect to the radius; the discriminant line must pass through the origin; the number of samples in each class must be the same.

2.4.1 Normalization

Normalization is a critical preparatory step for generating generalizable inferences from a dataset. Features that have different ranges, average values, variance, or other higher order statistics can confound applications when assumptions about the dataset and the mathematical model of the classifier are dissonant. Normalization strategies, for example, remedy non-Gaussian distributions through a whitening transformation.

Many classification algorithms assume that each feature is equally relevant to discriminating the samples with respect to the problem at hand. When normalization does not occur, features with large values or ranges can dominate metrics, dwarfing contributions from discriminatory features with smaller values or ranges. In such a case, overall classifier performance is constrained below any intrinsic dataset limits. An assumption of equal relevance between features is usually defensible as part of initial experiment scaffolding, but should not be left unquestioned. Further, when normalization transforms are used, it may be necessary to apply an inverse transformation to the analysis before the results can be applied. Some common normalization elements are:

- subtracting the mean or median.
- dividing by variance, or median absolute deviation (MAD).
- scaling to [0,1], [-1,1] , or [0, N] for $N \in \mathbb{R}_+$.

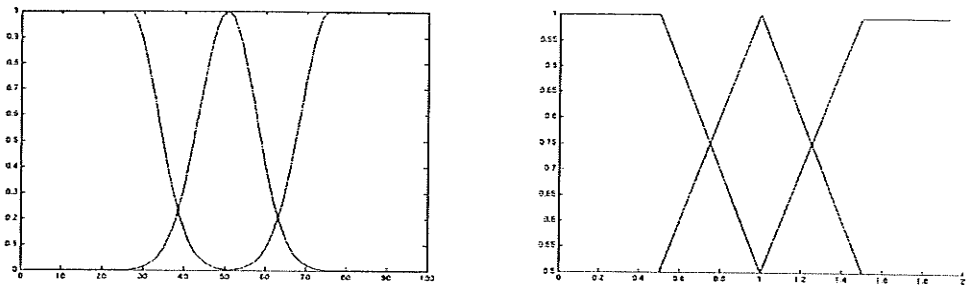
2.4.2 Transformations

Recall that a *basis* is a set of vectors which span a vector space. Transformations that simply exchange the original basis for another projection maintain the intrinsic dimensionality of the dataset. One problem related to the dataset dimensionality, or the number of sample features, is the curse of dimensionality. This curse originally noted the exponential growth of a hyper-volume as a function of dimensionality [Bellman]. In the field of pattern recognition, the curse refers to the exponential increase in difficulty of determining optimal discriminant functions when feature cardinality increases linearly. To avoid this imprecation, data can be processed with the blessing of dimensionality reduction, or feature selection, techniques. Adhering to robust design principles is one motivation for feature selection in classifier design. One such design principle limits the number of classifier parameters to a multiple of the number of samples per feature per class in the dataset. Two examples of transforms that exchange bases are principal component analysis and independent component analysis. A brief discussion of both techniques follows.

Fuzzy Sets

Set theory is at the heart of classification since the fundamental question is: to what class

does a sample or a group of samples belong? Fuzzy sets extend the binary categorization of classical set theory, in which an object either is or is not a member of a set. In a fuzzy set, objects are members *to a degree* and an object may belong to multiple sets to different degrees at the same time. This concept is explored more fully in the appendices. Fuzzy sets [Pizzi1] [Pedr2] [Wolf] [Zadeh] may be used to embed the samples into a higher dimensional space. While this initially seems counter-intuitive with respect to the preceding discussion on feature reduction, additional degrees of freedom in the resulting space may have desirable properties, such as differentiability or piece-wise continuity. These properties facilitate the determination of discriminative hyper-planes, with the caveat that they may not generalize well. Embedding the original feature space in a high dimensional manifold often enables simple discriminant function to be found through analytic methods. Fuzzy encoding may be viewed as feature membership in a set of feature groups. An obvious grouping uses membership in fuzzy sets based on feature magnitude. The original feature value is encoded by a vector denoting degrees of membership in a series of fuzzy sets, for example the fuzzy sets SMALL, MEDIUM or LARGE shown in Fig. 8.



(a) Gaussian fuzzy sets covering a value range. *(b) Triangular fuzzy sets covering a value range.*

Figure 8. Fuzzy sets denoting linguistic terms.

Removing Noise

Preprocessing operations may also expunge samples from the dataset in an attempt to eliminate or reduce the presence of noise. Many strategies exist to detect noise and then ameliorate the contaminated signals. Image filters can enhance contrast, detect edges, de-trend, and de-blur. Statistical analysis on signal intensity can also detect noise. For example, images with salt and pepper noise, isolated white and black pixels, can be improved by replacing the extreme pixel values by the average value of their neighbours. However, globally replacing pixel values by neighbour averages, as in lowpass filtering, reduces the salt and pepper noise but introduces a global blur.

Linear trends are one type of noise that occur in one-dimensional signals and may be addressed by de-trending, or removing the linear component of the signal. Identification of the trend $y=mx+b$ may be accomplished by linear regression.

Noise can also be addressed in the frequency domain. The Fourier Transform transforms the time series into the frequency domain of sinusoid amplitudes. Noise occurring at certain frequencies or ranges can be reduced or eliminated by band-reject and bandpass filters. Wavelet processing [Alex6] is also popular since wavelets provide a dual representation with temporal-frequency localization. High resolution localization enables the manipulation of frequency components at specific time intervals.

2.4.3 Metrics and Measures

Selecting an appropriate comparison method is critical in EDA. Metrics and measures exhibit different cost-functions (loss matrices) associated in assigning samples to the different possible classes. Metrics and measures can also embody relational rules between samples in terms of class labels. In this case, the decision rules change from general equations to heuristics dealing with sample cases.

As an example of Pearson correlation, Fig. 9 depicts a random sequence, a linear trend, and a scaled combination of the two. Correlation coefficients for the random sequence, with and without trend, are given between the random signal and the straight line. The coefficients are $\rho = 0.09$, for the random zero mean signal, and $\rho = 0.72$ for the signal combined with the trend. Others measures include: cosine, squared chord, squared chi squared.

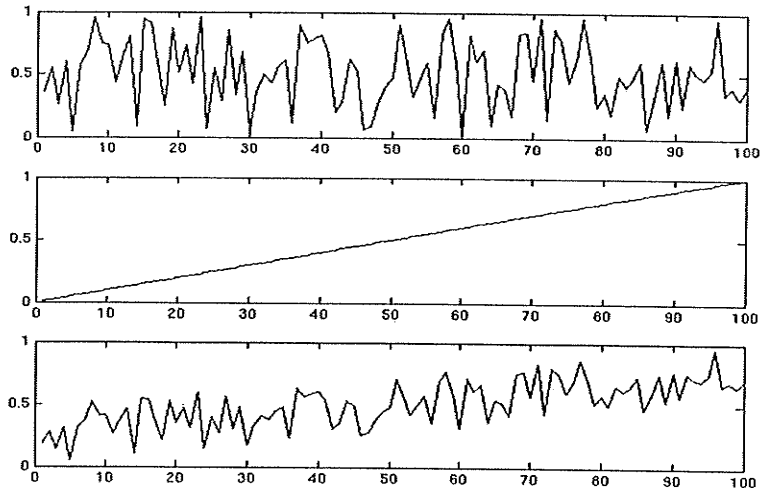


Figure 9. Trends in data.

(top) A random signal ρ =0.09, (middle) a linear trend ρ =1.0. (bottom) a random signal with trend ρ =0.72.

Thresholds on distance or similarity are often used to reduce the number of samples in a dataset or define a region of interest (ROI). A reduced dataset, X_s, where each sample has a minimum similarity to a prototype *y* is generated by a similarity threshold T

$$X_s(y)=\{x|S(x,y)>T\} \quad (4)$$

2.4.4 Experimental Concepts

Scientific theories are objective in the sense that the hypotheses and theorems are falsifiable. Falsifiability is the ability to devise and implement a test that could show the theory to be false. Hand in hand with falsifiability is the repeatability of experiments; an experiment performed at one laboratory must be able to be replicated at another similar

laboratory. When executing an experiment, the limitations or peculiarities of algorithms must be considered and addressed. When algorithms use random initialization, the initialization must be the same for comparable tests (the same random seed should be used), or a large number of tests should be executed. In supervised learning with training and test sets where the assignment of a sample to the training or test set is arbitrary, multiple training and test sets should be used. Using multiple training and test sets is called n-fold cross validation. The classifier design is validated by n unique training and test sets drawn from the same dataset. At its extreme, n-fold cross validation becomes leave-one-out (*LOO*) cross validation and each validation uses all but one sample in the training set. The remaining sample is the test set. Other approaches also use sampling to form conservative estimates of results. Monte Carlo techniques use random sampling to quantify properties. Sampling also takes place in the parameter space of the algorithm. Since many parameters will be independent of each other, an exhaustive testing of each combination may not be feasible. However, a grid may be constructed for integer or real parameters and individual nodes (a set of parameter values) evaluated. Each method above prepares to resolve the experiment hypothesis. Until an experiment has been performed the hypothesis exists in a state of tension. Experiments must resolve the hypothesis rejection/affirmation tension and decide whether the experiment supports the

hypothesis and to what degree. An experiment can fail to resolve the tension if it fails to reject or affirm the hypotheses at an acceptable level of confidence. Confidence intervals quantify the degree to which a given hypothesis is supported.

2.4.5 Classification

Supervised classification refers to a mapping from the samples to a set of class labels. Given class labels and unique samples, it is commonplace to design an exact mapping such that features of the samples discriminate samples in the same manner as the class labels[3]. A good classifier is distinguished from a mediocre or bad classifier by its ability to classify previously unseen samples correctly. This *generalized recognition ability* of the classifier signifies that the samples used to design the classifier contain generalizable discriminating information. The standard procedure for designing a supervised classifier is to divide two disjoint subsets: a training (design) set and a test set. The classifier is designed using the training set and is justified (validated) using the test set. Inferences are made about the relations of training sample features to the class labels. The veracity of those inferences is quantified by applying these discovered relations to the test set. If the relations also hold on the test set, a confidence level can be determined that the relations will also hold for previously unseen samples. This is, of course, an invaluable

3 A lookup table is always possible.

quality for field use of the classifier. The validity of feature-label relations, or the generalization ability of a classifier, is the predictive error rate of the classifier. It can occur that class labels are tarnished (samples have incorrect class labels). In such an instance methods dealing with probabilistic learning [Greb] or burnishing tarnished gold standards [Pizzi1] will be required.

Classifier Design Considerations

Factors to consider when deciding which classifier to apply or adapt to a problem are:

- Number of classifier parameters.
- Number of samples in each class.
- Time and other constraints on implementing the classifier.
- Generalization expectations.
- Existence and quality of data labels.
- End use conditions.

Design Demons

Two design problems are over-parameterization and over-generalization. Over-parameterization occurs when the model used has more parameters than are justified by the data [Tous]. A related problem is the curse of dimensionality; a small number of samples with a large number of features have many different hyper-planes that correctly discriminate between the samples classes. This is a common problem in sparsely populated, high dimensional spaces. Since many hyper-planes are equally

discriminatory, the risk of poor generalization is high. Small numbers of samples in individual classes may necessitate use of robust measures or combining similar classes.

Over-generalization is the application of accidental correlations among features in the dataset to unseen samples that may not exhibit this chance quality. This can occur if the sampling for the training set is biased to one mode of what is in fact a multi-modal distribution.

Iterative classification algorithms have additional design considerations. The initialization procedure must be well-defined and lead to convergence without compromising the expressiveness of the resultant discrimination function.

2.4.6 Implementation

Addressing a pattern recognition problem requires a careful selection of algorithms for data processing, pattern discovery and classification. For EDA investigations, the algorithms and their parameters should be guided by data-driven processes. Another consideration is the manner in which results are presented. Confidence in the generalization ability of the classifier is a critical factor before the system goes into field or production mode. Cluster analysis is discussed next.

3. Cluster Analysis

As complexity rises, precise statements lose meaning and meaningful statements lose precision.
Lotfi Zadeh

Clustering [Bezd1] [Rieg] partitions a dataset into subsets (clusters) and defines a representative, or centroid[4], for each cluster. Determining centroids is often the goal of analysis. For accurate representation, low sample-centroid distances are critical, regardless of cluster shape or sample density. Clustering algorithms can be divided into hierarchical and C-means types[5].

Hierarchical Clustering

Hierarchical clustering partitions a dataset using a splitting or merging heuristic. Splitting a dataset begins with all samples in a single cluster and a distance matrix of sample-centroid distances. At each step, the sample most distant from its centroid is split from that cluster and defines a new, singleton, cluster. After hierarchical clustering, a series of distance thresholds may be used to examine the induced clusters. Likewise, testing a hypothesis about the number of clusters in the dataset yields a distance threshold. The number of steps required for hierarchical clustering is the number of samples less one.

---

4 Centroids are often defined as an aggregate value, and, although they represent a cluster of samples, may, in fact, not occur in the dataset. This distinction is rarely crucial to experiments and use of medoids as centroids addresses the requirement to exist as a sample in the dataset.

5 K-means clustering is a hard clustering algorithm while C-means refers to fuzzy clustering.

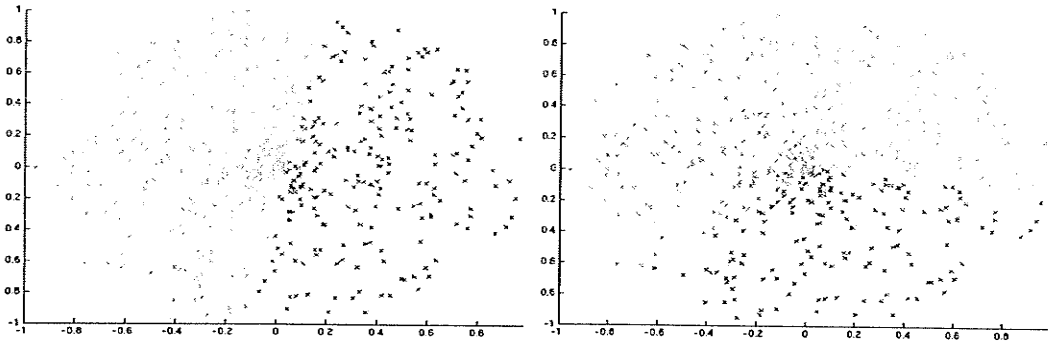
C-Means Clustering

In C-means clustering, functionals built around the sample-centroid distance form an objective function, which is then iteratively optimized. While C-means algorithms have C centroids throughout the optimization[6], centroid values change, as does sample membership in the clusters. Initial centroids are chosen to lie near the dataset centre of mass, as random samples or points in the feature-space, or based on external factors. A series of distinctly initialized runs is often used since C-means clustering is proven to converge to a local optimum or saddle point[7]. The number of iterations for convergence is not known beforehand, and termination criteria, such as U_Δ, may require tuning. Figure 10 shows two sets of FCM clusters resulting from different parameters. Criteria for determining convergence of the algorithm are monitored: the overall change in the membership partition is compared to a threshold and the number of iterations executed is compared to a maximum count. Either criteria terminates the algorithm. Changing the number of clusters, or any other parameter, requires that the C-means algorithm be re-executed. Benefits of C-means clustering with respect to hierarchical algorithms include: faster convergence for datasets with high dimensionality and tighter

6 This is true unless merge or split heuristics are also used, as in ISODATA.

7 Superficial differences between clustering runs may include permutations of the centroid order, alternately, the columns of membership matrix U are permuted.

clusters for globular data distributions.



(a) A single intrinsic cluster represented by two centroids.

(b) A single intrinsic cluster represented by three centroids.

Figure 10. Multiple representations of structure in a dataset.

3.1 Fuzzy C-Means Clustering

C-means clustering where samples have partial degrees of membership in clusters is called fuzzy C-Means (FCM)[8]. FCM is:

1. objective-function based, minimizing sample-centroid distances.
2. Constrained such that sample membership in all clusters sums to unity.
3. an alternating optimization algorithm where two sets of equations, for centroids and sample memberships, are updated in alternation.
4. unsupervised classification method (although supervised variants exist).

Recall that *unsupervised* refers to the case when only the data features, and not the class labels, determine the resultant classifier design. Clustering notation follows that of Bezdek [Bez1]. Let X be a n×f data matrix, n samples each with f features, with sample

8 In the literature, FCM is referred to as probabilistic clustering since the membership values sum to unity and can be interpreted as probabilities. When FCM is used to cluster time series, data it is referred to as fuzzy temporal clustering.

$x_i=[x_{i1},x_{i2}...x_{if}]$

$$(5) \qquad X=\{x_1, x_2 ... x_N\}^T \subset \mathbb{R}^D$$

There are C centroids

$$(6) \qquad V=\{v_1, v_2 ... v_C\}^T \subset \mathbb{R}^D$$

A C-partition matrix, or membership matrix, is defined as the n×C matrix

$$(7) \qquad U=[u_1 ... u_n]^T$$

The membership of sample x_i in centroid v_j is u_{ij} and indicates the degree to which sample x_i is associated with the cluster centre v_j. The vector $u_i=[u_{i1}\ u_{i2}\ ...\ u_{ic}]$ denotes the membership of x_i in all the C centroids. The maximum value of u_i determines the assignment of a sample to a cluster[9]. One benefit of allowing a sample to have partial membership in multiple clusters occurs when samples lie mid-way between two or more centroids. Sample-centroid distances will then be similar for two or more centroids; memberships should be correspondingly similar.

Figure 11 shows samples that lie at equal distance from two cluster centroids. Samples equidistant from both centroids have the same membership for each clusters.

---

9 Note that individual clusters may or may not correspond to sample classes. Samples with different class labels may in fact belong to the same cluster due to similar feature values. An indicator that a complex classifier may be needed to discriminate the samples occurs when class labels for a cluster are both trusted and heterogeneous. Membership values ultimately define a centroid label for the sample. A further step assigns class labels to clusters.

Classification is ultimately decided on a coin-toss and no differentiation is made based on overall distance between the samples located on the centre of mass mid-line.

FCM parameters, with default values, are shown in Table 2. Although m=2 is commonly used for general data [Bezdek], for noisy fMRI data, $1.1 \leq m \leq 1.3$, is recommended [Jarm].

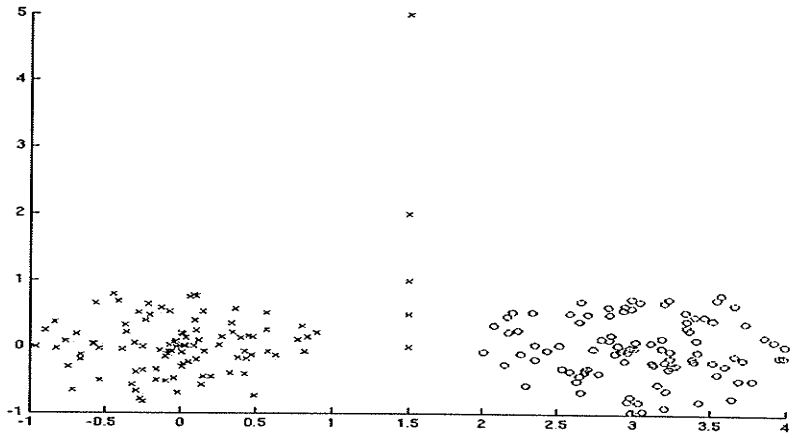


Figure 11. FCM outlier heuristics.

Table 2. FCM Parameters.

Name	*Symbol*	*Comment*
Fuzzification exponent	m=2	Controls shape of membership functions.
Number of clusters	C=20	Use of large values of C is standard.
Membership change threshold	U_{Δ}=0.001	Termination criteria for membership matrix.
Maximum iterations	T_{max}=50	Termination criteria for iterations.
Distance metric	$\|x,v\|$	Euclidean distance.
Membership change metric*	$\|U_t,U_{t-1}\|$	Euclidean distance.
* Note that the membership-change values can be exchanged for centroid-change values.		

It is worthwhile to discuss the effect the properties of the fuzzification factor m as it relates to the metrics used in the algorithm. Let $alg_{FCM}(U)$ denote FCM as it iterates on membership matrix U.

- $\lim_{m\to 0} alg_{FCM}(U)=alg_{HCM}(U)$ As m tends to 0, FCM tends to hard clustering.
- $\lim_{m\to 1} alg_{FCM}(U)=alg_{WM}(U)$ As m tends to 1, FCM tends to a simple weighted mean (WM) clustering.
- $\lim_{m\to \infty} alg_{FCM}(U)=U^*$ As m tends to ∞, FCM tends to assign equal centroid membership for each sample. Then

$$(8) \qquad u_{xv}=\frac{1}{C} \quad \forall u\in U^*, \forall x\in X, \forall v\in V$$

Algorithm Convergence

Bezdek [Bezd1] [Bezd2] [Hath1] [Kim] has shown that the convergence requirements for alternating optimization (AO) of the cluster algorithm are the same as the Karush-Kuhn-Tucker conditions. FCM is an alternating optimization problem because the optimal centroids are obtained by alternately updating both the centroids and memberships. AO requires that the objective function and any additional constraints be regular or differentiable. Since FCM constrains centroid memberships of a sample to sum to unity, Lagrange multipliers are used to convert the constrained optimization problem into an unconstrained optimization problem.

The trajectory of the centroid paths can be graphed as the algorithm converges. For datasets with globular sub-clusters and random initialization, the movement toward the global centre of mass is immediately apparent. Note that each globular cluster defines a convex region in the feature space. As multiple centroids converge to the same distribution mode (sample-gravity attraction), further iterations distribute the centroids within the mode (centroids repel each other). Algorithm iteration continues until convergence (the centroids are no longer changing) or until heuristic stopping criteria are met. Mathematically, the objective function reaches a local maximum or a saddle-point. Figure 12 shows centroid convergence over time when the centroids were initially define

near the dataset centre of mass.

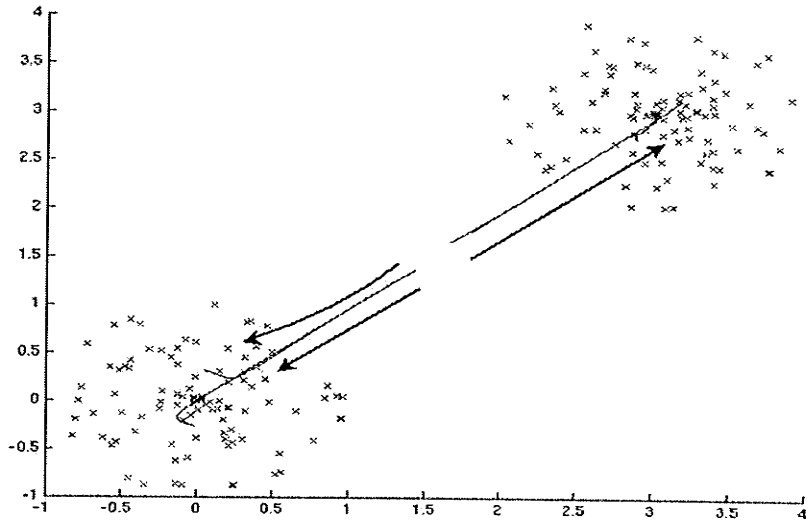


Figure 12. Centroid path convergence in feature-space.

FCM Algorithm

The iterative algorithm may be decomposed into two steps: initialization and convergence. The terminology used is listed in Table 3 and the algorithm is detailed in Table 4.

Table 3. FCM Algorithm Variables

Unlabelled dataset: $x \in X$ with $\|X\|=n$	Centroids: $v \in V$ with $\|V\|=C$, $1<C<n$
Fuzzification exponent: $m>1$	Objective function: J_{FCM}
Iteration limit: T_{max}	Objective function norm: $\|\circ\|_m$
Error termination criterion: U_Δ, V_Δ	Error norm: $\|\circ\|_\epsilon$

Table 4 refers to updating the centroids and the memberships.

Table 4. FCM Algorithm

Given: X, c, m, T, ϵ, $\|\circ\|_m$, $\|\circ\|_\epsilon$
Initialization: choose C initial cluster centres $V_0= [v_1, v_2, \ldots v_C]$
Loop: for t=1 to T Update memberships U_t using centres V_{t-1} Update cluster centres V_t using memberships U_t if $Et = \|V_t-V_{t-1}\|<\epsilon$ set $U_{final} = U_t$, $V_{final} = V_t$, $t = T$ fi end
Termination: Output final centres V_{final} and memberships U_{final}.

Update Equations for Clustering

The following steps determine the equations used in the FCM algorithm:

- Add Lagrange multipliers to remove any constraints on the objective function.
- Differentiate the modified objective function.
- Set the equation to 0 and solve for a system of equations in U and V.
- Relate the equations in time so that the most recent updates of U and V are

used. Iteration proceeds by updating U_t using V_{t-1}. V_t is updated using U_{t-1}. The relationship between the clustering update equations, distance function, fuzzy exponent *m* and convergence properties may now be stated: Given a general distance function $D_{ncA}= \| x_i-v_k \|$ and m>1, the objective function $J_{FCM}(U,V;X)$ is minimized by $(U,V)\in M_{FCM} | \mathbb{R}^D$ only if the update equations in Table 44 are used.

Objective Functions

Optimization problems quantify the degree of success for any solution in terms of an objective function. Optimization is straightforward when the objective function is differentiable. Recall that cluster analysis is concerned with discovering an optimal partition of the samples to a given number of clusters. The maximum number of iterations is used in conjunction with $U_\Delta=|U_t-U_{t-1}|$ to determine the number of iterations in which the algorithm is allowed to converge. A decrease in the value of U_Δ suggests that few not samples changed primary membership. By the same token, a measurement on the changes in the centroid values can also be used as a terminating criteria. Convergence is taken to mean that the centroids V_{final} (membership partition U_{final}) defined at the termination of the algorithm are sufficiently similar to the analytic solution V^* (U^*). That is, V_{final} and U_{final} converge asymptotically to V^* and U^*).

3.2 FCM Extensions

As noted previously, the literature on FCM is rich. There are a wide variety of both applications and extensions to FCM. For applications, see [Alex3] [Alex4] [Bozi] [Bruz] [Demi] [Devi] [Dimi] [Ermi] [Fan] [Liu] [Park]. For extensions, see [Alex5] [Bezd3] [Bobr] [Gath] [Gokc] [Gola] [Gust] [Hond] [Kole] [Kris] [Ng] [Pedr1] [Pedr2] [Pedr3] [Pedr4] [Pedr5] [Sark] [Seli] [Sgar] [Wang2] [Wu1] [Wu2]. Two common themes of variation are briefly noted that address critical limitations of FCM: cluster shape and relative sample importance.

Clustering Covariance Matrix

Clusters in FCM are hyper-spheres of the same size and, while it partitions the dataset, the intrinsic dataset structure does not always correspond to such size and shape assumptions. Ellipsoid clusters have been incorporated into FCM using a cluster covariance matrix A. This matrix is unique for each cluster and effectively scales distance along the axes [Raud]. Since A is used to compute distances, it is sensible that A be positive definite.

Secondly, recent clustering algorithms incorporate sample label information [Pedr3]. Since some sample labels are used, FCM becomes a partially supervised learning algorithm. This approach increases the contribution, or weight, of labeled samples to the

definition of the centroid. Both ideas, scaling distance and relative sample importance, recur in the development of FCMP. Many other clustering algorithms exist, many with adaptations that limit their utility for application to general data. Any one cluster algorithm may differ from another in terms of these algorithmic degrees of freedom: objective function, metric, centroid definition. A priori, dataset-specific knowledge may justify additional constraints on the clusters, such as their shape, variance and even the number of clusters. FCM is easily modified to add these constraints and ultimately leads to a more accurate representation of the data. For example, Gath and Geva introduced ellipsoidal clusters, but spherical shells, lines, parallelograms etc are all able to be defined by suitable objective functions [Höpp].

3.3 Cluster Validation

Validation methods [Halk] [Hath2] quantify the fitness of specific cluster parameters to the dataset and can be used to compare different parameter settings. Unfortunately, validation indices require that clustering be performed over a range of parameters (the number of clusters). Recent studies have shown that the Xie-Beni index consistently corresponds to good partitions for general data [Xie]. Other indices seem suited to particular types of datasets. HCM has it own validation indices. For comparison of

centroids generated by HCM to those generated by FCM it is useful to fuzzify the hard memberships (from HCM) or harden the fuzzy memberships (from FCM). Both HCM and FCM indices measure the fitness of clustering parameters to the dataset.

HCM Validation

Hard clustering uses scatter matrices to express the goodness of fit between the data and the parameters. These matrices measure the separability of the dataset and include the scatter matrix (variance of samples in a cluster)

$$S_C = \sum_{x_n \in \omega_c} (x_n - v_c)(x_n - v_c)^T \tag{9}$$

the within cluster scatter matrix (sample-centroid variance over all clusters)

$$S_W = \sum_{c=1}^{C} S_C \tag{10}$$

the between cluster scatter matrix (centroid distribution in space)

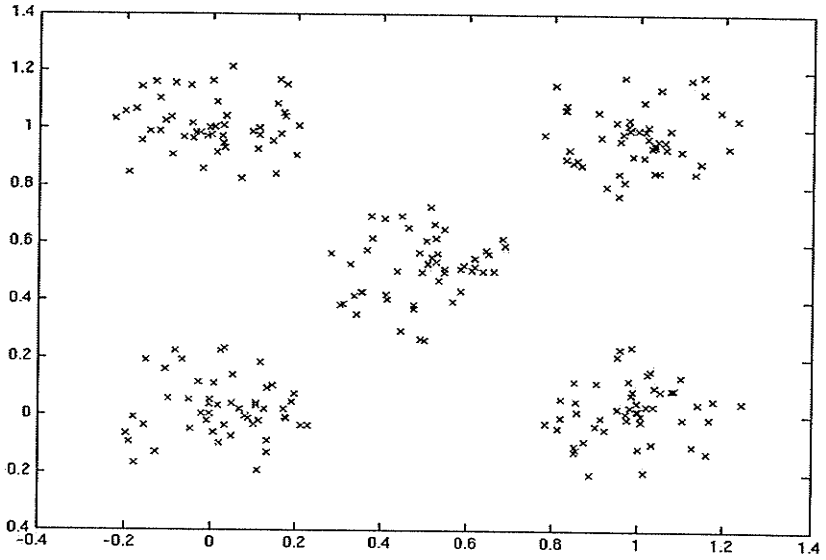
$$S_B = \sum_{c=1}^{C} (v_c - \bar{v})(v_c - \bar{v})^T \tag{11}$$

and the total scatter matrix

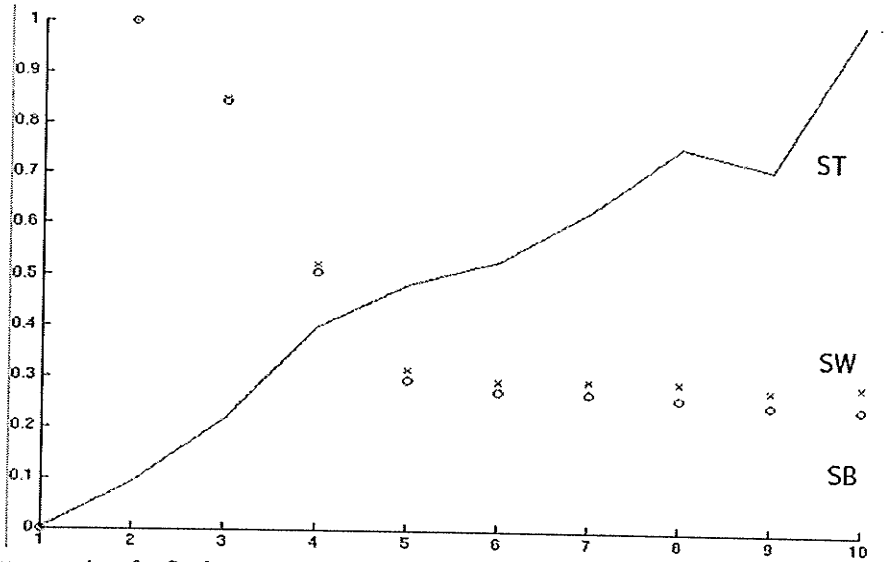
$$S_T = S_W + S_B \tag{12}$$

Figure 13 (a) shows the spatial distribution of data with five clusters. Figure 13 (b) shows scatter matrix values for the HCM algorithm. Note that the intrinsic number of

clusters is not indicated by a global minimum on the plot. Rather it is near the saddle point (point of diminishing returns). Adding another cluster decreases sample-centroid distances for the new cluster while removing samples from other clusters.



(a) Data for HCM validation indices.



(b) Scatter matrices for final centroids over a range of cluster values, C=3:10 for the dataset in (a). S_T, S_B and S_W are all normalized to [0,1].

Figure 13. HCM Validation Indices.

FCM and PCM Validation

Validation indices have also been developed to quantify the fitness of FCM parameters to the data. These indices are computed using the membership matrix. The Xie-Beni index has consistently proved itself to indicate an accurate number clusters for general data.

$$\text{Val}_{XB}(U,V;X)=\frac{\sum_{c=1}^{C}\sum_{n=1}^{N} u_{nc}^{2}\|x_n-v_c\|}{N(\min_i\{\|v_c-v_i\|\})}=\frac{(\frac{\sigma}{N})}{\text{sep}(V)},\ i\neq j \tag{13}$$

where

$$(14) \qquad \mathrm{sep}(V)=\min_j\{\|v_i-v_j\|\}$$

Others FCM validation indices are the partition coefficient (PC), partition exponent (PX), partition entropy (PE), Kwon (KW) and Fukuyama-Sugemo (FS) indexes.

The PC index rewards crisp (most unambiguous) partitions.

$$(15) \qquad \mathrm{PC}(U)=\frac{\sum_{x\in X}\sum_{v\in V}u_{xv}^2}{|X|}$$

The PE index interprets membership as a probability of class assignment. The entropy formula is then applied to the membership matrix.

$$(16) \qquad \mathrm{PE}(U)=\frac{\sum_{x\in X}\sum_{v\in V}u_{xv}\ln(u_{xv})}{|X|}$$

The FS index combines a compactness measure, J_m, with a degree of separation, K_m

$$(17) \qquad \mathrm{FS}(U)=J_m+K_m=\sum_{x\in X}\sum_{v\in V}u_{xv}^m\|x,v\|+\sum_{x\in X}\sum_{v\in V}u_{xv}^m\|x,\bar{V}\|$$

where V^* is the mean of all centroids.

The PX index is a measure for the number of partitions that classify all data better than the considered partition. An optimum is indicated for a single data assigned unambiguously. For the PC and PE, indices an optimum is indicated only for all data assigned unambiguously. Whether all memberships values are used depends on the

number of clusters. As the number of clusters increases, the number of large memberships decreases. Therefore, only the maximum memberships for each datum are used.

$$(18) \qquad PX(U)=-\ln\left(\prod_{x\in X}\left(\sum_{j=1}^{\lfloor \mu_x^{-1} \rfloor}(-1)^{j+1}\,{}_{j}^{c}(1-j\mu_x)^{c-1}\right)\right)$$

where $c=|x|$, $\mu_x=\max_{v\in V}(\mu_{xv})$ and $[x^{-1}]$ is the largest integer number smaller or equal to $1/\mu_x$. For PX, the use of all memberships depends on the number of clusters.

The KW index extends the Xie-Beni (XB) index to account for the monotonic decreasing tendency of XB with increasing c. This is implemented through a punishing function.

$$(19) \qquad KW(U)=\frac{\sum_{x\in X}\sum_{v\in V}\mu_{xv}^{m}\|x,v\|+\frac{1}{|V|}\sum_{v\in V}\|v,\bar{V}\|}{\min_{i\neq j}\|v_i,v_j\|}$$

Table 5 summarizes the range and descriptions of different validation indices. The use of these indices is examined in Chapter 8. Figure 14 shows a dataset with two clusters, each having a distinct cluster density. Validation indices differ on the inherent number of clusters on this dataset.

Table 5. FCM Validation Global Indices Summary.

Validation index	*Range*	*Description*
FS	Not given.	Measures compactness and separation.
KW	Not given.	Adds punishing function to XB as C increases.
PC	$\frac{1}{K} \leq PC(U) \leq 1.0$	Rewards crisp partitions or the most unambiguous.
PE	$0 \leq PE(U) \leq \ln(\lvert X \rvert)$ $0 \leq 1 - PC(U) \leq PE(U)$	Membership is probability of class assignment.
PX	Not given.	Requires large exponents. For uniform membership, PX = 0.
XB	Not given.	Measures sample-centroid distance and centroid distribution in sample-space.

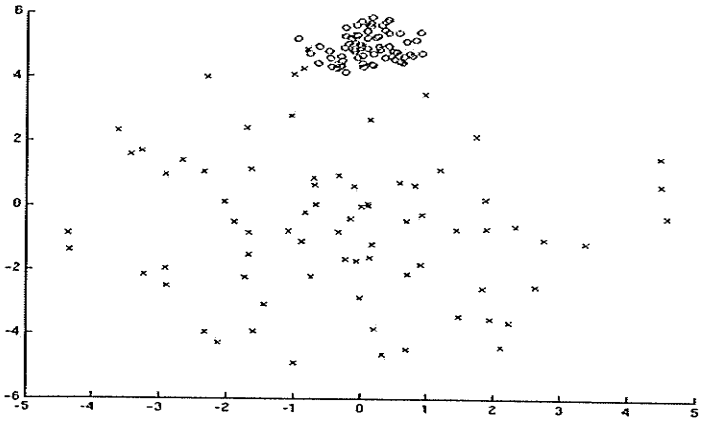


Figure 14. Dataset with two intrinsic clusters of distinct sample density.

Visual Cluster Validity

Another method to ascertain the number of intrinsic clusters in a dataset is to use Hathaway and Bezdek's [Hath2] visual cluster validity (VCV) index. VCV is a dis-similarity matrix based on a specific re-ordering of the samples in the dataset. Samples are associated with only one cluster (the fuzzy membership matrix is hardened). VCV is computed in two steps:

S1: A cluster (centroid) order is defined. The first centroid is chosen at random. An inter-centroid distance matrix determines selection of successive centroids in that the centroid closest to the previous selection is chosen. Ties are broken by a coin-toss.

S2: An ordering for the samples associated with each cluster is defined. Samples in each cluster are re-ordered using *inter-datum distances.* Hathaway and Bezdek note that the use of pair-wise Euclidean distances for inter-datum distances is best suited to well separated (or volumetric cloud) data. Therefore, VCV uses a pair-wise dis-similarity measure

$$(20) \qquad R_{ik}^{*} = \min_{1 \le j \le c} \left(d_{ji} + d_{jk} \right)$$

While many other validation indices require that the clustering algorithm be executed over a range of clusters, $C \in \{2,3,...N-1\}$, the VCV index describes the data structure

using a single set of centroids. It is suggested that a high value of C will aid VCV in depicting the intrinsic value of C. This intrinsic value is determined from the image by counting the number of dark blocks along the diagonal. Recall that dark (low) values of similarity are being depicted. A large (small) dark block on the diagonal indicates a large (small) cluster. Counting the number of dark blocks on the diagonal indicates the number of clusters while comparing block sizes suggests cluster size variance. Cluster overlap is indicated by dark (low) pixels off of the main diagonal. This illustrates samples between two clusters and can be used to identify clusters for merging.

The VCV index for a synthetic dataset composed of two distinct clusters (radius $r=1$, 100 samples per cluster, centres (x,y) = (0,0) and (5,5) respectively) is shown in Fig. 15 and clearly indicates two clusters. Hathaway and Bezdek's VCV [Hath2] is a prime example of EDA visualization for clustering. The VCV image of the sorted samples is used in an EDA manner; dark blocks (samples with low dis-similarity) suggest natural clusters.

3.4 FCM Analysis

For multi-dimensional data, the ability of a centroid to visually represent all samples in the cluster in a single plot is one of its most useful features. FCM analysis often involves plots of centroids, discovering the effect of adding more centroids, examining centroid distribution in space, developing heuristics for labeling samples, identifying outliers, and

identifying ambiguous assignment of samples (equal memberships in multiple clusters). Visualizations associated with FCM include: plots of validity indices vs. C, sample membership graphs, centroid changes; and for fMRI: voxel assignment maps.

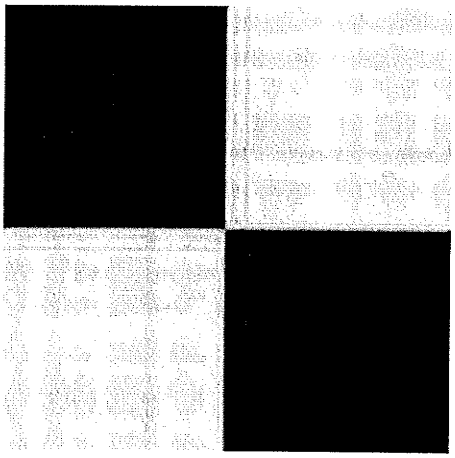


Figure 15. VCV index for dataset with two well defined clusters.

4. FCM with Feature Partitions

Facile est inventis addere.
(It is easy to add to things invented already.)
Latin proverb

FCM excels in eliciting intrinsic data structure, in part, because of the minimal mathematical model it introduces. As an EDA method, FCM is often used when authoritative data models do not exist, or are developing, and, despite varieties of FCM that exploit specialized structure, e.g. hyper-ellipsoids, a variety of FCM that leveraged known feature relationships was not found. When FCM lacks a mechanism to describe and to integrate feature relationships, information critical to eliciting overall data structure is omitted. In fMRI, and more generally, explicitly taking account of feature relations is beneficial. For example, datasets often have features acquired through different modalities, at different times, and under different conditions. Thus, individual features belong to distinct statistical distributions, differ in observation error and noise, and often have a *de facto* measurement or comparison method associated with them. In addition, the discriminatory ability of feature subsets is not known *a priori*. Concentration of discrimination is the driving force behind feature reduction techniques, especially in bioinformatics where less than one percent of sample features can successfully discriminate high dimensionality data [Pizzi5] [Pizzi6] [Pizzi7]. Embedding feature relationships in FCM can be used to consider features with respect to their

acquisition environment, preserve associations between features and distance concepts, and introduce non-linearities that aid discrimination. A formalism to encode feature relationships was developed to augment FCM with specific attention to fMRI analysis. Background theory for this chapter may be obtained from literature on time series: [Box] [Chat] [Digg] [Geye] [Rein]; space-time analysis [Melg]; and fuzzy concepts relating to spatio-temporal models [Cous] [Jack2] [Lowe] [Royc] [Sinc] [Torr].

Feature Partitions

FCM with feature partitions (FCMP) [Alex5] is a generalization of FCM that describes feature relationships and integrates contributions from distinct groups of feature. Due to the general nature of FCMP, classical and more recent fuzzy clustering variations [Pedr1] may be expressed as specializations of FCMP. Like FCM, FCMP is an iterative clustering algorithm that optimizes an objective function. Its novelty lies in the feature relations that are integrated (defined, ranked and combined) into the objective function. A feature partition is a formal mechanism that expresses relations between a single set of features and between sets of features by means of a triple, $\psi=\{ \mu, \nu, \rho \}$, consisting of a metric μ, a relevance weight ν, and a set of feature indices ρ. Particular features included in the subset (group) are encapsulated in ρ; μ defines distance between samples having only features in ρ, and ν is the weight, or relevance, of the distance μ

to the objective function. One benefit of using feature partitions is the ease at which existing analysis can be extended by an explicit feature relation and the resulting clarity of the effect of the relationship on the analysis.

4.1 Generalizing FCM

Pattern recognition literature abounds with FCM variations [Höpp][10] due to the versatility and robustness of FCM. However, individual variations may be understood as specializations of a more general clustering algorithm. Variations can then be considered as specialized instances of the general algorithm where algorithmic degrees of freedom have been constrained or fixed. Clustering notation is introduced which formalizes the generalized clustering algorithm FCMP. Dataset X consists of n f-dimensional samples, $x \in X$, $x=[x_1 x_2 \ldots x_f]$. Denote by V the set of C f-dimensional centroids that define cluster centres, which in turn partition dataset X. Thus, $v \in V$, $v=[v_1 v_2 \ldots v_f]$. Specific samples are referred to as x_i, $1 \le i \le n$; specific centroids are referred to as v_k, $1 \le k \le C$. U is a membership matrix, or cluster partition, detailing the (partial) membership of samples in

10 Höppner provides a formalized notation for cluster analysis. For data space S, $S \neq \varnothing$, and results space P, $|P| \geq 2$, A(S,P) is the analysis space defined as the collection of mappings U from a specific dataset X to a possible solution Y. That is, $A(S,P):=\{f \mid U:X \rightarrow Y, X \subset S, X \neq \varnothing, Y \in P\}$. Analysis spaces are evaluated by an objective function J, $J:A(S,P) \rightarrow \mathbb{R}$ and partial derivatives of J define update equations which converge to an optima. Both FCM and FCMP provide mappings from a data space to an analysis space, resulting in the mapping $U:X \rightarrow Y \in A(S,P)$. Despite the benefit of its formality, Höppner's notation conflicts to some extent with other clustering notations and is not used in this thesis.

the C clusters. Sample memberships are nonnegative and the memberships of a sample considered in all the clusters sums to unity,

$$(21) \qquad \forall x \in X \sum_{v \in V} u_{xv} = 1 \quad , \quad \forall v \in V \sum_{x \in X} u_{xv} > 0$$

Noting the above constraints, the FCM objective function minimizes the weighted sample-centroid distances

$$(22) \qquad J_{FCM} = J(U, X, V) = \sum_{x \in X} \sum_{v \in V} u_{xv}^{m} d^2(x, v)$$

where $d: X \times V \rightarrow \mathbb{R}$ is a distance metric and m is the fuzzy exponent defining membership function shape. FCMP extends this foundation using feature partitions as follows. Define a set of h, $1 \leq h \leq f$, feature partitions, $\Psi = \{ \psi_1, \psi_2 \ldots \psi_h \}$ where each feature partition $\psi_i = \{ \mu_i, \nu_i, \rho_i \}$ consists of a metric μ_i, weight ν_i, and a set of feature indices ρ_i. Alternately, feature partitions can be considered as sets of metrics, weights and feature indices as $\Psi^* = \{M, N, P\}$ where $M = \{ \mu_1, \mu_2 \ldots \mu_h \}$, $N = \{ \nu_1, \nu_2 \ldots \nu_h \}$, $P = \{ \rho_1, \rho_2 \ldots \rho_h \}$.

Partition Metrics and Weights

While FCM calculates distance using a single metric and considers all sample features, FCMP uses a generalized distance metric composed of a weighted sum of metrics over distinct feature indices. Each partition $p \in P$ uses a possibly unique distance function d_p.

They are combined in the cluster algorithm using the weighting parameter $\nu=[\nu_1 \ldots \nu_j \ldots \nu_p], \forall \nu_j, 0<\nu_j<1$. This results in a generalized distance

$$D(x,k)=\sum_{p\in P} \nu_p d_p^2(x',v') \tag{23}$$

Partition Feature Indices

Feature indices

$$P=\{p|p\neq\emptyset, p\subset\{1,2,..f\}, \cup p=\{1,2,..f\}\} \tag{24}$$

and a partition is said to be strict when $\forall p,q\in P, p\cap q=\emptyset$.

$$\sum_{p\in P} \nu_p=1, \quad \forall p, \nu_p>0 \tag{25}$$

The membership update equations exchanges its distance metric with a weighted sum of distances on feature partitions

$$u_{xc}=\left(\sum_{v\in V} \frac{\sum_{p\in P} \nu_p d_p^2(x',c')}{\sum_{p\in P} \nu_p d_p^2(x',v')}\right)^{|m-1|} =\left(\sum_{v\in V} \frac{D(x,c)}{D(x,v)}\right)^{|m-1|} \tag{26}$$

The centroid update equation remains unchanged

$$v=\frac{\sum_{x\in X} u_{xv}^m x}{\sum_{x\in X} u_{xk}^m} \tag{27}$$

The FCMP algorithm is the same as the FCM algorithm, given in Chapter 3 except that

the membership update equation first compute a general distance over all the feature partitions, see Table 6.

Table 6. FCMP Algorithm

Given: FCM parameters, feature partitions Ψ.

Initialization: Choose C centroids.

Loop:

for t = 1 to T

Update memberships, using centroids V_{t-1} and

$$u_{ik} = \left(\sum_{kk=1}^{C} \frac{\sum_{j=1}^{q} \nu_j \, d_j^2(x_i, v_k)}{\sum_{j=1}^{q} \nu_j \, d_j^2(x_i, v_{kk})} \right)^{[m-1]}$$

Update centroids, using memberships U_t and

$$v_k = \frac{\sum_{i=1}^{N} u_{ik}^m x_i}{\sum_{i=1}^{N} u_{ik}^m}$$

if (terminating criteria == TRUE)

$U_{final} = U_t$; $V_{final} = V_t$; t = T;

fi

end

Termination: Output final centroids V_{final} and membership values U_{final}.

FCM equations are provided in Chapter 4. The following section discusses FCMP parameters. Specific applications follow in this section on the utility of a general formalization.

FCMP Parameters

In addition to the FCM parameters, FCMP includes a non-empty set of feature partitions. FCMP uses the feature partitions to adapt to specific problems. When a feature partition collection contains only one feature partition it is proper to describe the algorithm as FCM. A feature partition is composed of a triple: a metric $\mu: \mathbb{R}^G \times \mathbb{R}^G \rightarrow \mathbb{R}^+$, a weight $\nu \in \mathbb{R}$, and a set of feature indices P={p}. This work was initiated by an investigation of the proper initialization of FCMP parameters with respect to fMRI data. Initialization concerns the number of feature partitions to use, the assignment of weights to rank partitions, and the assignment of metrics for each partition. Many methods may be used to select feature partitions. They include:

1. Entropy based grouping – Appropriate for partially supervised learning. Choose n best discriminating features. Partitions may be formed using the increasing, discriminatory ability of collections of features.
2. Statistical Heuristics – Best for unsupervised learning. Rank the features by variance or other statistical properties.
3. Local domain – Useful for unsupervised or partially supervised learning. An example application is fMRI analysis. If there are time instances (temporal values), form a partition by selecting activated and unactivated epochs. Allow lag times for the activated epochs and form another subset.

Metrics may be chosen to ameliorate noise detected in the data or by external associations between features and metrics. For example, it is customary to use the Pearson correlation

coefficient for fMRI time series analysis. This completes our review of FCMP parameters. Further examples demonstrate their use in practice.

4.2 FCMP Specialization

Moving from an abstract (general) formulation to a concrete (specific) adaptation is known as specialization and eliminates algorithmic degrees of freedom from the abstract formulation. Specialization constrains the formula in a particular application of the algorithm. Expectations about the generalized performance of specializations should be tempered by the fact that specialization is essentially a dataset-specific process. The casting of the abstract (or meta-) formulation to a particular dataset aims at a local optima. For example, feature partitions determined to be optimal for one dataset have no necessary relation to optimal partitions in a different dataset. Several examples of optimal parameters are examined for specialization instances.

4.2.1 Utility of a General Formalization

The utility of the algebraic expression of FCMP is shown in the following concrete applications. The format of each application is summarized in Table 7. Additional notes and diagrams follow each example. These examples of utility deal with advantages of using FCMP over FCM in problems of robust clustering, preprocessing, and partial supervision.

Table 7. FCMP Utility of Concept Outline

Concept Name	Topic area.
Problem Description	A specific area where FCMP offers an advantage over FCM.
FCMP Specialization	Details on the specialization of FCMP to fit the problem.
Utility of FCMP	Description of advantages.
Example	Concrete application for the specialization of FCMP.

Application: Robust Clustering

Table 8 details the utility of FCMP specialization with respect to a robust clustering application. Robust clustering is designed to reduce the effects of outliers or noise on the centroids. This is done through preprocessing operations such as outlier detection tests. Outliers can then be eliminated from the dataset.

Table 8. Robust Clustering Utility Synopsis.

Concept Name	Robust Clustering
Problem Description	How to adapt FCM to use different, especially robust, metrics ?
FCMP Specialization	Several cluster algorithms using robust metrics are formulated using FCMP. Simple extensions exist using the feature partition triple.
Utility of FCMP	FCMP has a mechanism to substitute arbitrary metrics for different feature subsets.
Example	Several robust clustering algorithms described by Leski are formulated using FCMP.

Norms and metrics can also be used to reduce the impact of outliers. One such metric is the ϵ-tolerant metric, which defines samples within a distance tolerance of ϵ to be equivalent. Leski [Lesk] presents the following approach to robust clustering. Define an ϵ-insensitive metric or norm where

$$|t|_\epsilon = \begin{cases} 0 & \text{if } |t| \le \epsilon \\ |t| - \epsilon & \text{if } |t| > \epsilon \end{cases} \tag{28}$$

Then the cluster objective function is

$$J_\epsilon = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m |x_k - v_i|_\epsilon \tag{29}$$

This applies to each of the robust clustering algorithms [Leski] which follow: ϵ-FCM ,α-FCM, β-FCM. It can also be used with Fuzzy c-median (FCMedian). The general form

of FCMP entails the specialization as follows: 1 feature partition contains all features; a robust metric is used; the feature partition has a weight of 1.

Application: Preprocessing

Table 9 details the utility of FCMP specialization with respect to a preprocessing application.

Table 9. Preprocessing Utility Synopsis

Concept Name	Inclusion of Preprocessing Operations in the Algorithm Formula
Problem Description	Incorporate algebraically common preprocessing strategies with the clustering algorithm.
FCMP Specialization	Feature indices and weights correspond to standard pre-processing operations.
Utility of FCMP	The feature subsets and weighting of FCMP allow many preprocessing steps to be modeled. A more compact notation for preprocessing and clustering algorithms results.
Example	1. FCMP incorporates Principal Component Analysis 2. FCMP incorporates Independent Component Analysis 3. FCMP incorporates Feature Selection

Example 1: FCMP Incorporates Principal Component Analysis

Designing a pattern recognition or EDA system often requires manipulating the raw data before the clustering process begins. PCA is one common transformation and projects the samples onto axes of maximal variance. Preprocessing with PCA is incorporated by FCMP through using the eigen-values λ of the dataset eigen-vectors Λ as partition

weights. In the notation of feature partitions: $\nu=\lambda$. The number of partitions corresponds to the number of principal components used. The number of components used is often determined by

a) a goal of accounting for at least a certain percentage of the total variance.

b) isolating projection axes that discriminate between the samples.

Each feature partition contains all the feature indices and requires that the samples be projected onto the principal components. Of the triple ψ, only the metrics μ are left to define as parameters.

Example 2: FCMP Incorporates Independent Component Analysis

ICA defines independent components and a mixing matrix which describes how the original signals may have been combined. A whitening matrix, which makes the distribution Gaussian, is often used. The mixing matrix can be used to define a partition similar to that in the previous example using PCA. Each feature partition ψ includes all feature indices, $p_i=\{1,...f\}\forall i\}$, and the ICA mixing matrix is used to transform the data from the original feature space to the independent component space. Typically, the number of independent components is much less than the number of features.

Example 3: FCMP Incorporates Feature Selection

Determining which features to use in a partition can be determined using an exhaustive

search, heuristics, or randomized selection. Each method results in a binary value for a feature (0=not selected, 1 = selected). A binary vector defines the selection of features. This vector defines feature indices in the feature partition. Some methods, such as GMDH [Farl], generate new features by including feature products in combination. This new feature is the product of two or more existing features. The partition in both these cases is the feature index vector multiplied element-wise by the feature selection mask.

Application: Partial Supervision

Table 10 details the utility of FCMP specialization with respect to a partial supervision application.

Table 10. Partial Supervision Utility Synopsis

Concept Name	Partially Supervised Clustering
Problem Description	How can knowledge of class labels be exploited by FCMP?
FCMP Specialization	Samples with labels have an increased weight or cause the algorithm to switch from a robust metric to a regular metric.
Utility of FCMP	Changing metrics for samples with trusted labels increases the contribution of these samples to the centroid. Other methods can be used such that samples with labels will receive a higher weighting in the feature partition. Clusters that have samples with a variety of trusted labels can be split into clusters with samples having the same trusted label.
Example	Similar to partially supervised FCM of Pedrycz and Waletzky.

Let $\Omega=\{\omega_1, \omega_2, ...\omega_C\}$ denote the set of C class labels. Let ω_0 denote the class assigned to outliers, classification rejections, unknown and ambiguous samples. Let $\Omega^*=\Omega\cup\omega_0$ be the set of all possible class labels. Let $\omega(x)$ denote the class label associated with sample x. Define a dataset of unlabeled samples $X=\{x \mid (x)=\omega_0\}$ and a dataset of labeled samples $Y=\{x \mid (x) \in \Omega\}$. Consider the following metric μ_{XY}

$$\mu_{XY}(a,b)=\begin{cases}\mu_1(a,b) & \text{if } a\in X, b\in Y\\ \mu_2(a,b) & \text{if } a,b\in X\\ \mu_3(a,b) & \text{if } a,b\in Y\end{cases} \tag{30}$$

Label information is integrated by switching between different metrics. Alternately, consider feature subsets that are discriminatory with respect to the labeled samples. Let

$\Gamma \in \{\gamma_1, \gamma_2, \ldots \gamma_D\}$ be the set of discriminating features. Let p_1 be the feature indices in feature partition ψ_1 that are all discriminatory, $p_1=\{p \mid p \subset \Gamma\}$. Let feature partition ψ_2 contain only indices of non-discriminatory features, $p_2=\{p \mid p \cap \Gamma = \varnothing\}$. Let weights ν_1, ν_2 be associated with ψ_1, ψ_2 and assign more weight to the discriminating features. When this occurs, $\nu_1 > \nu_2$ and knowledge of class labels is effectively exploited[11].

Application: Small Signal

Table 11 details the utility of FCMP specialization with respect to a small signal detection application. The solution to small signal detection is similar to that of exploiting class labels. It is a question of selecting the appropriate metric. A probe is defined as a finely-tuned function which determines the presence or absence of a localized signal. Probe localization may be spatial, temporal, in the frequency domain, a combination of domains (space-time, time-frequency etc) or may be defined by heuristics. Let us consider features acquired over a time interval. For example, a time series of n instances from an fMRI study. Denote activated epochs as E_A and the unactivated epochs as E_U. Define feature partitions ψ_1, ψ_2 with feature indices

11 Note that, depending on the relative number of features in p_1 and p_2, even the case $\nu_1 < \nu$ may show improvement in terms of discrimination. This can occur when there are few discriminating features. In general, the formula is $|p_1|\nu_1 > |p_2|\nu_2$ for two classes.

Table 11. Small Signal Utility Synopsis

Concept Name	Small signal discovery
Problem Description	Detect a small signal only on the activated epochs while also clustering on the total temporal similarity of time courses and spatial proximity.
FCMP Specialization	Focus on transient and minute phenomenon through the use of probes can be implemented in FCMP by designer metrics.
Utility of FCMP	FCMP extends its clustering scope by tuning metric sensitivity.
Example	Feature partitions are: 1. spatial 2. temporal (all epochs) 3. activated epochs. The activated epochs are correlated to the small signal; time courses are measured against the stimulus paradigm taking into account spatial proximity.

$$p_1=\{p|p\in E_A\} \tag{31}$$

$$p_2=\{p|p\notin E_A\} \tag{32}$$

Assign weights such that $|p_1|v_1>|p_2|v_2$ to augment the activated epochs. This method can also be used to discount features whose values have been contaminated. For labels that exhibit uncertainty, this method may also be applied.

Finally, consider a probe to be a thresholded metric or similarity function such that the output is binary signifying presence (1) or absence (0) of a phenomenon. One difficult problem in fMRI analysis is searching for small signals that are not be linearly related to the paradigm. These small signals presumably reside in only a few percent of the time

courses and will not form their own cluster with a general cluster algorithm. Detecting a cluster with a centroid highly correlated to the paradigm is insufficient to also identify time courses containing this small signal. However, a series of probes may be assembled to detect a variety of non-linear small signals. When the signal is detected in the clustering process (say, after a cluster has defined a centroid sufficiently similar to the paradigm), the probes, being based on the centroid, will be able to detect the small signal. Heuristics can be devised to change metrics when the probe indicates that the small signal is present. In this manner, the clustering process of the entire dataset is combined with a search for small related signals.

4.3 fMRI Data Analysis with FCMP

The following applications of FCMP to problems using fMRI data show the versatility of a generalized clustering algorithm and the practicality of various specializations, including that of adding spatial context to fMRI. Consider a partition of the sample features into spatial ($S=\{x,y,z\}$) and temporal features ($T=\{t_1, t_2, \ldots t_n\}$) for n time instances of the data. In our previous notation, the feature partition P is

$$P=\{S,T\}=\{\{x,y,z\},\{t_1,\ldots t_n\}\} \tag{33}$$

and the respective partition weights are $\nu=\{\nu_S, \nu_T\}$. Denote distances (metrics) for the

partitions as d_S and d_T. Substituting these specific values into the expanded formula, the FCMP objective function adapted for fMRI is

$$J = \nu_S \sum_{x \in X} \sum_{v \in V} u_{xk}^m d_S^2(x,k) + \nu_T \sum_{x \in X} \sum_{v \in V} u_{xk}^m d_T^2(x,k) - \lambda \left(\sum_{x \in X} \sum_{v \in V} u_{xk} - 1 \right) \tag{34}$$

For a particular sample x, the objective function is

$$J_x = \nu_S \sum_{v \in V} u_{xk}^m d_S^2(x,k) + \nu_T \sum_{v \in V} u_{xk}^m d_T^2(x,k) - \lambda \left(\sum_{v \in V} u_{xk} - 1 \right) \tag{35}$$

The membership update equation is

$$u_{ik} = \left(\sum_{kk=1}^{C} \frac{\nu_S d_S^2(x_i, v_k) + \nu_T d_T^2(x_i, v_k)}{\nu_S d_S^2(x_i, v_{kk}) + \nu_T d_T^2(x_i, v_{kk})} \right)^{(m-1)} \tag{36}$$

The interpretation of the objective function is this. Minimize the temporal distance of the sample to the time centroid and minimize the spatial distance to the spatial centroid based on the weights ν_S and ν_T. Spatial proximity and temporal similarity are considered at each iteration of the cluster process at specified levels of integration.

Parameters and Typical Use

Consider the remaining parameters of FCMP with spatial context. The euclidean metric is normally used with spatial features. A variety of metrics may be applicable to temporal features. In practice, a distance metric based on the Pearson correlation coefficient is often used for temporal features. Variations on the theme of temporal

distance include considering only activated epochs (epochs where stimulus is being applied), including anticipation and relaxation responses of the subject [this would consider portions of the unactivated epochs both proceeding and following the stimulation epochs], or allocating increased weight to activated epochs.

A fuzzy exponent, m, close to but greater than 1, tends to reduce the effects of noise. This is significant since common SNR values for fMRI studies are 2-5[Jarm]. Cluster algorithms that use cluster merging (eg. EvIdent® [Pizzi3]) often initialize the algorithm with more clusters than the analyst expects in the dataset. This ensures that a better local optimum will be found by the algorithm as the sample space is searched more extensively. Since FCMP extends FCM, default values for FCM parameters are repeated in Table 7.

Table 12. FCM fMRI Default Parameters

m = 1.1 to 1.3	*C = 10 to 40*
maximum iterations = 20 to 40	U =0.0001

Extending Spatial Context in FCMP

Additional information may be implicitly included in the FCMP functional by considering the spatial neighbours of a sample time course. Consider the effect of replacing the sample by a representative time course that shares both spatial proximity

and signal similarity. Let γ denote a mapping from a given sample index i to the index of a neighbour j which best represents the neighbourhood around sample x_i. For each sample x_i, form its set of neighbours, where the neighbourhood is defined in terms of both spatial and temporal similarity. The functional now has the form

$$J_\gamma(U,V;X)=\sum_{v\in V}\sum_{x\in X} u^m_{\gamma(x)v} D^2_{\gamma(x)v} \tag{37}$$

where $\gamma(x)$ is the neighbourhood (spatial) representative for sample x_i. Note that there are two additional degrees of freedom in this modification, namely:

1. the definition of a neighbourhood for a sample time course
2. the selection of a representative time course from the neighbourhood.

This alternate formulation appends a term to the FCMP objective function

$$J_{m,\alpha,F}(U,V;X)=\sum_{k=1}^{K}\sum_{n=1}^{N} u^m_{nk} D^2_{nk}+\alpha\sum_{n=1}^{N} F(x_n, N_e(n)) \tag{38}$$

where $F(x_n,N_e(n))$ is a function of sample x_n and its neighbourhood $N_e(n)$.

One possible implementation of F is a weighted sum of the neighbourhood scatter matrix:

$$F=\beta_i \sum_{j=1}^{|N_e(i)|} (x_i-x_j)^T(x_i-x_j)=S_c(N_e(i)) \tag{39}$$

Such an objective function is said to contain spatial and temporal terms.

Incumbent Duties in Algorithm Development

When an algorithm is developed, or extended, the content and context it contributes to data analysis must be examined, especially with respect to its implementation and relevance. Some general content and context questions are posed with notes pertaining to FCMP.

1. Is the algorithm restricted to specialized datasets? [General matrices.]
2. What is distinct/novel in the algorithm? [Encoding feature relationships.]
3. How can the algorithm be extended? [E.g. Dynamic partition weights, ν (t).]
4. What implementation constraints exist? [FCMP is iterative; random initialization.]
5. Can the algorithm be parallelized? [Yes, the distance calculations.]
6. In what sense is the algorithm optimal? [Results is intrinsic structure guided by feature relations.]
7. Can critical parameters be determined a priori? [Choice of fuzzy exponent, m, can be guided by FCM studies on similar datasets.]
8. What parameter space region is critical for a robust analysis? [Examine each feature partitions in isolation; combine partitions allowing different partitions to dominate the distance metric.]
9. What data-models informed parameter selection? [Use of Pearson Correlation in fMRI guided choice of spatial metric.]
10. Does the use of this algorithm impact experiment design? [No.]
11. What constitutes a valid benchmark for this algorithm? [Any algorithm that partitions the dataset with or without feature relationships.]
12. Are there optimal parameters that are generalizable? [Unknown.]
13. In what sense is the algorithm, or subsequent analysis, sensitive or robust?

[Random initialization, poor parameter space sampling; minimal data model.]

14. How can anomalous or spurious results be detected? [Examine validation indices over multiple runs.]

From an engineering standpoint, developing or extending an algorithm is driven by a concrete application of the algorithm. FCMP was developed for the spatio-temporal analysis of fMRI data. Chapter 5 describes the fMRI datasets examined.

5. Magnetic Resonance Imaging

Nothing exists except atoms and empty space;
everything else is opinion.
Diogenes Laertius

Magnetic resonance imaging (MRI) [Brow] [Kupe] is a non-invasive imaging modality. The ability of MRI to contrast various soft tissues has led to new imaging applications of the brain, abdominal organs, the musculo-skeletal system, breast, heart and blood vessels. It is the *de facto* standard for biomedical imaging. Different echo sequences, magnetic strengths, goal-specific techniques (e.g. contrast agents) and coils provide high resolution spatial images. Functional magnetic resonance imaging (fMRI) examines blood flow intensity changes produced by a structured stimulus. A stimulus may be any physical change produced near, on, or in the subject. Common stimuli include cognitive, visual, tactile, or auditory impressions. The stimulus is applied over an interval of time (the activated epoch), and then the subject is allowed to rest (the unactivated epoch). Intensity values are recorded continuously over the alternating epoch pairs. Typically active and inactive epochs contain multiple intensity records. MRI and fMRI studies generate voluminous amounts of data at each acquisition. A common dataset size is 60-100 megabytes. MRI and fMRI analysis tests for order and relation in the presence of multiple noise sources with the cognizance that novel information underlies expected

patterns. Since its commercial deployment, MRI and fMRI has become methodologies of choice for investigating the structure of the human body and its functional behaviour respectively. In particular, neural activation studies present challenges for both standard pattern recognition and data analysis techniques. This chapter examines fMRI theory and describes the datasets used in the experiments.

Nuclear Magnetic Resonance

The phenomenon of nuclear magnetic resonance (NMR) involves the interaction of static and oscillating magnetic fields. For a volume of tissue outside of a magnetic field, the spin of constituent protons are randomly distributed and yield a near null net field. Within a static magnetic field, B_0, the protons precess around B_0. Perpendicular to the field, the spin orientations remain randomly distributed. Parallel to B_0, the coupling of the static field and the spin orientations produces the so-called Zeeman interaction, an energy difference between parallel and anti-parallel spins. The lower energy orientation, parallel, has the larger proton population and is characterized as a Boltzmann distribution. Equilibrium between the parallel and anti-parallel spins is known as the induced magnetization M_0. Irradiation of an object in a static magnetic field by an magnetic field, oscillating at precession frequency, rotates the magnetization of the object

into the transverse field. Magnetic resonance occurs between an external magnetic field and nuclei with non-zero magnetic moments or non-zero spin[12]. Hydrogen (H)[13] has spin 1/2. Hydrogen has a high gyro-magnetic ratio γ and is thus sensitive to magnetic fields. The presence of hydrogen in both blood[14] and fat makes imaging of *in vivo* tissues viable.

5.1 Image Acquisition and Noise

MRI data acquisition occurs while the subject lies in a static magnetic field generated by a superconducting electromagnet. Currently, clinical magnetic field strengths range from 1 to 5 Tesla. The application of short radio-frequency (rf) pulses to the magnetized tissue causes the tissue to absorb and subsequently re-emit the energy. Note that the frequency of the rf pulse is matched to the energy difference between the spin up, parallel, and spin down, anti-parallel, orientations. Energy absorption by the nuclei at the resonant frequency is resonance absorption. The time between energy absorption and re-emission is known as the relaxation time. Two main relaxation times are used to define imaging intensities: *T1* and *T2*. The *T1*, or spin-lattice relaxation, time is the time necessary for the z component of M_0 to return to 63% of its original value following an rf pulse. *T1*

12 A non-zero magnetic moment is equivalent to spin. All nuclei having an odd atomic number or odd atomic weight will have spin.

13 The most abundant isotope of hydrogen, ^{1}H, accounts for 99.985% of all hydrogen.

14 Blood is composed of a liquid called plasma and suspended cells such as red and white blood cells (erythrocytes and leukocytes) and platelets. Approximately 90% of plasma is water [Brow].

measures the rate at which spins return to their original configuration. The $T2$, or spin-spin relaxation, time is the time required for the transverse component of M_0 to decay to 37% of its initial value via irreversible processes. The TC activation levels acquired relate the intensities of the deoxygenated blood to the stimulated activity. Figure 16 shows the mean intensity acquired from the s05 dataset. Regions of high average intensity appear lighter in the image.

MRI Noise

There are many compromising factors in MR imaging:

1. Echo sequences, the patterns of rf pulses applied to the subject, have associated noises.
2. Motion artifacts are introduced through various sources. Aperiodic motion blurs image regions containing the moving tissue. Peristaltic motion adds noise to otherwise stationary tissue. Periodic motion generates ghost images. Flow artifacts contort the image depending on their flow velocity.
3. Misregistration, or pixel misalignment, occurs as equipment bias drifts.
4. Radio-frequency transmitters or coils can introduce spurious signals. Magnetic field inhomogeneities reflect, scatter and amplify spurious and legitimate signals.

Reducing the Impact of Noise

For noise localized in the frequency domain, standard bandpass filters may be used. Wavelets are also an effective signal representation for amelioration.

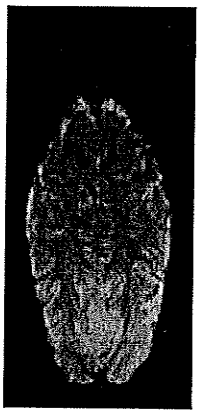


Figure 16. Mean intensity image of dataset s05.

Example of single slice fMRI acquisition.

With respect to fMRI, a noise TC $\xi(t)$ can be identified when a time-shifted copy of itself, $\xi(t+T)$, is not significantly correlated to the original TC. Therefore, a correlation threshold ρ_0 is used to eliminate noise TC. If $\rho(\xi(t), \xi(t+T)) < \rho_0$, $\xi(t)$ is purged from the dataset. Consideration of a desired specificity and sensitivity determine the value of ρ_0. Another means of identifying noisy TC assumes that the TC has an approximately uniform spectral density. A statistic SP is defined as the power in the spectral peak divided by the average power, which is a scaled Fisher's g statistic. A threshold, SP_0, is defined to determine which TCs are analyzed.

$$\Pr(SP > SP_0) = 1 - [1 - e^{-SP_0}]^m \tag{40}$$

Finally, noise TC can be identified by recourse to an expert who authoritatively defines an activation paradigm. An ideal boxcar train demarcates time intervals where the

subject is undergoing response to stimuli or is resting. Thresholding correlation coefficients for TC based on the paradigm eliminates irrelevant TC from further consideration.

5.2 MRI Analysis

MRI analysis is a complex and developing field. Some core concerns of analysts are now listed. Literature on neuron activation analysis has been specifically consulted. The following are references of MRI studies: [Bitt] [Boak] [Bran] [Brie1] [Brie2] [Chen] [Clare] [Cox2] [Fort] [Hark] [Laza] [Jack1] [Miln] [Mega1] [Mega2] [Mega3] [Sava] [Shen] [Wall] [Wein]. A discussion of MRI analysis would be remiss if it did not mention, however briefly, the biological background and several important developing fields and concepts. Such topics as brain atlases, hemodynamic responses models, and neural tissues are now briefly mentioned.

Viability of a General Brain Atlas

Human brains differ in size and neural composition, for example, grey and white matter, change over time for individuals due to illness and disease. General brain models of tissue types and functional inter-relations are being developed but are still at an early stage and tend to be patient specific. Some, such as the Talairach-Tournoux system,

utilize stereotaxic coordinates, which do not necessarily refer to a specific sulcal location. Thus, an atlas for use with neural activation studies is not always available or advisable [Cox1]. That said, the mapping of neural functions and understanding of sulci and gyri are important auxiliary concepts with which to analyze fMRI data. As an example of their future use, enhancement to region growing algorithms could be realized by considering tissue-type directed growth vectors.

Tissue Topology

Another method to reduce noise considers the topological nature of biological tissues. It assumes that no more than two types of tissue types overlap. This gives rise to 7 main types of tissue present in a scan [Drebin] as shown in Table 13. The rf absorptive and reflective properties of the various tissues and their interfaces can augment a HR model.

Table 13. Voxel Classes

Homogeneous Voxel Content	Heterogeneous Voxel Content
(1) air	(2) air/fat
(3) fat	(4) fat/tissue
(5) tissue	(6) tissue/bone
(7) bone	

Robust Statistics

Finally, robust statistics are used to insure analytic inferences from noise. For example,

the Pearson correlation coefficient and its associated metrics, which provide implicit signal normalization, are commonly used to measure similarity for neural activations.

Basic Neurology

Current biological taxonomies list 18 classes of brain tissue [Berg]. As a resource for the reader, brain tissues related to neural activation fMRI studies are listed in Table 14.

Table 14. Brain Tissue Classes

Brain Tissue Classes	Function
Brain stem	Motor and sensory pathway.
Cerebellum	Balance, posture, eye movement.
Carpus callosum	Message transmission from one side of the brain to the other.
Frontal lobe	Abstract thought, emotion, attention, partial eye, muscle, smell.
Hypothalamus	Physical reaction, sexual urge, temperature regulation, hormonal process.
Occipital lobe	Reading / vision.
Parietal lobe	Tactile sensation, stereo-gnosis (understanding form through touch); proprioception (acting on internal stimulation); some visual functions.
Spinal cord	Source of sensation and movement.
Temporal lobe	Music, some vision, memory (vision, auditory and other).

Neuron firings are oscillatory with an individual rate of 800 Hz and a group rate at 40 Hz [Wang]. Consideration of local neuron firing conditions may allow highly resolved frequency filtering for future studies.

5.3 fMRI Analysis

fMRI analysis extends MRI analysis, such as consideration of tissue types and noise

reduction, to a temporal series of MR images. One popular fMRI analysis technique is clustering [Barr] [Baum1] [Baum2] [Baum3] [Baum4]. The main purposes of fMRI analysis are [Frist]:

1. Identification of areas reacting to stimulation.
2. Region growing.
3. EDA for novel activations relating to stimulation paradigm.

To these ends, a variety of signal and image processing operations are marshaled. These operations include: trend removal, temporal-based noise reduction, outlier rejection, thresholding, fitting of statistical models, and paradigm matching. These operations result in documentation as regions of interest, histograms, voxel assignment maps, typical noise TC, outlier spatial locations, intensity images, TC clusters, trends, clusters, typical TC and novel activation candidates. A discussion of the main differences between MRI and fMRI analysis is presented and brief discussions on several of the more common operations follows.

Hemodynamic Response Models

Consideration of both non-homogeneous hemodynamic responses (HR) [Ford] and acquisition-specific noise in hemodynamic response models is a formidable task. This difficulty was previously given as justification for the use of EDA techniques, of which

ICA has been used with some success [Duann]. The following is presented as an example of an HR model. The HR function is modeled as *h(t)* where the observed values *x(t),* are given as

$$x(t)=\int_0^{\infty} h(u)s(t-u)du \tag{41}$$

The noise-free response x(t) is the convolution of the hemodynamic response h(t) and an external stimulation s(t) [Wors]. Noise remains to be included in the formula. An example of the considerable effect of noise is shown in terms of the Pearson correlation of activated TCs. Correlation coefficients for TCs with of 3 second scan intervals may only be as high as 0.4 especially in relevant cortical regions. Using a correlation threshold at that level tends to include many outliers and necessitates an augmenting selection method for TC analysis. Filtering is a common method to reduce signal noise. Low pass filtering, removing the low frequency components of a signal, can eliminate interference from magnetic field inhomogeneities and rf coil noise. A high pass filter reduces the impact of slower trends and shifts in bias. Filtering using an autoregressive model is also used [Wors]. Filtering in the wavelet domain [Alexander] also increases the SNR of fMRI signals.

5.3.1 Trend Removal

Causes of trends in time series include instrument drift and patient movement such as respiration and cardiac pulsation. Methods of dealing with drift include:

1. Identifying trendy TC by an EDA technique and removing them from the dataset.
2. Correcting or de-trending the TC using linear or non-linear methods.

Trends are detected by examining the correlation of the TC with a straight line of unity slope [Jarm]. If the correlation magnitude is greater than a threshold, the TC is determined to be trendy. For a straight line *L*, the correlation of the TC to the line is given as $\rho(TC, L)$. When $|\rho| > \rho_0$ the *TC* is rejected. The threshold ρ_0 is determined by specifying a desired false positive rate using the student-t distribution.

$$(42) \qquad \rho_0 = \frac{t_0}{\sqrt{t_0^2 + D}} \quad \text{where} \quad D = \frac{(T-2)}{DRF}$$

and DRF is a dimensionality reducing factor, modifying the effective degrees of freedom.

Regions of Interest

An analyst often restricts her attention to a localized region in the dataset by defining one or more regions of interest (ROI). fMRI allows definition of ROI corresponding to anatomical structures. Alternately, TC that satisfy a heuristic may be examined, e.g. TCs with activations in the 90% percentile of intensity. ROI are generated by thresholding

intensity statistics of the TCs, for example considering only TCs with above average mean intensity. A correlation threshold may be used if a paradigm is associated with the dataset. ROI may also be defined by a region growing algorithm that has been seeded at specific spatial locations. Segmentation methods also may be used to partition the image, for example the watershed algorithm.

5.3.2 Time Course Normalization

The stimuli of fMRI studies may elicit intensity patterns that vary in magnitude and phase depending on the nature of the stimulus, neural functionality and inter-connectedness, tissue types, and signal path. In order to associate temporal response patterns in the dataset that exhibit magnitude and phase variance, normalization methods are employed. That Pearson correlation uses an implicit normalization partially explains its popularity in analysis. A brief note of explicit normalization methods is provided for completeness.

Normalization by Subtraction: Subtract a constant value from the *TC* , for example, the mean or the median. Comparison between *TCs* is then a baseline comparison.

Normalization by Division: Divide by median. Comparison is based on the percentage of intensity values with respect to a baseline.

Composite Normalization: Subtract median (mean) and divide by MAD (standard deviation). This may increase noise (decrease SNR).

Order Normalization: Rank order the *TC*. Each *TC* is replaced by a sequence of

integers which represent the relative magnitude order for the time instance. This process tends to eliminate outliers.

5.4 Datasets

This study of the effectiveness of FCMP on fMRI datasets examines both synthetic and *in vivo* data sets. The structure of the synthetic datasets are discussed first, followed by a discussion on the *in vivo* datasets. Table 15 describes the format used to summarize the dataset.

Table 15. Dataset Description Format

Name	Dataset Name used in thesis.
Number of samples	total: number in class 1 (description) : ... : number in class N (description)
Labels	Class descriptions.
Description	Salient features of data. Importance of study.
Source	How was data generated or acquired?
Comments	Additional information.

As an aid to understanding the datasets involved and the subsequent cluster analysis, basic fMRI statistics are presented for each dataset. Descriptions of the datasets include: mean intensity images, spatial distributions and regions of interest, intensity and correlation histograms, typical time series, comments on noise and outliers, the activation paradigm and the main focus of analysis with respect to the particular dataset.

Synthetic Datasets

The following datasets were synthesized to demonstrate the feasibility of the FCMP algorithm and to quantify its contribution to fMRI analysis.

- **Syn1** demonstrates the spatio-temporal trade off possible with FCMP on a synthetic fMRI dataset. By design, sample spatial coordinates belong to one of two well separated clusters. One spatial cluster has been associated with TCs correlated to the paradigm while the other has been associated with noise TCs.
- **Syn2** demonstrates the ability and distinctiveness of region growing algorithms. Regions of interest with spatially unique descriptors provide a means to evaluate region growth between the algorithms.
- **Syn3** demonstrates clustering in two dimensions which facilitates the direct viewing of the clusters and cluster partitions.
- **Syn4** demonstrates basic fMRI clustering on a small scale. The dataset consists of a small number of noise TCs and TCs correlated to a stimulus paradigm.
- **Syn5** demonstrates a fMRI dataset where a small signal is present on some of the TCs that are correlated to the paradigm but only on the activated epochs.
- **BaumNull** is a null (no activation stimulus or paradigm) fMRI data acquisition where synthetic activated TCs have been added at specific locations. BaumNull is a hybrid dataset (part synthetic, part *in vivo*) and serves as a transition point for analysis of FCMP.

Synthetic datasets were generated by M. Alexiuk. Dr. R. Baumgartner generated the BaumNull dataset. Each synthetic dataset is now discussed.

5.4.1 Syn1 Data

Syn1 is a synthetic fMRI dataset consisting of two spatial regions that are each associated with only one temporal characteristic. One region has only noise TCs. The other region

only has TCs correlated to the activation paradigm. These TCs are denoted as degraded or paradigm-correlated. Multiple variants of this dataset have been generated at a specific *SNR* levels. The noise TCs are uniform random values in [0,1]. The Syn1 dataset facilitates a controlled demonstration of feature partition interaction in terms of signal degradation and spatial proximity since the spatial and temporal feature partitions are independently generated and the spatial-temporal associations are correlated by design. The degraded TCs are composed of a copy of the stimulus paradigm with random additive noise. The spatial data is uniformly generated, with respect to radius and angle, in two circular areas. Spatial centres, or region means, are c_1=(0.1,0.1) and c_2=(0.9,0.9). Both regions have the same radii, r_1=r_2=0.1.

Syn1 consists of 100 noise TCs and 100 degraded TCs. FCMP uses two feature partitions with Syn1, consisting of spatial (x,y) coordinates and temporal intensities. The activation paradigm is defined as [0101010100]. This paradigm is uniformly expanded when the number of intensity values exceeds the number of paradigm epochs. Recall that an epoch is either activated or unactivated and may contain multiple sampling instants. A paradigm TC maps the stimulus to the number of sampling instants. Thus, successive time instances may belong to the same stimulus epoch. The dataset is the concatenation of the spatial data to the temporal data. The degraded TCs are spatially associated with

c_1; noise TCs are spatially associated with c_2. See Fig. 17 For spatial coordinates. Figure 18 displays typical TCs for each spatial centre.

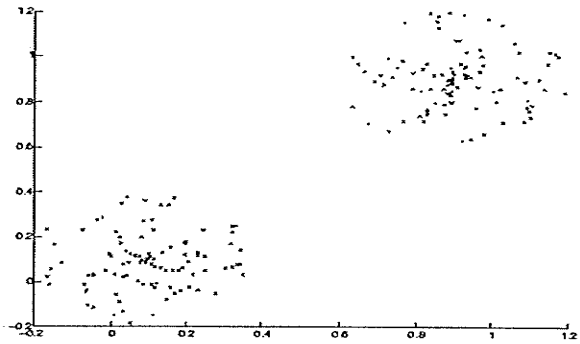
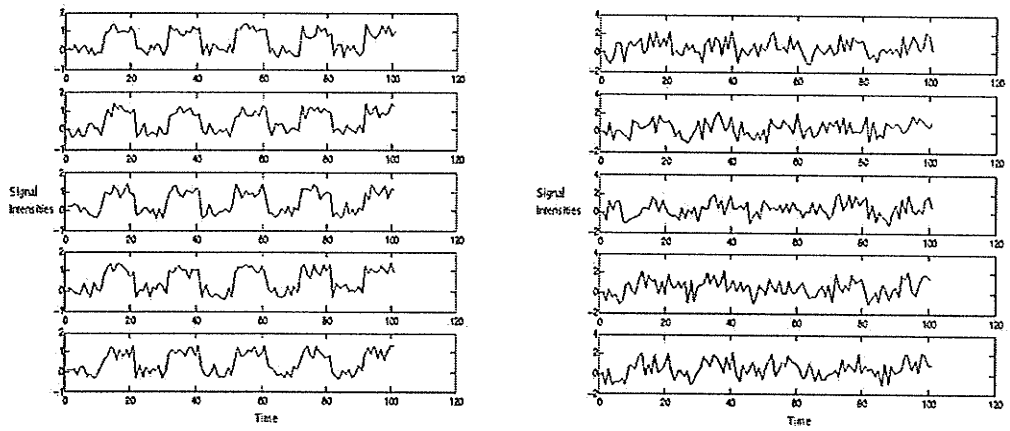


Figure 17. Spatial distribution of Syn1 TCs.



(i) Typical TCs that are correlated to the paradigm at SNR=10dB.

(ii) Typical TCs that are correlated to the paradigm at SNR=1dB.

Figure 18. Typical paradigm correlated TCs in Syn1.

5.4.2 Syn2 Data

The Syn2 dataset is designed to evaluate region growing algorithms in an fMRI context. That is, spatial regions of interest are defined and each spatial point has an associated TC.

TCs are significantly correlated to an activation paradigm, [0101010101], and the TCs are considered as part of the region growing process. The spatial regions used to initialize the region growing algorithms are shown in Fig. 19 and represents a single z- or xy-plane. Methods to generate a region of interest, or interest mask include thresholding of intensity values or correlations in an *in vivo* dataset. The overall (X,Y,Z,T) dimensions of the dataset are (100,100,1,10).

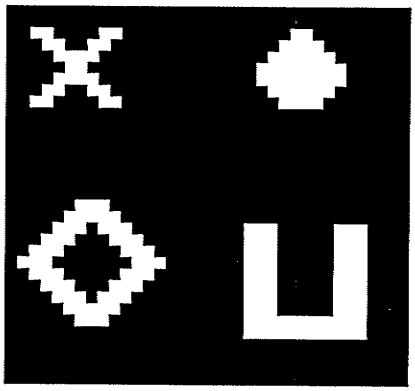


Figure 19. Syn2 ROI.

5.4.3 Syn3 Data

The Syn3 synthetic dataset is used to demonstrate basic clustering principles by using two dimensional samples. Thus, the data distribution in space and the cluster centres are readily viewable and provide low-dimensional justification of such clustering concepts as validation indices, objective functions, and distance measures. The data consists of xy coordinates of samples belonging to two circular distributions, see Fig. 20. The two

circular distributions are well separated in space, meaning that the distance between any two samples in a single distribution is smaller than the distance between any two samples in both distributions.

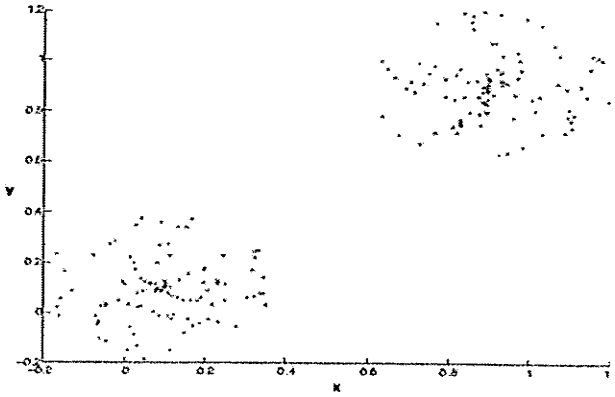
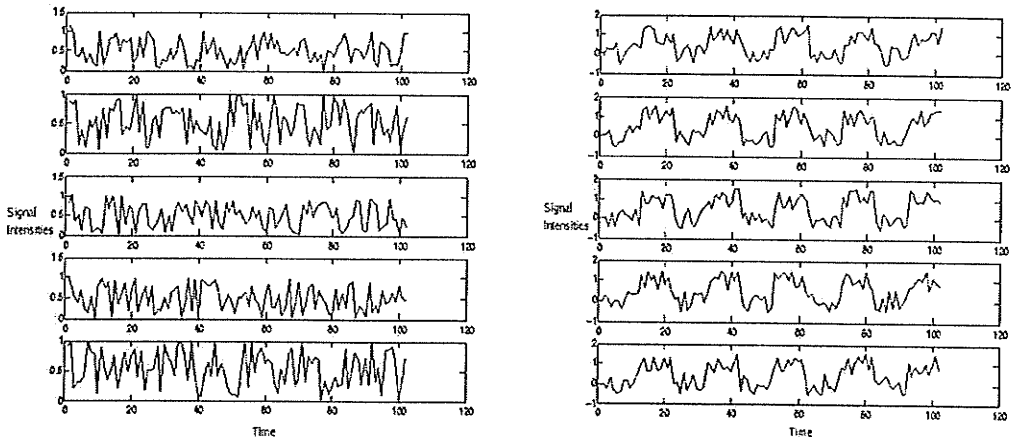


Figure 20. Spatial distribution of Syn3.

This distribution corresponds to that of the spatial coordinates in Syn1.

5.4.4 Syn4 Data

Syn4 is a synthetic fMRI dataset that has been significantly reduced, in terms of numbers of samples, from the *in vivo* datasets examined in this thesis. Syn4 is used to demonstrate basic clustering concepts, especially induced cluster hierarchy, with TCs representative of the main TCs categories present of *in vivo* datasets. Syn4 consists of 5 pure noise TCs and 5 TCs that are significantly correlated to an activation paradigm, [0101010]. The TCs are shown in Fig. 21, (i) and (ii).



(i) Noise TCs in Syn4. *(ii) Degraded TCs in Syn4.*

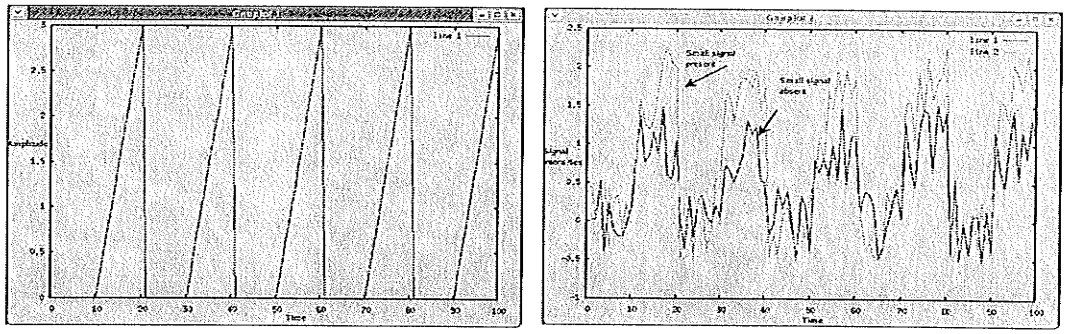
Figure 21. Typical TCs in Syn4.

5.4.5 Syn5 Data

Syn5 is a synthetic fMRI dataset that has many of the same parameters as Syn1. The dataset is composed of fifty noise TCs and fifty TCs that are degraded with respect to the activation paradigm. Each class of TCs is associated with 2-dimensional spatial features located in a circular radii of 0.3 with centres (0.1,0.1) and (0.9,0.9) respectively. The number of temporal features is 100. The SNR level of the degraded TCs is 5 with respect to the activation paradigm [0101010101]. Epoch lengths span 10 sampling instants.

The difference from the Syn1 dataset is that 10 degraded TCs have been modified to include a small signal only on the activated epochs. Ten such TCs were selected with random spatial coordinates and have had a ramp function added to their values, seen in Fig.22 (i). A comparison between a degraded TCs with and without the ramp function is

seen in Fig.22(ii).



(i) Ramp small signal added to activated epochs of degraded TCs.

(ii) Comparison of degraded TCs with and without the presence of the small signal.

Figure 22. Small signals in Syn5

5.4.6 BaumNull Data

Dr. R. Baumgartner, of IBD, et al. synthesized the BaumNull dataset to demonstrate the high type1 (false positives) error rate associated with simple correlation analysis of fMRI data [Baum4]. Two groups of TCs simulate the extreme case when TCs correlate highly to an activation paradigm but when the groups do not correlate at all to each other. The groups contain 46 and 26 TCs respectively and signal intensities over 120 time instants are used. A simulated hemodynamic response used two parameter gamma functions to generate responses to the activations in the respective TC groups. Thus, this dataset may be considered to have two activation paradigms.

A mean intensity image for BaumNull is shown in Fig. 23. Figures 24 and 25 show the mean intensity image after it has been thresholded at fifty and sixty percent of the

maximum value, respectively. The elongated and convoluted regions of interest generated by this thresholding are common in fMRI studies and correspond to the anatomical sulci structure. When neural regions are stimulated, the maxim that spatially proximal regions are temporally correlated must be modified to consider proximity in terms of anatomical structure, and not simply Euclidean distance.

Figure 23. BaumNull coronal plane, z = 0.

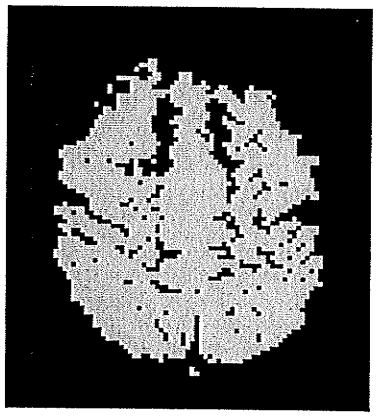


Figure 24. BaumNull thresholded intensity image. Threshold of fifty percent maximum intensity.

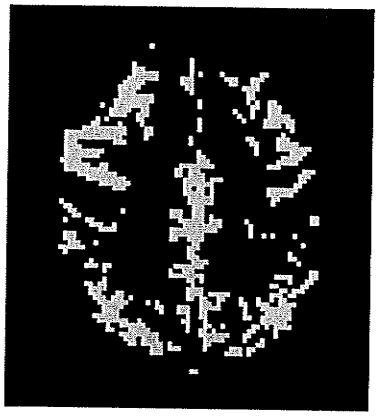


Figure 25. BaumNull thresholded intensity image.

Threshold is set at sixty percent of the maximum intensity.

Figure 26 shows the distribution of correlations throughout the BaumNull plane, z=0. It is apparent that large numbers of noise TC, those outside of regions of interest and even the entire brain, are significantly correlated to the activation paradigm.

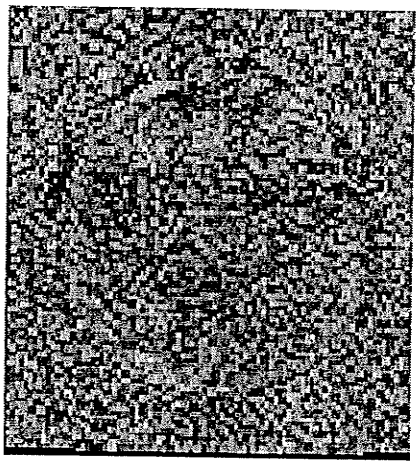


Figure 26. BaumNull correlation plane, z = 0.

However, the distinguishing features of these voxels is the small number of their

neighbours that also share significant correlations to the paradigm. A correlation histogram, shown in Fig. 27, shows the positive values of correlation. Few of the TCs in the dataset have correlation values that would be considered significant in fields outside of fMRI analysis, which would lie in the range [0.4, 0.8].

Basic statistics for BaumNull include: data value range: [0,1000], a mean value of 135.603, variance of 51305.7. The first stimulus paradigm is of length 12 with pattern [000001000000]. The second stimulus paradigm is of length 12 with pattern [000100001000]. The (X,Y,Z,T) dimensions of the dataset are (128, 128, 1, 120) with a corresponding number of elements, 1966080.

Noise TCs lie outside of the two artificial activated regions of interest. Outliers may be considered to be TCs outside of the ROI but that are somewhat correlated to one of the two activation paradigms. The focus of analysis for this dataset will be the interaction of distinct activation paradigms, not correlated to each other, in a synthetic, but typical, fMRI dataset.

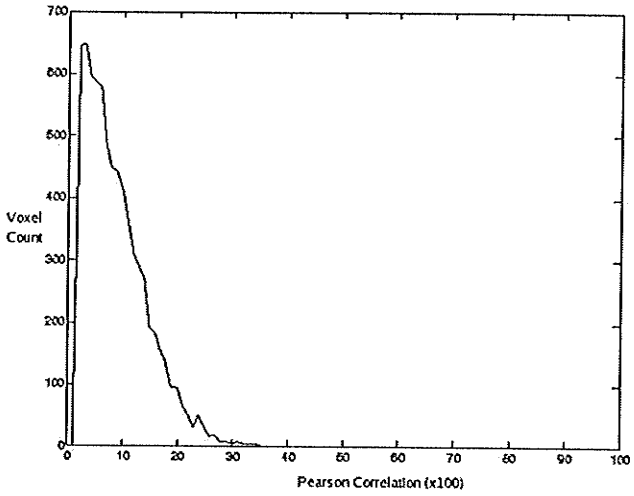


Figure 27. Correlation histogram for BaumNull, z= 0.

5.5 In vivo Datasets

Neural activation data has been acquired from various studies performed at IBD, as well as at external locations. A list of all *in vivo* datasets used in this study follows.

S05 – checkered visual stimulation.
Sample4d – tactile (finger-tapping) stimulation.
Halx – Tourette's syndrome study.

For consistency in inter-data comparisons, all *in vivo* datasets have been scaled to [0,1000]. Each *in vivo* dataset is discussed in detail in the following pages.

5.5.1 S05 Data

S05 is an *in vivo* neural activation study acquired at IBD. The activation paradigm is based on a checkered visual stimulation presented to the subject. Neuron activations for a single z-plane or slice were recorded at 42 distinct time instants. The dataset is

composed of TCs with (X,Y,Z,T) dimension of (128, 256, 1, 42). A mean intensity image is shown in Fig. 28. The visual cortex, lower centre, and a lobe in the right hand lower side, show high levels of signal intensity.

Figure 28. Mean intensity image for S05.

A correlation histogram for S05, Fig. 29, shows significant correlation values with respect to the activation paradigm, while a correlation image, Fig. 30, highlights correlation in the visual cortex and a proximal lobe. The correlation plane, Fig. 31, shows the spatial locations of high correlation, indicated by lighter values, which significantly includes the visual cortex and the aforementioned lobe.

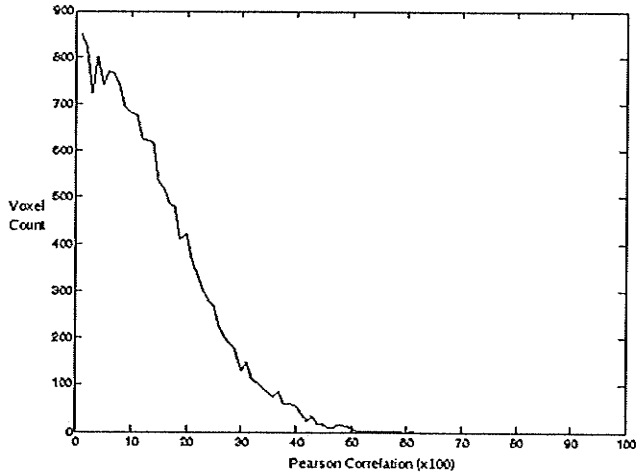


Figure 29. Correlation histogram for S05.

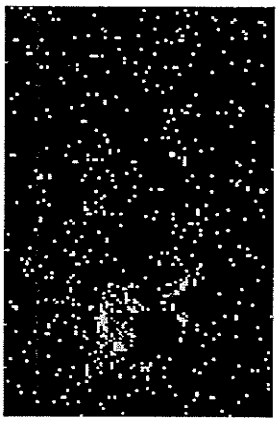


Figure 30. Correlation locations in S05.

Locations with Pearson Correlation greater than 0.3 are shown.

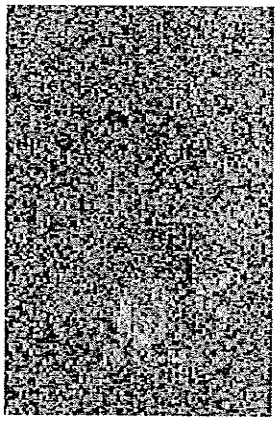


Figure 31. Correlation Plane for S05.

Thresholding the correlation plane, Figs. 32 and 33, generates regions of interest, spatial regions where all voxels exhibit a significant level of correlation to the activation paradigm. Typical TCs with above average intensity values for S05 are seen in Fig. 34. There is obviously high levels of noise present.

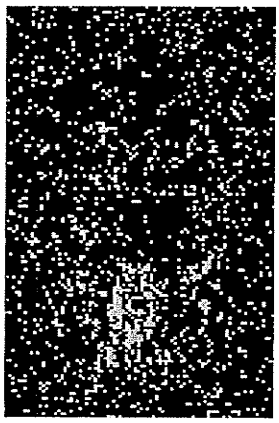


Figure 32. Thresholded correlation plane S05. Pearson Correlation threshold is set at T_{Φ} =0.2.

Figure 33. Intensity thresholded ROI of S05. Threshold is sixty percent of the maximum intensity.

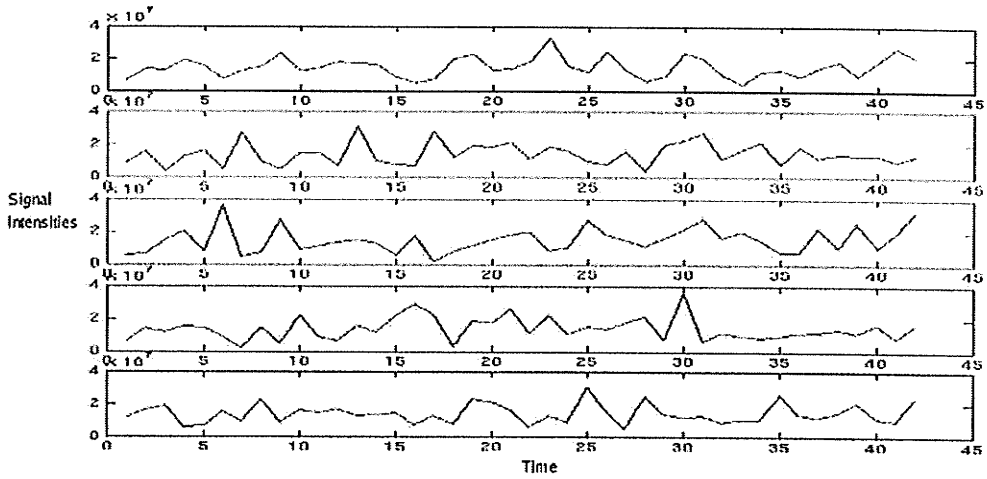


Figure 34. Typical TCs in S05 with greater than average intensity.

Figure 35, which displays typical TCs with above average correlation values for S05, also indicates the presence of high noise levels. Of interest in some fMRI analysis is the variability, or intensity variance, of blood flow as indicated by intensity values, see Fig. 36. This variance may correspond to the repeating pattern of the stimulation. Changes in variance between activated and unactivated epoch pairs can indicate learning mechanism

in the neural activations, such as anticipation.

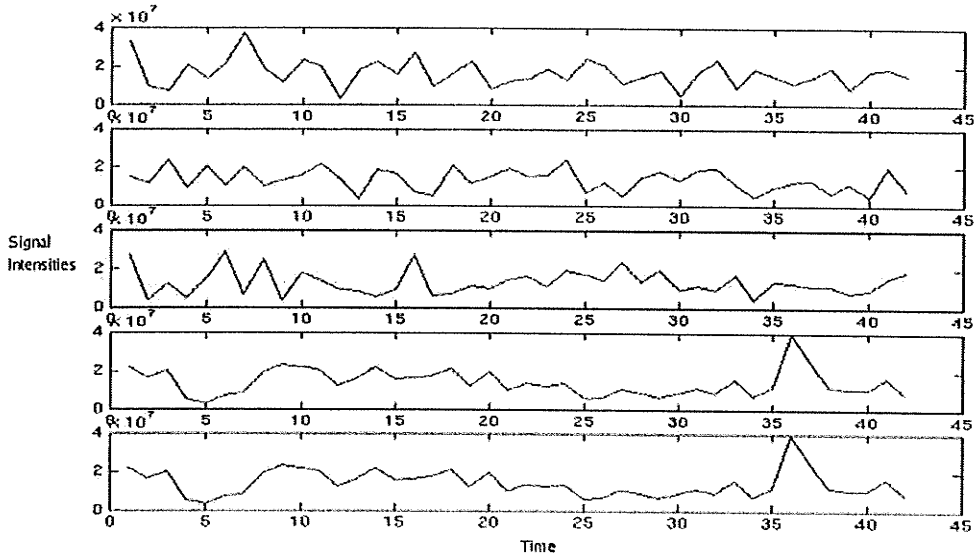


Figure 35. Typical TCs in S05 with above average correlation to paradigm.

Figure 36. Variance of intensity for S05.

The (X,Y,Z,T) dimensionality of the dataset is (128, 256, 1, 42) and the number of elements is 1376256. The activation paradigm is 010110. When considering noise and outliers in the S05 dataset, note that significantly correlated values are found outside of

the visual cortex. The focus of analysis for S05 is the visual cortex and the adjacent lobe.

5.5.2 Sample4d Data

Sample4d is a neural activation study based on a motor stimulation, finger tapping. It was acquired at IBD and is different from S05 in that sample4d contains multiple Z slices. A mean intensity image, for the xy-plane z=0, is given in Fig. 37. Only the epochs corresponding to active epochs were considered for this rendering. Regions of interest due to intensity variance are shown as light patches in Fig. 38. A correlation image, Fig. 39, of sample4d indicates one large ROI at location (60,200). A correlation histogram for sample4d, Fig. 40, is comparable to the correlation histogram for S05. Basic statistics for the intensity values include a range of [0,1000], a mean value of 56.0085, variance of 14327.4. The stimulus paradigm is [0101010101010]. The (X,Y,Z,T) data dimension is (128, 128, 4, 110) and the number of elements is 7208960. Noise and outliers in the dataset should be considered as in S05, namely, those samples with significant mean intensity, high intensity variance, or high correlation values that are spatially isolated.

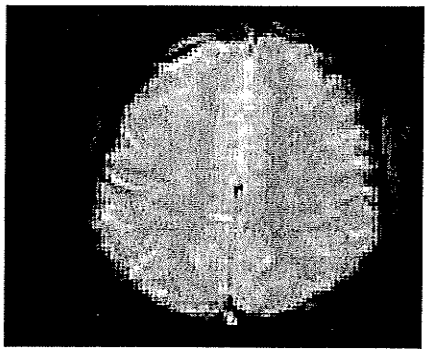


Figure 37. Mean intensity for active epochs for sample4d, z = 0.

Figure 38. Sample4d intensity variance image, z=4.

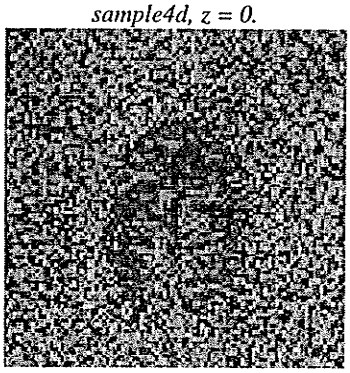


Figure 39. Sample4d correlation plane.

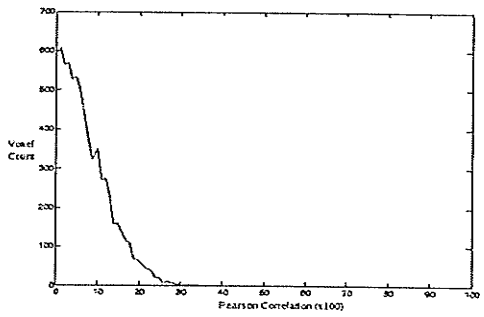


Figure 40. Sample4d correlation histogram.

5.5.3 Halx Data

Halx is an *in vivo* dataset used at IBD that was acquired in a Tourette's syndrome study. The dataset is composed of fMRI TCs with (X,Y,Z,T) dimension (128, 128, 16, 128). The study poses challenging problems to analysis due to the variable, context-dependent nature in which Tourette's syndrome is manifested, namely tics. Persons with tics exhibit sudden, brief and isolated movements, and experience noises or sensations that are involuntary. The most common cause of tics is Tourette's syndrome, which may also be related to behavioural idiosyncrasies. Tics are variable in intensity, location and in

their susceptibility to voluntary suppression. Tics increase with excitement and decrease with distraction. While most cases of Tourette's syndrome are genetic, it can also be caused by neuroleptics, head trauma, encephalitis and cocaine abuse. Mean intensity images for the Halx dataset are given in Figs. 41, 42 and 43 for xy-planes z=0, z=5 and z=14 respectively. These planes were chosen to sample the different cross-sections acquired and the ROI associated with each plane. Figure 44 shows the intensity variance for the xy-plane z=14. Note the highly localized, apparently symmetric structure of the ROI.

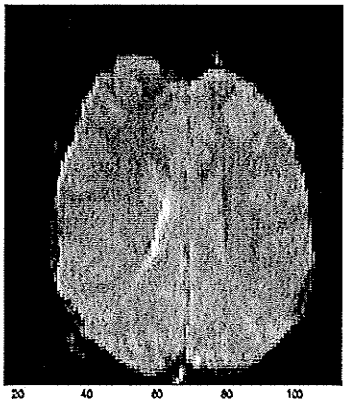


Figure 41. Active mean intensity for Halx, z =0.

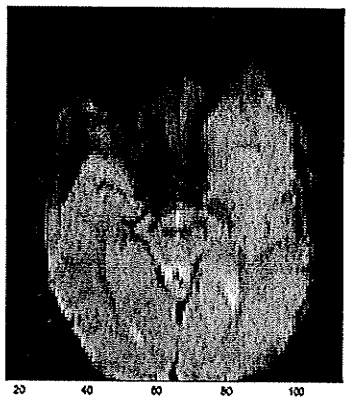


Figure 42. Active mean intensity for Halx, z = 5.

A correlation image for Halx is given in Fig. 45. Of note is the dark patterns apparent around locations (60,20) and (60,80). The correlation histogram, Fig. 46, for Halx shows that the this dataset does not have as many highly correlated TCs as did S05 and

sample4d.

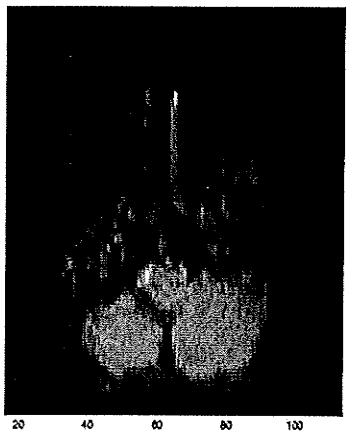


Figure 43. Halx active mean intensity plane, z = 14.

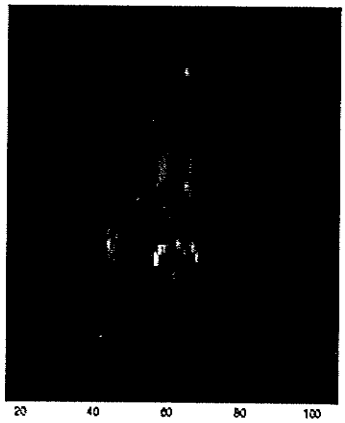


Figure 44. Halx active variance intensity plane, z = 14.

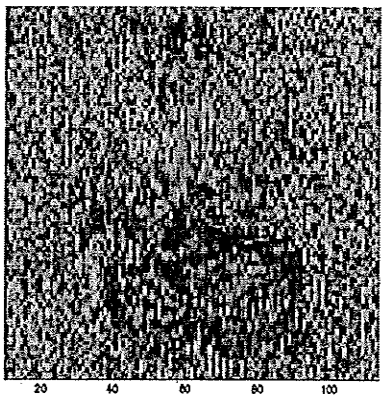


Figure 45. Halx, correlation plane for z = 14.

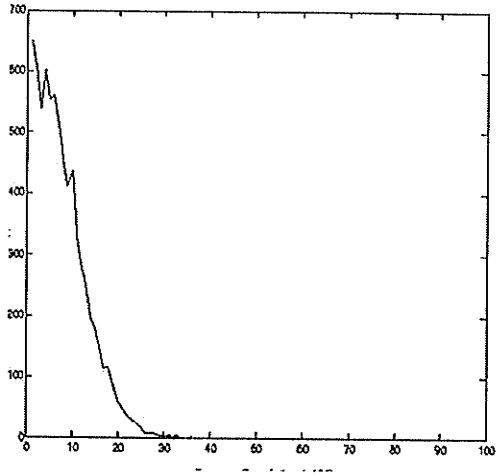


Figure 46. Halx, correlation histogram for z= 14.

Typical TCs in the Halx dataset are seen in Fig. 47, where the TCs have high average intensity, and in Fig. 48, where the TCs have high relatively correlation. As mentioned, the (X,Y,Z,T) data dimension is (128, 256, 16, 110). The paradigm stimulus for this

dataset is [010101010101].

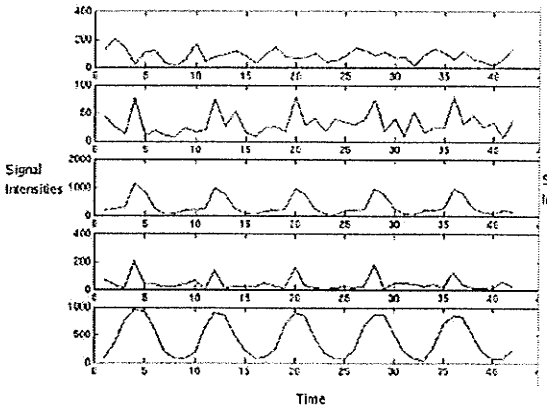


Figure 47. Typical high intensity TCs in Halx.

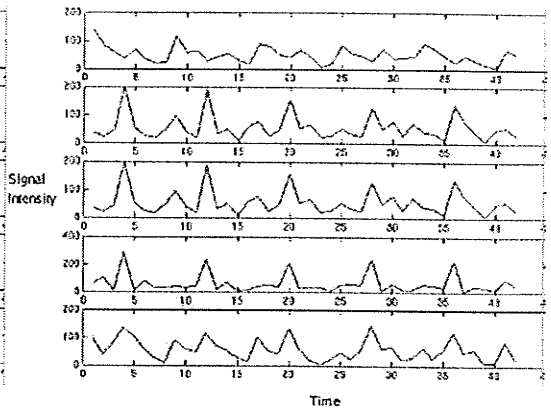


Figure 48. Typical high correlation TCs in Halx.

EDA and fMRI

Neural activation studies examine the spatial regions (volume elements or voxels) associated with paradigm-correlated TCs. Regions of interest are defined in space for time sequences of importance. Such a designation increases the importance of other spatially proximal TCs. While clustering algorithms define centroids, they can also define spatial partitions on the fMRI images, using the cluster membership values as voxel memberships in various partitions.

6. Benchmarks

Test everything.
Hold on to the good.
1 Thessalonians 5:21

Since the scope of interest in fMRI analysis can vary, a wide variety of algorithms have been developed to cover niche applications. The ability of the algorithm to address noise in the dataset and to elicit meaningful relations from the data are fundamental requirements for each niche. When developing an algorithm for a novel application, the existing knowledge base of pattern recognition techniques is examined, as a rule. Pattern recognition experts do so because, as Huber stated, *"there is no panacea in data analysis"*. This battery, or salvo, approach to problem solving proceeds despite the foreknowledge that many algorithms in the knowledge base have limiting weaknesses. However, each algorithm failure adds a point on a receiver operator characteristic (ROC) curve for the problem that focuses both the adaptation of existing algorithms and new development. In order to compose a set of benchmarks for FCMP, a number of basic algorithms, industry standards and algorithms noted in recent literature have been selected, namely:

Basic Benchmarks: Cluster Merging [Romes] and FCM [Bezd1]

Industry Standards: EvIdent® [Jarm]

Academic and Research Literature:

CHAMELEON [Kary] and
Fuzzy Seeded Region Growing (FSRG) [Viva].

Other notable algorithms include: model based fMRI approaches such as: AFNI [Cox2] [Cox3], Statistical Parametric Mapping (SPM) [Fris1], and Medx [Boak]; and model free approaches such as algorithms in the FMRIB Software Library (FSL) [Smith]. Each benchmark algorithm is examined in the following pages. Table 16 details a template which is used to summarize benchmark characteristics.

Table 16. Benchmark Summary Template

Description	A characterization of the algorithm.
Strengths	Particular advantages of the algorithm for a specific task or in general.
Weaknesses	Particular dis-advantages of the algorithm for a specific task or in general.
Goal Orientation	The heuristic or objective function that guides the algorithm to convergence and completion.
Degrees of Freedom	Algorithmic degrees of freedom. A description of the parameter space of the algorithm.

An example use of the algorithm is provided for each benchmark algorithm. Related notes and discussions follow in each section. For algorithms where it is appropriate, a synthetic fMRI dataset, Syn4, is used. It consists of 10 samples: 5 noise, and 5 signals

that are significantly correlated to a boxcar stimulus paradigm. To demonstrate algorithm use, several measures are computed, among them the mean square error (MSE) and Pearson correlation of the centroids to the paradigm. For EvIdent®, results are shown for s05, an *in vivo* dataset examined in Chapter 7. To demonstrate region growing on a small scale, another dataset, Syn2, is used. Syn2 is a synthetic fMRI dataset with (X,Y,Z,T) dimensions of (100,100, 1, 20). In the XY plane, four distinct regions are designated as ROI, being correlated to a stimulus paradigm [0101010101]. Regions represent spatially connected, temporally similar TCs. All other TCs in the dataset, in regions outside those already defined, are noise TCs. FSRG and region growing processes, which utilize FCM and FCMP, are demonstrated on Syn2. The spatial characteristics of the ROIs highlight the abilities of the different region growing processes.

6.1 FCM

FCM, as outlined in Chapter 3, is a robust algorithm well suited to unsupervised learning problems. FCM[15] is used as a benchmark due to its successful history in discovering data structure, its minimal mathematical model, the variety of validation methods for its results, and its ease of adaptability to external constraints such as dependencies in the data. Although non-convex regions pose problems for vanilla FCM, it is a standard

15 When FCM is restricted to the standard spherical clusters, as it is in this thesis, the algorithm is referred to as *vanilla* FCM.

benchmark in pattern recognition. Recall that FCM can be considered a specialization of FCMP where the following algorithmic degrees of freedom are fixed: only one feature partition exists, and the Euclidean metric is used. Table 17 provides details on FCM. Fig. 49 shows the distribution of samples for Syn3. Centroids are generated by FCM for C=2,3,4 and 10 respectively and listed in Table 18. The VCV matrix for the dataset, shown in Fig. 50 , displays two dark regions along the diagonal, indicating two inherent clusters. Validation indices are another means to determine the most representative number of clusters. These indices are shown in Fig. 51 and indicate two clusters.

Table 17. FCM Algorithm Summary

Description	A clustering algorithm which defines C spherical clusters in a dataset and assigns degrees of membership for each sample in every cluster.
Strengths	Always finds C clusters. Iterative convergence often provides a fairly representative partition after a small number of iterations.
Weaknesses	The optimal value of C is unknown *a priori.* Use of validation indices to confirm the optimal value of C are to some extent dataset-dependent. Clusters are always spheres. Generally, only the Euclidean metric is used.
Goal Orientation	Maximizes the inter-cluster variance while minimizing the intra-cluster variance.
Degrees of Freedom	Number of clusters; fuzzy exponent; initialization.

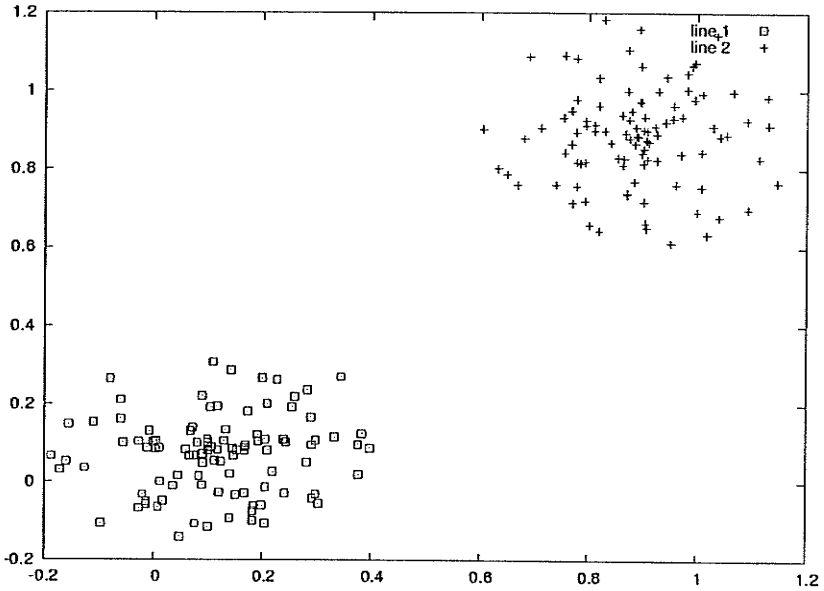


Figure 49. Sample distribution of dataset Syn3.

Table 18. FCM Centroid Locations for Syn3

C	***Centroid Locations***
2	(0.175335 , 0.130495), (0.821875, 0.814863)
3	(0.740489, 0.72867), (0.1523, 0.106273), (0.743041, 0.73138)
4	(0.821892, 0.81488), (0.821854, 0.81484), (0.175258 , 0.130414), (0.175417, 0.13058)
10	(0.780952, 0.771524), (0.782082, 0.772724), (0.157706, 0.111915), (0.157709, 0.111918), (0.781896, 0.772527), (0.157709, 0.111919), (0.782481, 0.773148), (0.782534, 0.773204), (0.782078, 0.77272), (0.157715 0.111925)

Validation indices for clusters generated by FCM with C=2,3...20 are shown in Fig. 51. The indices indicate from 2 to 7 clusters in the dataset, determined by points of diminishing returns for each index over the range of C values.

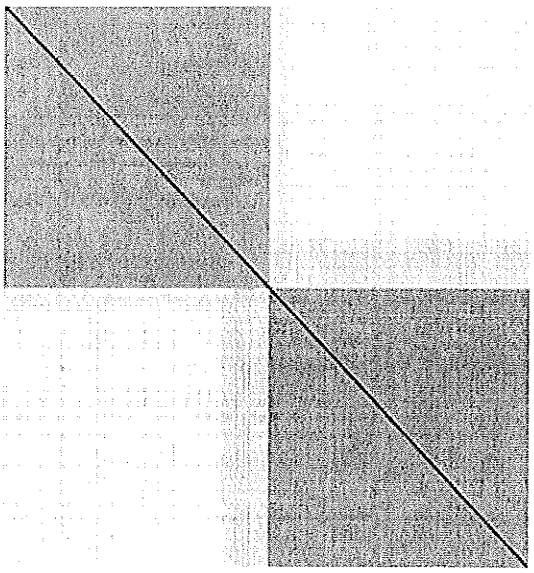


Figure 50. VCV matrix associated for dataset Syn3.

Two intrinsic clusters are indicated.

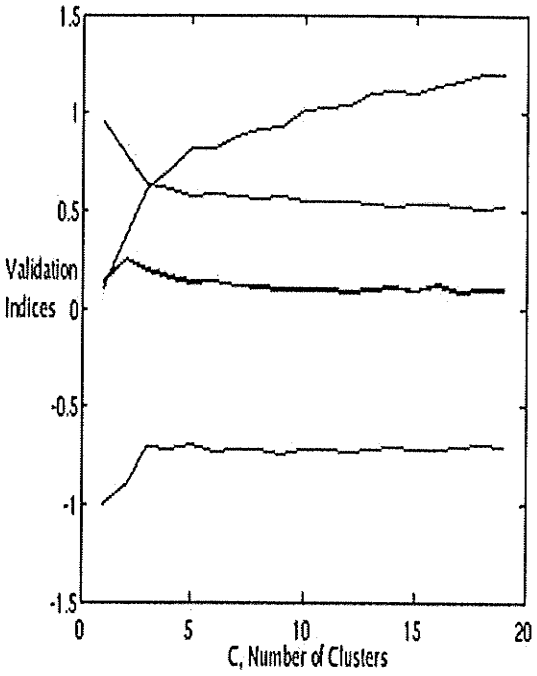


Figure 51. Validation indices for Syn3.

The indices PC, PE, FS and XB have been normalized to [-2,2].

6.2 Basic Cluster Merging

Cluster merging considers the proximity between single samples, a sample and a group, and groups of samples. Samples and groups of samples are merged until the total number of desired clusters is met. This method was not discussed in Chapter 4 on clustering since the algorithm is straightforward. A distance matrix is defined over the sample space. Proximal samples are merged and their centroid is computed which replaces their respective entries, rows and columns, in the original distance matrix. The algorithmic

degrees of freedom of this basic hierarchical method are: choice of metric; centroid definition. Adding new samples to the cluster requires that the cluster centroids be updated. Hard cluster memberships are produced but thresholds are easily set to reveal partitions for an arbitrary number of clusters. Table 19 provides details on the benchmark basic cluster merging.

Table 19. Basic Cluster Merge Algorithm Summary

Description	The dataset is merged sample by sample using a sample-centroid distance. Centroid definitions are updated after each new merger.
Strengths	After merging all samples, the cluster structure for an arbitrary number of clusters, C, is determined simply by appropriate thresholding the induced hierarchy. An added benefit for analysis is that the order in which samples are merged is preserved by the algorithm.
Weaknesses	Hard sample memberships.
Goal Orientation	Merge the sample closest to any centroid with the cluster of that centroid. Update the centroid definition. Continue until all samples are merged or until a specified number of clusters is achieved.
Degrees of Freedom	Centroid definition; metric.

6.3 EvIdent®

EvIdent®, from EVent IDENTification [Jarm], is a current research and industry tool for the EDA of fMRI data. It is introduced in Chapter 7 as a state of the art algorithm for fMRI analysis but is included here in order to consolidate the discussion of benchmark algorithms. EvIdent® is a model-free, 3D, EDA application that locates regions of

activation, and detects artefacts and trends. It was developed at the National Research Council, Institute for Biodiagnostics (NRC-IBD) and has found acceptance in the research community. Numerous publications [Baum2] [Jarm] [Pizzi2] attest to its benefits. The featured analysis tool associated with EvIdent® is known as Exploring Regions of Interest with Cluster Analysis (EROICA) [Jarm]. EROICA is a *monolithic algorithm* which utilizes dataset-specific statistical pre-processing of the TCs. Such pre-processing tasks as rejecting outliers and noisy time courses sometimes eliminates 95% of the original TCs. This data-screening significantly reduces execution time. Fig. 52 shows a screen shot of EvIdent® after a dataset has been clustered with EROICA. Table 20 provides details on the benchmark EvIdent®.

Table 20. EvIdent® Algorithm Summary

Description	FCM clustering is augmented by a complex of statistic-based pre-processing heuristics which reduce noise and speed execution.
Strengths	Trendy samples removed. Fast execution time. Industry acceptance.
Weaknesses	Removes samples. Initial results can have significant false positive rates.
Goal Orientation	Remove trends; remove noise; produce cluster analysis on remaining data.
Degrees of Freedom	Preprocessing parameters. FCM parameters. EROICA parameters.

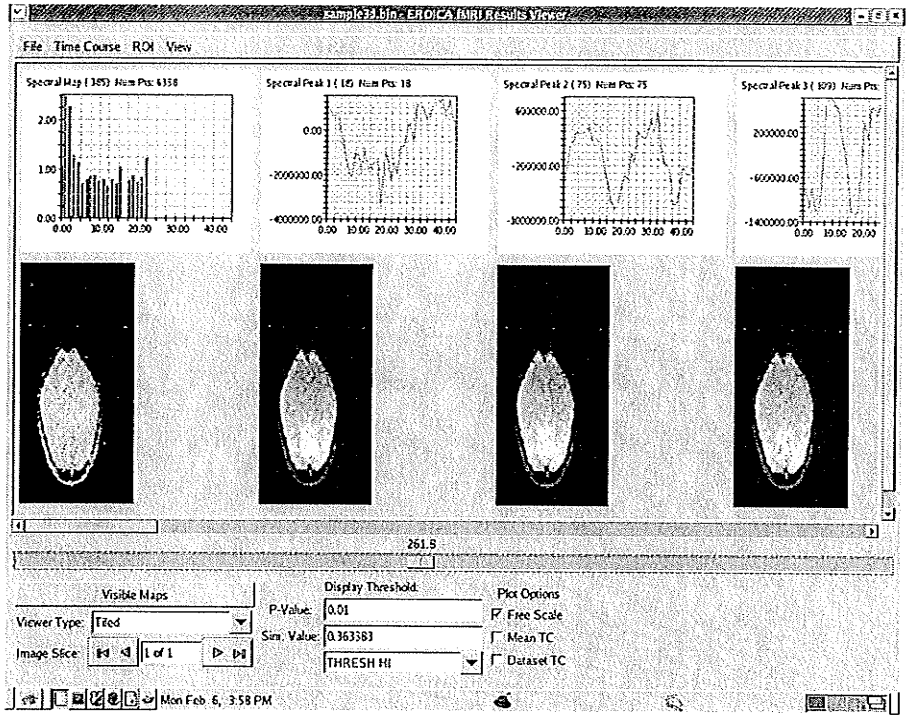


Figure 52. EvIdent® display in Scopira.

Default EROICA settings were used to generate these results from the s05 dataset.

6.4 CHAMELEON

CHAMELEON [Kary] is a hyper-graph partitioning algorithm that uses a dynamic cluster model. Karypis notes that existing hierarchical methods use static models or do not take into account idiosyncrasies of individual clusters. Clusters are idiosyncratic when they have unique shapes and size with respect to other clusters in the dataset representation. He suggests that a dynamic method which considers how distances, and nearest neighbours groupings, change as the algorithm converges, will capture such cluster uniqueness. The CHAMELEON approach consists of iteratively comparing

clusters C_i and C_j in terms of their relative interconnectivity, *RI*, and their relative closeness, *RC*. Table 21 provides details on the benchmark CHAMELEON.

Graph Theory Background

CHAMELEON is designed to partition hyper-graphs. A *hypergraph* is a graph where edges are shared by two or more nodes (vertices); both the edges and vertices may be weighted. The *min-cut bisector* of a graph is the minimum sum of edge-weights cut when the graph is partitioned into two sub-graphs and is used to define *RI* and *RC*.

A graph is defined by how individual *TC*s relate to each other. The CHAMELEON algorithm was designed for use with sparse graphs in fields such as very large scale integrated (VLSI) circuit design and theoretical computer science. These sparse graphs do not arise naturally in fMRI studies but may be constructed artificially as follows. A parameter k is introduced which limits the definition of adjacency for *TC*s. This limit constrains the topology to be sparse.

Table 21. CHAMELEON Algorithm Summary

Description	A heuristic based recursive partitioning algorithm which utilizes a dynamic cluster model.
Strengths	A dynamic cluster model has demonstrated the ability to characterize idiosyncratic cluster structure.
Weaknesses	There are many parameters.
Goal Orientation	Partition the dataset using a recursive splitting and agglomeration of the samples using graph theoretic definitions of interconnectedness and closeness.
Degrees of Freedom	CHAMELEON parameters.

The cliquish distance matrix (which relates each sample to every other) is replaced by a k-nearest neighbour distance matrix which is comparable to those in circuit design. Thus, a comparison between FCMP and CHAMELEON considers the use of ϵ-tolerant metrics, such as

$$\begin{aligned} d_\epsilon(x,y)&=0, \text{if } d(x,y)<\epsilon \\ d_\epsilon(x,y)&=d(x,y), \text{otherwise} \end{aligned} \tag{43}$$

where d is a metric. In this way, the distance matrix commonly used in clustering is replaced by a sparser matrix. The FCMP generalized metric is used to compute a sparse version of the inter-sample distance matrix dm. This matrix is thresholded such that each sample is connected only to its K nearest neighbours, and is denoted dm_{knn}. Thus, the

nearest neighbour ϵ-tolerant metric is

$$d_\epsilon(x,y)=d_{knn}(x,y), \text{if } x \text{ and } y \text{ are nearest neighbours} \quad (44)$$
$$d_\epsilon(x,y)=0, \text{otherwise}$$

CHAMELEON then recursively splits the dataset into sub-clusters. Fragmented sub-clusters are merged based on RC and RI thresholds. Readers interested in details on the precise effects of CHAMELEON parameters are advised to consult [Kary]. The following example shows results of CHAMELEON analysis on the Syn3 dataset. Typical parameters used include: 10 desired clusters, separate threshold consideration in merge mode, closeness threshold 0.4, connectivity threshold 0.4, number of nearest neighbours 3, Euclidean metric, minimum cluster size in splitting phase 5. Due to the heuristics involved in both the CHAMELEON algorithm and the current Scopira interface for fMRI analysis, the desired number of clusters is not always achieved. Eight clusters were defined by the algorithm with the clusters containing [85,10,7,15,19,6,30,28] samples respectively.

6.5 Fuzzy Seeded Region Growing

FSRG is a spatio-temporal, data-driven post-processing operation designed to enhance the structural boundaries of activated regions. Fuzzy seeded region growing (FSRG) is one of the few methods which incorporates knowledge of spatial proximity between TCs.

This enhancement is gained without the cost of adding false positives or outliers to the region of interest. Table 22 details the FSRG algorithm.

FSRG incorporates this information in a two stage process. Stage 1 of FSRG is the actual region growing where seeds are augmented into robust activation regions using spatio-temporal considerations. The second stage produces fuzzy memberships of the samples in the grown regions with respect to the average TC of the region. This fuzzy membership provides a method to view the centre of the activated region using activation similarity.

Table 22. Fuzzy Seeded Region Growing Algorithm Summary

Description	Region growing in 3d space based on temporal similarity and spatial proximity.
Strengths	Considers multiple domains (spatial and temporal) to provide a better overall representation of the dataset.
Weaknesses	Regions must be seeded by an independent method before they may be grown.
Goal Orientation	Grow regions of temporally similar samples based on heuristics for growth size and direction.
Degrees of Freedom	Regions seeds. Growth direction and size heuristics. Temporal similarity metric. Spatial distance metric.

Figure 53 shows the initial seed regions and Fig. 54 shows the augmented regions after FSRG has been applied. Note that the mean TC to which candidate voxels are considered

changes as TCs are added to the FSRG ROI.

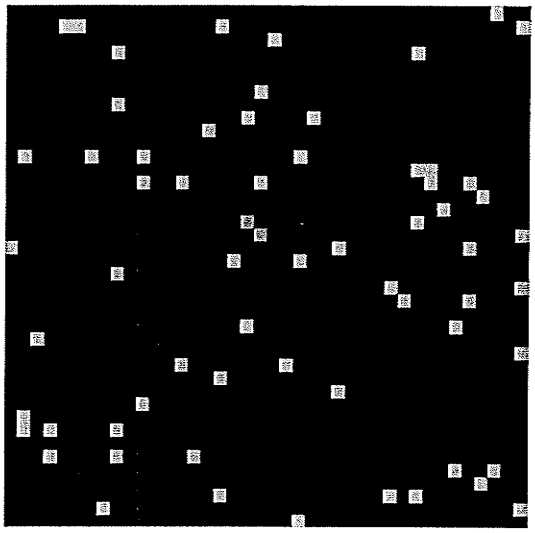


Figure 53. FSRG seed points.

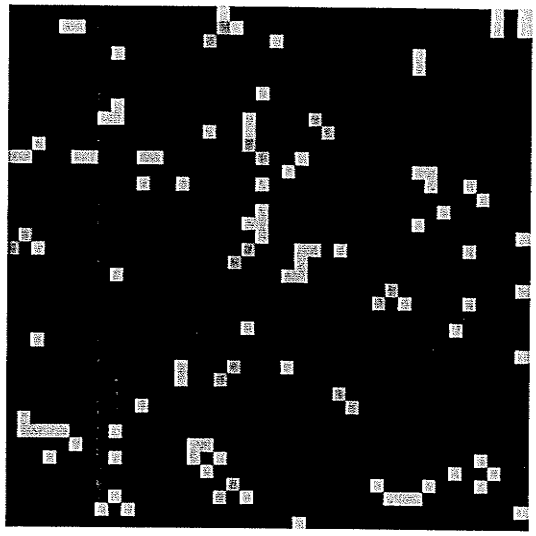


Figure 54. ROI generated by FSRG. Seed points to left were used.

A possible variant of region growing is noted here with a region growing process that involves FCMP.

Region Growing Variant

Region growing methods require a seed, or sample, which serves as the initial point for a collection of samples. A region representative is similar to a cluster centroid; it is a sample used to characterize the region. A region descriptor is any value or statistic that characterizes a two dimensional, three dimensional, region such as the Euler number or chain code for a region. Region growing accuracy for fMRI may be defined as:

- Regions that correspond to anatomical components.
- Regions that supply important information to the researcher such as ROI.
- Regions that meet the expectations of an expert or gold standard.
- Alternate heuristic criteria. E.g. spatially proximal, temporally-similar TCs should be clustered together.

Region Growing using FCMP

In order to compare FSRG with FCMP, it is necessary to formulate a region process that involves the characteristics of FCMP with respect to its feature partition integration. Table 23 outlines such a process.

Table 23. Region Growing with FCMP

Description	FCMP initializes seed voxels. Regions are grown based on a spatially-ordered temporal similarity.
Strengths	Spatial ordering of temporal information discovers local formations.
Weaknesses	Computation associated with clustering is incurred before regions are grown.
Goal Orientation	Grow spatial regions of temporally similar voxels.
Degrees of Freedom	Temporal and spatial metrics. Region growing parameters such as neighbourhood connectedness.

The robustness of region growing algorithms may be defined as the degree to which the regions do not significantly change as TC features vary slightly. Robustness could be measured in terms of SNR of the TCs to the stimulus paradigm. Besides using FCM centroids as seeds, one may generate seeds by thresholding the mean image intensity.

Common pre-processing includes mathematical morphology, and image filters (high, low, Sobel, median) [Gonz]. Region growing rules may be diverse. A dynamic region growing heuristic may be defined where growing occurs only along the strongest path from origin to edge. This method should map out sulci with correlated activations.

The basic steps to region growing after FCMP are:

1. Rules to generate neighbourhoods.
2. Similarity measure used in generating neighbourhoods.
3. Distance metric used for maximum radius of neighbourhood.

Neighbourhood Definition

A priori regions may be considered for potential sample substitutes. For example, a polygon may be used, samples within a maximum radii may be considered, or a region may be grown dynamically until a certain number of samples have accumulated. Let $N_e(n)$ be the neighbourhood of sample x_n, i.e.

$$N_e(n)=\{x_j | S(x_n,x_j)>T_S \wedge d(x_n,x_j)<T_d\} \tag{45}$$

where $d(°,°)$ is a distance measure, T_S is the similarity threshold, T_d is a distance threshold.

$$\|N_e(n)\|>T_{N_e} \tag{46}$$

Summary

This completes an examination of the benchmark algorithms used to quantify and qualify the distinct contributions made by FCMP to fMRI data analysis. The software implementing the algorithms, library interfaces, data I/O, synthetic fMRI dataset generation, and related utilities was written in C++ as an application specific software kit extension to the Scopira framework [Demk]. The Scopira framework is available under the GNU Public License (GPL) from www.scopira.org.

7. Experiments and Results

No amount of experimentation can ever prove me right; a single experiment can prove me wrong.

Albert Einstein

If your results need a statistician then you should design a better experiment.

Ernest Rutherford

Ten experiments examine the use of FCMP on fMRI datasets and assume familiarity with TC analysis [Cox2] [Pres] [Royer] and fMRI, as per Chapter 5. Experiment structure consists of: a general hypothesis, one or more datasets, a main algorithm with a set of parameter values, one or more benchmark algorithms, with associated parameter values, and a set of performance measurements. In all cases, FCMP is the main algorithm, although in some experiments, such as region growing, it is followed by additional steps. Thus, each experiment quantifies FCMP performance against one or more benchmarks, using one or more datasets, over a defined set of parameters. Comments follow on the composition of the experiments.

Choice of Benchmark

Benchmarks represent basic and well understood algorithms, *de facto* industry standards, and algorithms of note in recent literature and are listed in Table 24. Some experiments compare FCMP with a single benchmark with respect to FCMP while a *salvo* approach, using many benchmarks, is also used. EvIdent®, as an industry standard for fMRI analysis, is a particularly important benchmark. Recall that FCM using only temporal

features is the most common current use of fMRI data.

Table 24. Summary of Algorithms

Acronym	*Algorithm*
FCM	Fuzzy C-Means
FCMP	Fuzzy C-Means with Feature Partitions
CHAMELEON	CHAMELEON
Cluster Merging	Basic cluster merging
EvIdent®	Event Identification
FSRG	Fuzzy seeded region growing
EROICA	Exploring Regions of Interest using Cluster Analysis

Choice of Datasets

A variety of synthetic and *in vivo* datasets, see Table 25, are used to compare the efficacy and robustness of FCMP and the benchmarks. Synthetic datasets were designed to accentuate differences between algorithms through use of somewhat idiosyncratic structures, a range of noise levels, partially correlated feature partitions, and superimposed, unanticipated novelty. *In vivo* datasets provide industry-norm data acquisitions with concomitant noise types and levels.

Table 25. Summary of Datasets

Name	*Type*	*Comments*
Syn1	Synthetic	Dataset with correlated features.
Syn2	Synthetic	Data to demonstrate region growing.
Syn3	Synthetic	Data to demonstrate basic clustering concepts.
Syn4	Synthetic	Small scale fMRI data.
Syn5	Synthetic	Small, unanticipated signal on the activated epochs.
BaumNull	Hybrid	Gamma function activation injected into a null fMRI scan.
S05	*In vivo*	Checkered visual stimulus.
Sample4d	*In vivo*	Tactile stimulus.
Halx	*In vivo*	Tourette's syndrome study.

Choice of Parameters

Parameters constitute degrees of freedom for the experiment that may constrain the optimization of the algorithm. This is significant when parameter optimization tends to be dataset-specific. Exhaustive examination of parameter-spaces is rarely feasible and extends far beyond the thesis scope. Limiting parameter evaluations does have the benefit of simplifying the variables in the investigation. Justification for the parameters evaluated is given in each experiment synopsis. Details on algorithm parameters are given in Chapter 4 for FCMP and Chapter 6 for the benchmarks.

Experiment Synopsis

Each experiment is characterized by its hypothesis. In order to provide a consistent structure for experiment synopses, a common table format is used, see Table 26. Each experiment augments its synopsis with further relevant details, and introduces functions, statistics, and plots requisite for interpretation of the experiment results.

Table 26. Experiment Synopsis Format

Experiment Component	*Description*
Name	Uniquely describes the experiment in the thesis and is used to relate the description with the results and accompanying figures and tables.
Description	Adds details as to method, dataset, and algorithm parameters.
Independent Variables	List of the experiment factors, typically the parameters used and any preprocessing. Details the parameter sampling grid, if one was used.
Dependent Variables	List of dependent measurements of performance measurements.
Hypothesis	Postulates an expected outcome of the experiment.
Main algorithm	Details any auxiliary processing used with FCMP.
Benchmarks	List of algorithms used for comparative purposes to FCMP.
Datasets	List of datasets used.
Validation	Provides an explanation or interpretation of the performance measures with respect to the experiment. Fixes the means by which different algorithms are to be compared.

Tables 27 lists the experiments conducted in this thesis. Detailed experiment descriptions are provided sequentially in this chapter.

Table 27. Summary of Experiments

Index	*Datasets*	*Benchmarks*	*Comments*
1	Syn1	FCM	**Concept partition** with respect to various SNR levels.
2	Syn1 Syn3	FCM	**Utility of validation indices:** PC,PE,FS,XB, as well as HCM metrics.
3	Syn1 Syn3 BaumNull *in vivo*	FCM	**Utility of VCV** matrix for determining the number of intrinsic clusters.
4	S05 Sample4d	FCM	**Optimal partition weights** with respect to validation indices on *in vivo* datasets.
5	Syn1 Syn2 Syn3 S05	Cluster Merging CHAMELEON	Examination of **induced cluster hierarchy.**
6	Syn2 BaumNull *in vivo*	EvIdent®	**Region definition** test the algorithm robustness with respect to outliers and spatial distribution metrics.
7	Syn2 *In vivo*	FSRG EvIdent®	**Region growing** based on TC seeds generated by FCM, FCMP and global statistics.
8	*In vivo*	EvIdent®	**Activated epochs** have increased weight and a separate partition in the clustering process.
9	Syn5	FCM EvIdent®	Test of ability to **detect novelty**, eg of small signals on activated epochs.
10	*In vivo*	EvIdent® FSRG	**Bridge voxels** define global thresholds for fMRI datasets.

Discussion of the experiment results conforms to the following format:

Executive Summary: A concise rendering of experiment significance.

Contribution: Components of original work are noted.

Overview: Extends the experiment synopsis.

Validation: Describes the method by which the experiment results are evaluated, confirming or countering the hypothesis. It may also impose a caveat on interpretation of the results.

Discussion: Experiment results, such as notes, tables and figures, are integrated.

The order of the experiment results are:

(1) Concept Partition, (2) Validation Indices, (3) Utility of VCV,

(4) Optimal Partitions, (5) Induced Hierarchy, (6) Region Definition,

(7) Region Growing, (8) Activated Epochs, (9) Novelty Detection, (10) Bridge Voxels.

Generally, results for a single experiment are nested in the following structure:

- According to dataset, with order: focus, synthetic, hybrid, *in vivo*.
- According to algorithm, with order: FCMP, benchmarks.
- According to algorithm parameter values, with order: increasing resolution or specificity.

As a rule, in depth details and analysis are provided for only one dataset per experiment, the focus item above, while results for other datasets are summarized in a few sentences. Comments that extend beyond a single experiment are noted at the end of this chapter. Experiments are now discussed in turn.

7.1 Concept Partitions

The concept partition experiment, see Table 28, examines Syn1, a synthesized fMRI dataset where spatial features form two distinct clusters. Each spatial cluster corresponds to a specific set of TCs. One set of TCs are pure noise while the other is composed of TCs that are correlated to the paradigm at a specific SNR level.

Executive Summary At high SNR levels spatial and temporal features can, individually, discriminate the two classes of samples. Integrating both types of features in FCMP

marginally enhances centroid-paradigm MSE and correlation. However, as SNR levels decreased, an optimal feature partition integration of the spatial and temporal features was observed.

Contribution An optimal feature partition integration, $[\nu_S\ \nu_T] = [0.1, 0.9]$, was found for FCMP, as measured by MSE and the centroid-paradigm correlation.

Overview This experiment examines a parameterized synthetic dataset, Syn1, over a range of SNR values: {50,30,20,10,5,2,1}. See Fig. 55 for typical TCs.

Validation MSE and correlation of the centroids are compared to the paradigm.

Discussion

Syn1 Results Several expected trends were confirmed by this experiment. As parameter C increases, the overall MSE decreases, regardless of SNR for the correlated TCs and the spatial weight ν_S. As the noise in the dataset increases, the minimum MSE of the resulting centroids increases. However, for an increase in the number of clusters, both FCM and FCMP are more likely to generate additional noise clusters. The following relation is noted between SNR, MSE and ν_S: as $\nu_S \rightarrow 0$ the MSE depends increasingly on the SNR. That is, TCs with higher temporal SNR have lower MSE for the same values of ν_S.

Table 28. Concept Partition Synopsis

Experiment Component	*Description*
Index: 1	Concept partition.
Comment	For the synthetic dataset described above, how does FCMP compare to FCM over a range of SNR values?
Independent Variables	SNR value of degraded TCs {2,5,10,20,30,40}. Number of samples in the two clusters {100}. Spatial distance between spatial cluster centres {1}. Metric applied to each partition {Euclidean, Pearson distance}. Weight applied to each partition {0, 0.1, 0.9, 1.0}.
Dependent Variables	The mean square error (MSE) of the centroid from the cluster of degraded TCs to the stimulus paradigm. Same as above but using the Pearson Correlation of the centroid to paradigm.
Hypothesis	FCMP will achieve a lower MSE and a higher correlation.
Main algorithm	FCMP
Benchmarks	FCM
Datasets	Syn1
Validation	Minimum MSE of centroid to paradigm. Maximum correlation of centroid to paradigm.

As reliance on the low noise feature partition is increased, that is $\nu_s \rightarrow 1$, MSE decreases. This occurred for almost all cases; for the synthetic dataset at low SNR, a local optima was noted. At this combination of feature partitions the incorporation of degraded signals is beneficial. This inflection point, against the general trend of decreasing MSE for increasing ν_s, was noted for low SNR and occurred at $\nu_s=0.9$ as shown in Fig. 56. Significantly, a range of SNR values exhibit this inflection and $\nu_s=0.9$ seems optimal for

this particular dataset and feature partitions. The slope of the trends in Fig. 56 suggest a critical SNR value at which ν_s has a significant ameliorative effect, 2-5 SNR.

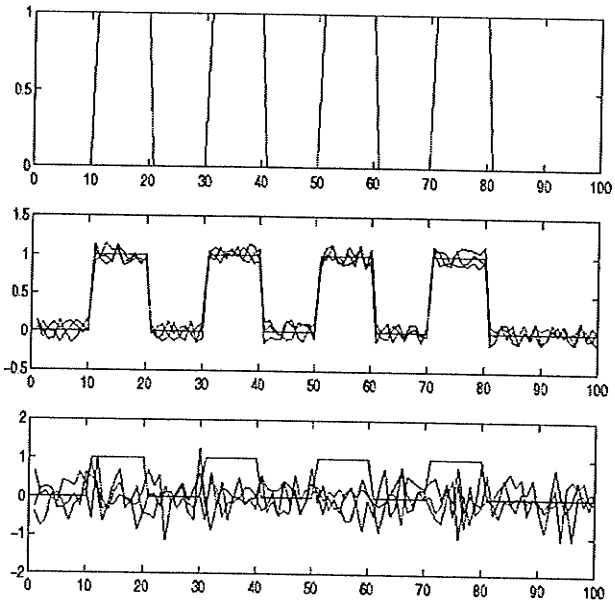


Figure 55. TC in Syn1.

top) paradigm TC. middle) degraded TCs.
bottom) noisy TCs.

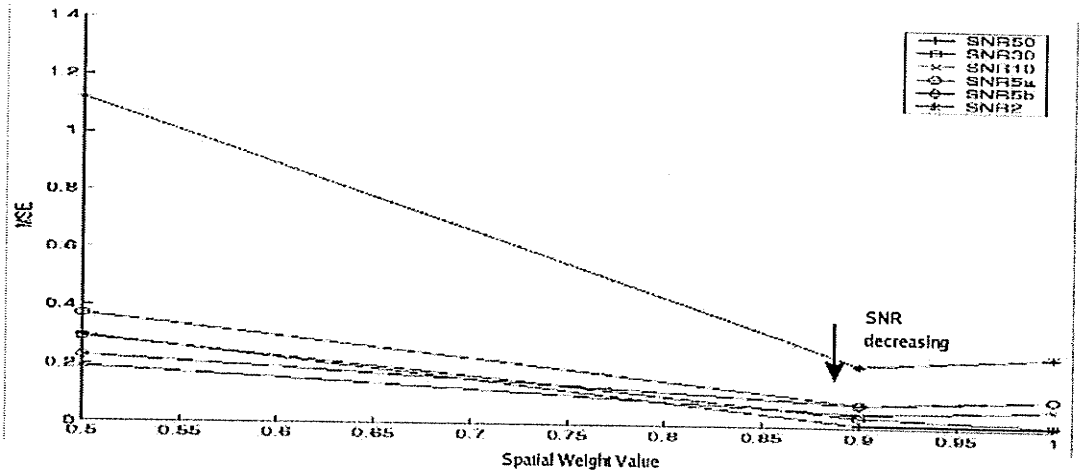


Figure 56. Partition weights vs MSE.

7.2 Validation Indices

Chapter 4 introduced several cluster validation indices. We now determine their application with respect to fMRI datasets, see Table 29. Synthetic and *in vivo* datasets are examined.

Executive Summary Validation indices are used to compute the optimal number of clusters in an fMRI dataset. The Xie-Beni index, XB, is shown to be most efficacious at indicating intrinsic structure.

Contribution Validation indices are evaluated on FCMP centroids of fMRI data.

Overview The utility of validation indices in conjunction with FCMP is examined on fMRI datasets. The Xie-Beni index, which has shown most potential for validation of general data with FCM, also shows the most potential with FCMP and fMRI data.

Validation The number of clusters indicated by the validation index is compared to a gold standard or to the designed structure of the synthetic datasets.

Discussion

For each dataset, plots of the validation indices, {PE, PC, PX, FS, XB}, are presented for different number of clusters and different feature partitions.

Table 29. Validation Indices Synopsis

Experiment Component	*Description*
Index: 2	Validation Indices
Comment	Interpret how the various validation indices should be used on fMRI datasets. Determine the relation of clustering algorithms to validation indices generally.
Independent Variables	FCM parameters, FCMP parameters.
Dependent Variables	Validation indices: FS, PC, PE, PX, XB.
Hypothesis	The commonly used set of FCM validation indices are not general enough to give consistent results for both synthetic and *in vivo fMRI* datasets.
Main algorithm	FCMP
Benchmarks	FCM
Datasets	Syn1, Syn3.
Validation	Validation index corresponds to an expert opinion (or gold standard) relating to the data.

Syn1 Results As the SNR between the paradigm prototype and the correlated TCs decreased, the validation indices diverged in their agreement as to the intrinsic structure (number of clusters) in the dataset. XB was the preferred index and was most consistent with respect to changes in SNR.

Syn3 Results For the simple two cluster structure in Syn3, all validation indices performed well and correctly indicated the presence of two clusters.

7.3 Visual Cluster Validity

The visual cluster validity (VCV) [Hath2] index is an indicator of data structure. By first sorting the samples based on their cluster assignment, a matrix of all inter-sample relations reveals dataset structure. The experiment is outlined in Table 30.

Executive Summary The VCV index displays intrinsic cluster structure and is compared to other validation indices. The VCV index corresponds well to gold standard clusters.

Contribution Evaluation of VCV on FCMP clusters of fMRI data.

Overview The VCV index uses sorted, inter-sample relations to visually represent data structure.

Validation Correspondence of the VCV index to gold standard cluster structures.

Discussion

Basic pre-processing operations reduce the number of TCs under consideration. Thresholds for average intensity and correlation to a stimulus paradigm are used.

Table 30. VCV Synopsis

Experiment Component	*Description*
Index: 3	Visual cluster validity.
Comment	Determine the effectiveness of the VCV index for clustering fMRI data.
Independent Variables	FCM, FCMP parameters.
Dependent Variables	VCV index.
Hypothesis	The VCV index detects the intrinsic data structure for FMCP centroids more accurately than for FCM centroids (since the FCM centroids ignore the spatial features).
Main algorithm	FCMP
Benchmarks	FCM
Datasets	Syn3, BaumNull, *in vivo*.
Validation	Correspondence of the data structure as suggested by the VCV index and the parameters used to generate the synthetic data.

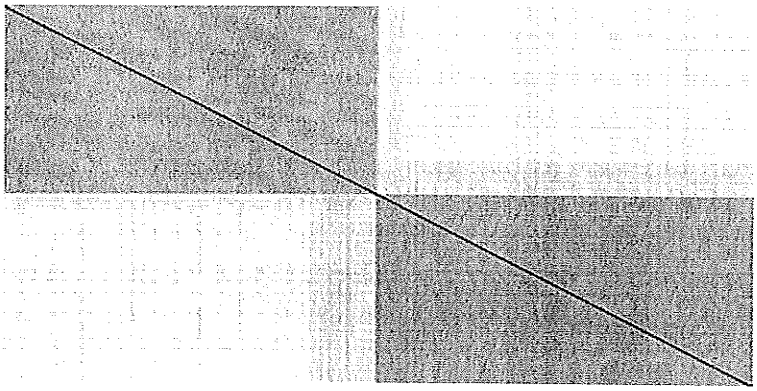


Figure 57. VCV matrix for two distinct clusters.

Dark regions on diagonal indicate two inherent clusters.

Thresholds for average intensity are set as percentages of the maximum intensity in the dataset. Thresholds for TC correlation are set to select only those TCs with marginal and

significant correlation. The number of samples in S05 remaining after intensity and correlation thresholding are shown in Tables 31.

For each reduced dataset, two *VCV* matrices are computed, using FCM and FCMP centroids respectively, and the number of inherent clusters determined by heuristics. Recall that, unlike other validation indices, the *VCV* index is computed once per cluster algorithm, usually with a large number of clusters (20).

S05 Results Figures 58 and 59 show the VCV matrix for intensity thresholded samples where the initial centroids were defined using FCM and FCMP respectively. Figures 60 and 61 show the VCV matrix for correlation thresholded samples with initial centroids were defined using FCM and FCMP.

FCM Results Intensity thresholds produced consistent VCV indicators for two clusters, of unequal sizes, over all threshold values. Correlation thresholds produced a large dark region indicating a single cluster. It is noted that this single cluster is slightly larger than the largest indicated by the intensity thresholds. As the number of samples increases, that is the intensity threshold becomes lower, the single cluster grows in size.

Table 31. Thresholding Effect on S05

Correlation Threshold	*Number of samples*	*Intensity Threshold (%)*	*Number of samples*
0.18	4397	0.34	4103
0.25	2015	0.47	2054
0.27	1585	0.51	1507
0.3	1085	0.55	1056
0.35	537	0.61	519

FCMP Results Intensity thresholds exhibited resolution control in the number of clusters for FCMP results, see Table 32. At high thresholds only one cluster was discernible; at lower thresholds, with more samples, two clusters were indicated. Use of correlation thresholds, also in Table 32, show similar results.

FCM vs FCMP The dark regions in the FCMP VCV images were generally more homogeneous, meaning that light lines, indicating outliers, were less common, than for FCM. FCMP also provided useful information for cluster merging. As noted, dark patches off the diagonal indicate overlap between sample groups. These off-diagonal regions appeared more often using FCMP.

Intensity vs Correlation For both FCM and FCMP, VCV generated from correlation thresholds showed greater cluster size variance. VCV is efficacious in determining the number of sub-clusters in fMRI datasets. FCMP was shown to produce

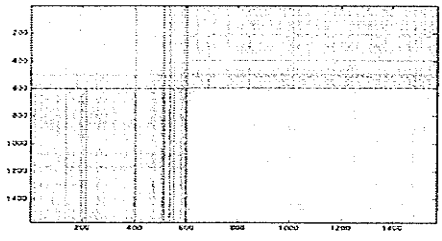


Figure 58. VCV Intensity image FCM, ν_s=0.59.

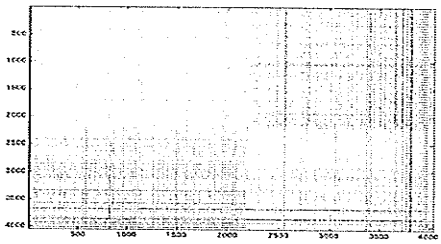


Figure 59. VCV Intensity image FCMP, ν_s=0.4.

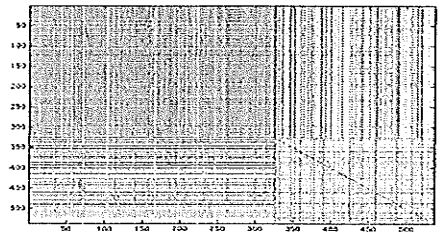


Figure 60. VCV Correlation image FCM, ν_s=0.35.

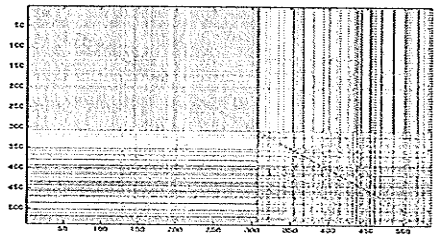


Figure 61. VCV Correlation image FCMP, ν_s =0.35.

Table 32. VCV Indices for FCMP Clusters

Intensity Threshold	***FCM***	***FCMP***	***Correlation Threshold***	***FCM***	***FCMP***
0.34	2	1	0.18	1	1
0.47	2	1	0.25	1	1
0.51	2	1	0.27	1	1
0.55	2	2	0.3	1	1
0.61	2	2	0.35	1	1

finer structure that related to the level of integration of spatial context. A binarized VCV matrix (bVCV), seen in Fig. 62, is constructed by thresholding similarity values in the VCV matrix and allows the analyst to estimate size robustly since the cross-hatching noise is reduced.

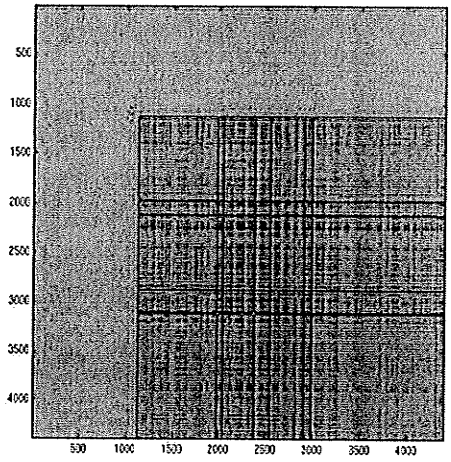


Figure 62. Binarized image of VCV.

Syn1 Results As the SNR decreased, the VCV matrices indicated the presence of an increased number of clusters along with a corresponding decrease in cluster size.

Syn3 Results VCV correctly indicated the presence of two clusters.

Baumnull Results The spatial localizations of the synthetic activations were detected in the VCV matrix.

Sample4d Results The VCV matrix indicated four main clusters over the time periods.

Halx Results One z-plane (z=14) was examined for this dataset. Six clusters were indicated by the VCV matrix. The clusters were fairly consistent in size and shape over all time periods.

7.4 Optimal Partitions

In order to determine the existence of optimal weighting between feature partitions, a series of weight combinations is examined for each dataset, see Table 33.

Table 33. Optimal Partition Synopsis

Experiment Component	*Description*
Index: 4	Optimal Partitions
Comment	The partition weight parameter offers a continuous method of integrating feature partition information (and minimizing feature partition noise) when two partitions exhibit correlation.
Independent Variables	Dataset. Number of samples in the clusters. Weight and metric applied to each partition.
Dependent Variables	MSE of centroid to paradigm. Correlation of centroid to paradigm.
Hypothesis	FCMP will exhibit a dataset-specific optima with respect to the partition weights.
Main algorithm	FCMP
Benchmarks	FCM
Datasets	S05, Sample4d.
Validation	Minimum MSE of centroid to paradigm. Maximum correlation of centroid to paradigm.

Executive Summary A high resolution parameter-space for FCMP partition weights is examined with respect to the S05 and sample4d datasets.

Contribution It is shown that an optimal partition weight exists for a particular dataset. See the SNR table in experiment 1 results.

Overview The partition weight parameter offers a continuous method of integrating

feature partition information, as well as minimizing feature partition noise, when the two partitions exhibit correlation.

Validation MSE and correlation compare centroids to the paradigm.

Discussion

S05 Results The optimal two-feature partition consisted of a spatial features with weight 0.8 and temporal features with weight 0.2, using the Euclidean and Pearson Correlation metrics respectively.

Sample4d Results The optimal two-feature partition consisted of a spatial features with weight 0.7 and temporal features with weight 0.3, using the Euclidean and Pearson Correlation metrics respectively. The change in weights with respect to the results from S05 may be related to the increased number of z-planes available for sample4d, (4 vs 1).

7.5 Induced Hierarchy

In fMRI region growing, temporal cluster analysis and region (spatial) analysis can be integrated to determine a hierarchy for region growth. Spatial and temporal features are integrated in that temporal similarity directs spatial association. Details are provided in Table 34.

Executive Summary Hierarchies induced by different clustering algorithms are

compared. FCMP shows versatility in integrating feature relevance with respect to the objective function. FCMP shows an intermediate performance, between that of FCM and CHAMELEON, in characterizing the idiosyncratic nature of individual clusters in a dataset.

Contribution Sample hierarchies from FCMP and CHAMELEON are compared.

Overview Comparison of partition matrices generated by basic cluster merging, CHAMELEON and FCMP.

Validation Correspondence of induced hierarchy to a gold standard organization.

Discussion

It was noted that phase II of the CHAMELEON algorithm, the agglomeration phase, was infrequently used as the parameters specified a maximum graph partition size. This is due to our clustering objectives which specify a range for the allowable number of clusters and was true for all datasets examined.

Syn1 Results With respect to the designed association between spatial coordinates and temporal intensity patterns, decreases in SNR had the expected outcome of increasing the mixture (entropy) of the sample hierarchies.

Table 34. Induced Hierarchy Synopsis

Experiment Component	*Description*
Index: 5	Induced hierarchy.
Comment	Comparison of induced hierarchies in the clusters and partition matrices as generated by basic cluster merging, CHAMELEON and FCMP.
Independent Variables	FCMP parameters; CHAMELEON parameters; basic merging parameters.
Dependent Variables	Hierarchical organization. Ability to express idiosyncratic structure of individual clusters.
Hypothesis	FCMP generates a hierarchical organization of the data that more accurately corresponds to a conceptual understanding of the data. FCMP captures idiosyncratic structure across feature partitions.
Main algorithm	FCMP
Benchmarks	FCM, basic cluster merging, CHAMELEON.
Datasets	Syn1, Syn2, Syn3, BaumNull, *in vivo*.
Validation	Correspondence of resultant hierarchy to inherent organization of samples.

Syn2 Results Due to the geometric structure of the ROI in this dataset, FCMP results could be enhanced through the use of a metric that exploited the spatial distribution of the ROI. however, even if the structure was known a priori, such results would be ungeneralizable to the other datasets in this study. For this dataset, CHAMELEON generated sample hierarchies with fewer errors than FCMP across the feature partitions studied.

Syn3 Results FCMP outperformed CHAMELEON using the Euclidean / Pearson correlation metrics. The Euclidean metric was advantageous for use with FCMP as a

radially symmetric distance measure was optimal for the circularly distributed spatial coordinates.

S05 Results FCMP again outperformed CHAMELEON using the Euclidean / Pearson correlation metrics over a range of spatial weight values (0.7-0.9). While the clusters in s05 do not have a circular shape, the can be well approximated by the FCM default cluster shape and the large number of clusters used generated a sample hierarchy with fewer errors than CHAMELEON.

7.6 Region Definition

Regions of interest generated by different algorithms are compared with respect to neural activation studies with known stimuli, see Table 35. The main algorithm pseudo-code is: FCMP generates centroids. For each centroid, its surrounding neighbourhood is grown. If not all samples are captured, remove all samples in the region from consideration and cluster again. Grow from the resulting centroid.

Executive Summary A comparison of ROIs generated by FCMP and FSRG is made. Region growing using FCMP show exclusion of spatial outliers.

Table 35. Region Definition Synopsis

Experiment Component	*Description*
Index: 6	Region definition.
Comments	Compare ROI associated with the dataset as they are generated by different algorithms. Examine the EDA potential and robustness of FCMP versus an industry benchmark.
Independent Variables	Region seed generation; method of representing regions; region growing algorithms; region growing algorithm parameters; definition of accuracy.
Dependent Variables	Resultant regions.
Hypothesis	Region seeds generated by spatio-temporal clustering provide a more accurate region, in terms of region representation, than FCM on TC data only.
Main algorithm	FCMP
Benchmarks	FCM for seed generation. Mean intensity for seed generation. FSRG.
Datasets	Syn2, BaumNull, *in vivo.*
Validation	MSE and correlation of region representative to paradigm. Compare region representatives for all region growing algorithms for all datasets.

Contribution A comparison of FCMP to FSRG.

Overview Compare the regions of interest, e.g. visual cortex, associated with the dataset as they are generated by different algorithms. Examine the EDA potential and robustness of FCMP versus an industry benchmark. FCMP generated regions of interest that compared well those generated by FSRG.

Validation Compare region representatives for all region growing algorithms for all

datasets.

Discussion

S05 Results An intensity image, Fig. 63, shows regions with high average intensity as dark. Notable is the visual cortex region and a motor or memory region located closer to the bottom of the image. A histogram of the correlation values, Fig. 64, of the TC to the paradigm TC show the significant level of noise in the dataset.

Figure 63. S05 Mean intensity coronal image.

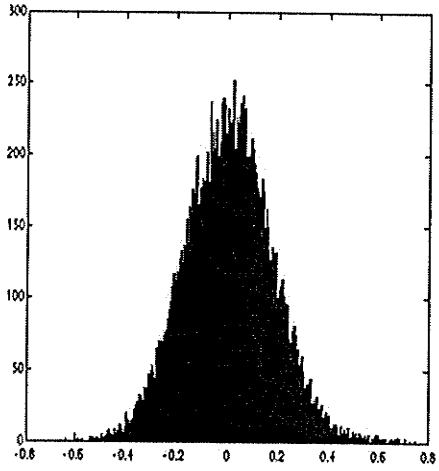


Figure 64. S05 correlation histogram.

An examination of the histogram values in S05 shows that regions with TCs highly correlated to the paradigm will be small. Also, that any region growing operations in the region of interest will suffer loss of continuity (will have a number of included holes / a higher Euler number) unless they accept marginally correlated TCs that are nonetheless

are spatially proximal. Examining the spatial distribution of correlated voxels, Fig. 65, a correspondence is seen between the high average intensity voxels. Note that (i) the visual cortex has a significant number of highly correlated voxels, (ii) some noise TCs (outside of the subject's body) have a significant correlation, see Fig. 66, and (iii) a second separate region of correlated voxels exists, again, near the bottom of the image. Several TCs are displayed in Fig. 67. They show the stimulus paradigm (010110) mapped to the 42 sampling instants, TCs in the visual cortex and noise TCs.

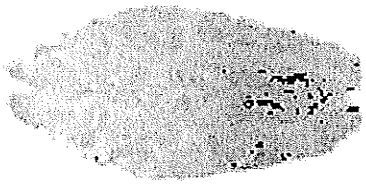


Figure 65. S05 TCs with significant correlation.

Figure 66. S05 outlier and missed TCs. Dilation of coordinates from FCM are shown.

Using mathematical morphology, the spatial locations of significant correlation in the dataset can be dilated. This dilation takes into account only spatial topology. Thus voxels with lower temporal correlation have been included in the darker areas. Such an operation has increased the spatial continuity of the region of interest.

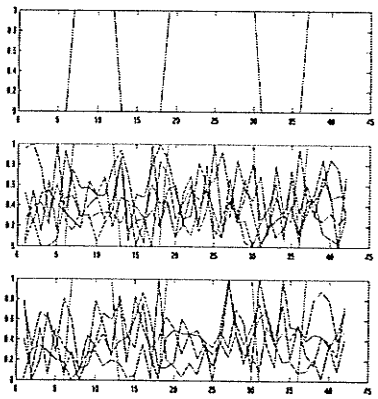


Figure 67. fMRI TCs (s05).

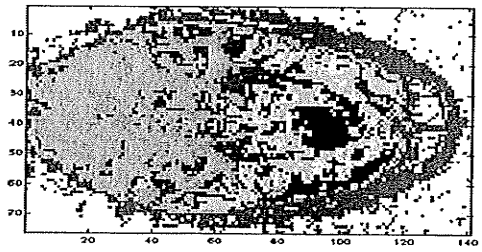
top) paradigm. middle) visual cortex TCs.
bottom) noisy TCs.

Applying FCMP to S05 Previously the S05 dataset was examined at a general level using basic pre-processing operations. It was seen that spatial continuity of the ROI could be achieved using spatial topology without any consideration for temporal similarity. FCMP will address this by defining spatial and temporal feature partitions. One problem for FCMP is that it is not obvious how to relate the spatial and temporal domains. This is solved by executing a series of cluster experiments, each with a different weight between the partitions. Results are shown in Fig. 68. Note that when one feature partition weight is 0 and there are only two weights, the results are exactly that of FCM with the same metric function. When $\nu_S=0$ one expects to have spatially disconnected regions; when $\nu_S=1$ one expects to have spatially convex regions, Fig. 68d.

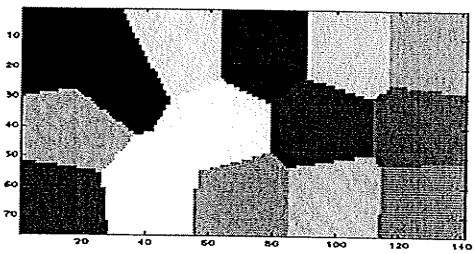
When $0 \leq \nu_S \leq 1$ interesting affects can be observed in the resulting voxel assignment maps. Temporal similarity and spatial proximity combine to produce spatial regions that exhibit degrees of spatial continuity. The centroids of these regions show corresponding increasing correlation (for regions correlated to the visual cortex) or decreasing correlation (for noise regions). Based on MSE of the centroids to the paradigm, $0.6 \leq \nu_S \leq 0.9$ values gave best results. Examining just the visual cortex region, Fig. 69, one can detect the increase in spatial continuity as the spatial feature partition weight increases.

Syn2 Results Each ROI was enhanced by an increased spatial weight with the effect being most noticeable at the interior and exterior corners respectively.

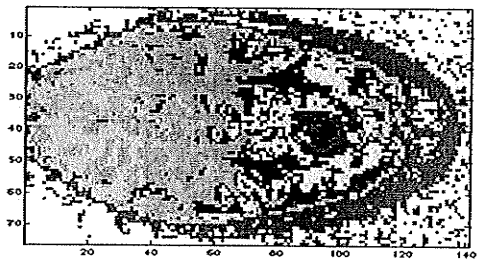
Baumnull, Sample4d, Halx Results Increasing the contribution of spatial features enhances spatial continuity and size of ROI related to areas of maximum intensity, intensity variance and correlation. Also, spatial outliers were noticeably reduced as spatial weightings increased in all datasets.



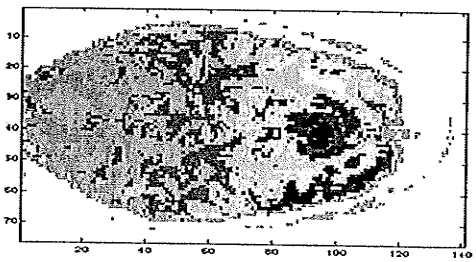
(a) C=15, ν_s=0.



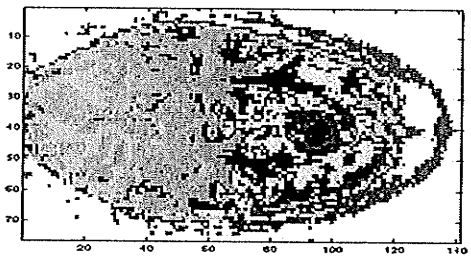
(d) C=15, ν_s=1.



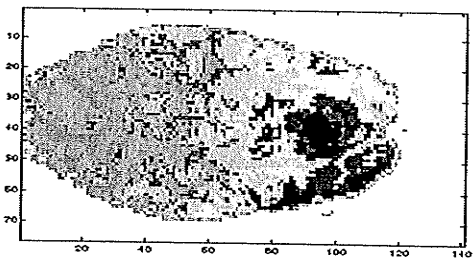
(b) C=15, ν_s=0.2.



(e) C=15, ν_s=0.6.



(c) C=15, ν_s=0.4.



(f) C=15, ν_s=0.9.

Figure 68. S05 FCMP spatial maps with various parameters.

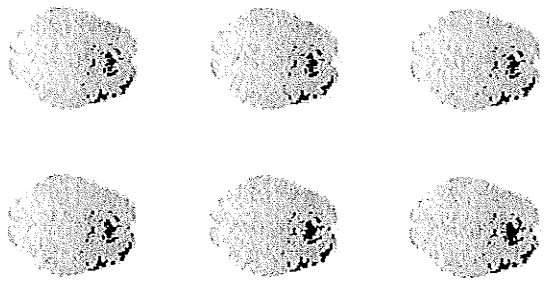


Figure 69. Visual cortex voxels (S05) as spatial weight increases.

7.7 Region Growing

Consideration of fMRI ROI is often based on methods that pare down the dataset to exclude TCs when they no longer meet inclusion criterion. On the other hand, region growing processes extend consideration to unselected voxels and TCs implementing a dynamic inclusion criterion with respect to region growth. Table 36 outlines a comparison of different region algorithms on fMRI data.

Executive Summary FCMP is adapted as a region growing processes to use its spatio-temporal integration as a dynamic inclusion criterion.

Contribution The results of FCMP are compared to an industry standard, EvIdent®.

Overview The utility of using FCMP as part of a region growing process is compared to industry benchmarks.

Validation Comparison of ROI to gold standards.

Table 36. Region Growing Synopsis

Experiment Component	*Description*
Index: 7	Region growing
Comment	Compare regions grown by different methods on fMRI datasets.
Independent Variables	Region seeds. FSRG parameters.
Dependent Variables	Generated ROI.
Hypothesis	The region growing method using FCMP will be comparable to FSRG since they both take into account spatio-temporal information.
Main algorithm	FCMP based region growing.
Benchmarks	FSRG
Datasets	Syn2, *in vivo.*
Validation	ROI analysis of gold standard.

Discussion

S05 Results TCs proximal to the ROI are seen in Fig. 70. They were not associated with the visual cortex by EvIdent® but were by FCMP over a range of partition weights, ν_S, $0.6 \leq \nu_S \leq 0.9$.

Figure 70. TCs not in S05 visual cortex.

TCs spatially proximal to the ROI, in this case the visual cortex, but that were not selected as significant by EvIdent® are shown in Fig. 71. These TCs are denoted as *proximal rejections.*

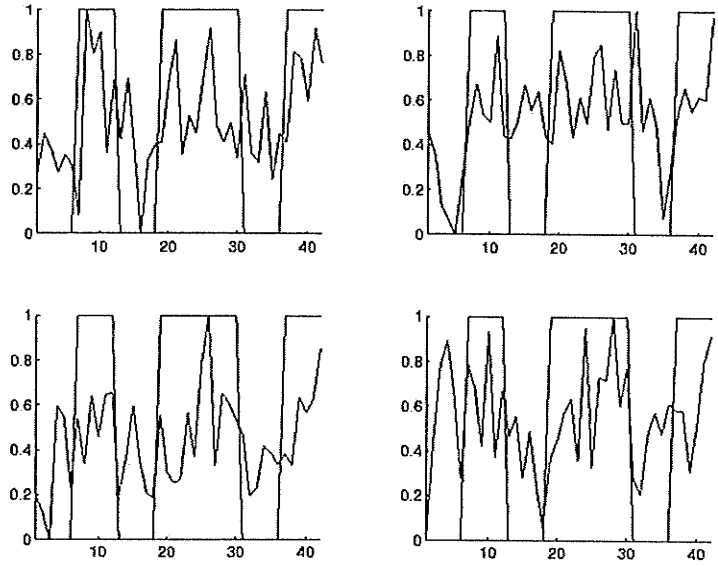


Figure 71. TCs proximal to ROI but not selected by EvIdent®.

Figure 72 shows TCs that EvIdent® identified as being significantly related to the stimulus paradigm although they were not spatially proximal to the ROI. These TCs are denoted *outlier inclusions.* Table 37 provides coordinates and correlation values for both the proximal rejections and outlier inclusions of EvIdent® with respect to FCMP. Given the histogram of correlation values for the datasets, these TCs can be considered significant. This correlation coupled with spatial proximity to the ROI make them excellent candidates for further examination. Therefore, FCMP has a role to play in the

EDA of fMRI not yet filled by industry standards.

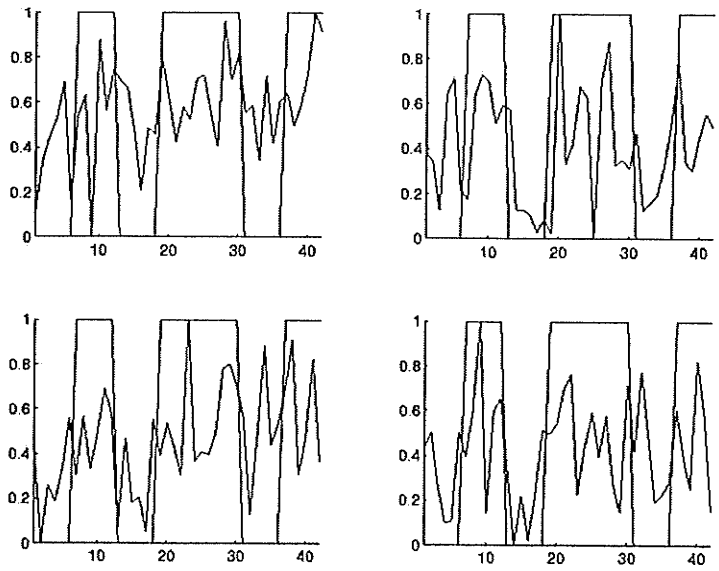


Figure 72. Outlier TCs selected by EvIdent®.

As part of this experiment, the average distance from a sample assigned to the visual cortex cluster over the partition weights was measured. This distance was also considered when the furthest 10% of the samples in that cluster were ignored. The average distance decreases as ν_S increases and that ignoring the farthest samples in that cluster has an appreciable effect on the distance computation over all feature partition weight values.

FSRG is well suited to growing regions that matched gold standard ROI and is constrained to grow in a spatially contiguous manner; clustering algorithms such as EvIdent® and FCMP may group temporally similar samples that are not contiguous.

Table 37. Evident® Test Cases in S05 TCs

	(x,y) coordinates	*Pearson Correlation*
Outliers	(87,196)	0.4043
	(89, 210)	0.4022
	(95, 58)	0.4156
	(104, 179)	0.4097
	(108,189)	0.4001
Missed	(57, 185)	0.4780
	(60, 204)	0.4529
	(64, 197)	0.5309
	(76, 205)	0.4071

One method of incorporating the spatial-temporal tuning advantages of FCMP with the region growing ability of FSRG is to use FCMP as a seed generator for FSRG. Seeds generated in a default manner, that is, by thresholding intensity or correlation values, and seeds generated by first running FCMP on the dataset, generated seeds in the same general regions while the optimal feature partitions of FCMP tended to better match the gold standard regions. However, optimal partitions are not known a priori.

Syn2 Results EvIdent® was able to detect the correct ROI and should be considered better than FCMP on this dataset as the appropriate feature partition values that one should use with FCMP are not known beforehand. FSRG correctly discovered the ROI and using FCMP as a seed point generator did not improve results. However, on a dataset with a less structured geometry, FCMP will tend to find similar regions across bridge voxels, as mentioned in the next section.

7.8 Activated Epochs

Interest in the activated epochs of an fMRI study are significant since the discovery of novelty related to the stimulus enhances understanding of the complex functional relationships present in neural activation studies. Since the signal characteristics of the novelty are not known *a priori*, analysis algorithms must allow general detection of unspecified relationships, or allow rapid and flexible tuning of their objective functions allowing the analyst to direct the objective function. This second case is examined with respect to FCMP where the objective function is progressively modified in order to highlight novelty in the dataset. Table 38 details the experiment and comparison will be made between FCMP and FCMP with increased weight on the activated epochs.

Executive Summary The flexibility of FCMP feature partitions are used to enhance sensitivity of the clustering algorithm to sample properties in the activated epochs.

Contribution Demonstration of selective sensitivity augmenting normal analysis.

Overview Feature partitions in this experiment include: spatial coordinates, activated temporal features, and unactivated temporal features. Activated epochs are determined with reference to the paradigm and receive an increased weight with respect to the unactivated epochs.

Validation Identification of a centroid corresponding to the paradigm.

Discussion

Sample4d Results When only the activated epochs are examined, a slight change in the range of spatial weights that produce ROI, comparable to gold standard ROI, is observed. While activated epochs seem of some benefit to generating ROI from an fMRI dataset, it is conceivable that unactivated epochs may at times contain significant identifying characteristics, for clustering or classification. Activation or stimulus type is one possible criterion for determining whether the unactivated or activated epochs is more significant for the analysis at hand.

Table 38. Activated Epochs Synopsis

Experiment Component	*Description*
Index: 8	Activated epochs.
Comment	Consideration of data acquisitions should be weighted according to their relevance in the study. Blood flow during activated epochs can be more informative in many circumstances.
Independent Variables	FCMP parameters. EvIdent parameters.
Dependent Variables	Cluster results.
Hypothesis	FCMP returns significant benefits when integrating the activated epochs at a higher level of relevance to the clustering process.
Main algorithm	FCMP
Benchmarks	EvIdent®
Datasets	*in vivo*
Validation	MSE of centroid to paradigm. Correlation of centroid to paradigm.

7.9 Novelty Detection

The ability of different algorithms to detect a novel signal in an fMRI dataset is examined, see Table 39. Only a few *TC*s in the dataset will contain the novel signal and detection will be confounded in that *TC*s will be spatially distributed.

Background

A *novel TC* refers to one in relation to an external standard (eg, the applied stimulus) and which does not account for, or occur in, the majority of the samples in the dataset. Novelty can also refer to the existence of properties of which the researcher had not

anticipated and for which the experiment was not intentionally designed to capture. In Syn5, the presence of a small signal on the activated epochs is examined. Its discovery under a fine parameter grid approach using FCMP is detailed.

Table 39. Novelty Detection Synopsis

Experiment Component	***Description***
Index: 9	Novelty detection.
Comment	Determine whether the given algorithms can capture a novel *TC*. Under what conditions is this possible?
Independent Variables	Algorithm parameters. Number of novel *TC*s in the dataset. Significance of novelty with respect to the metric used.
Dependent Variables	The clusters containing novel TCs as determined by the various algorithms.
Hypothesis	FCMP can be tuned to capture novel *TC*s more effectively than the other algorithms. FCMP will capture more novel TCs more often.
Main algorithm	FCMP
Benchmarks	FCM, EvIdent, CHAMELEON.
Datasets	Syn5
Validation	MSE and correlation comparisons of the centroid of the cluster containing the most novel *TC*s and the novelty paradigm.

Executive Summary EDA is challenging when the analysis modes of pattern matching and novelty discovery are conflated. Novelty may be characterized as existing among only a small number of samples and as being derived through a higher order transformation from a basic archetype. Both characterizations are explored on synthetic

fMRI data.

Contribution A special instance of FCMP is tested which compares favorably to other methods.

Overview Determine whether given algorithms can capture a novel TC, and under what conditions. FCMP is shown to be able to detect novel small signals at low SNR levels. The performance of FCMP surpasses that of FCM and EvIdent®.

Validation A centroid is detected that correlates to the novelty.

Discussion

Note that the varying SNR relates to the novel signal only and not to the TCs that are correlated to the paradigm, as was the case in other experiments.

Syn5 Results EvIdent® demonstrated the ability to detect the presence of novel signals in a small number of locations over a range of SNR values (40-2 dB). FCMP demonstrated a comparable ability only when the novel signals existed in a single location and shows the limit of a global spatial constraint on the distance measure. It is proposed that, when pattern matching and novelty discovery are being pursued simultaneously, the objective functions or distance metrics associated with the novelty contain the fewest number of constraints. In this case, the distance metric for the novel

TCs should have ignored spatial location and attempted to match only temporal similarity while the distance metric for the correlated TCs was correct in constraining the TCs to also be spatially proximal.

7.10 Bridge Voxels

FMRI analysis of activated regions found by FCM in neural activation studies often detects distinct yet proximal areas with similar centroid *TC*s. When spatial information is incorporated into a region generating analysis, as in the FCMP algorithm, these regions are often merged. It is of interest to what degree the spatial features must be integrated into the region generating algorithms before they are merged. Alternately, a mathematical morphology approach can determine *bridge voxels.* Table 40 outlines the experiment.

Current fMRI analysis generally depends on user-defined intensity thresholds or spatial locations to define regions of interest (ROI). Mean intensity values are commonly used to generate candidate ROI for subsequent analysis. The general validity of such methods suffer from the variety of circumstances under which fMRI data may be acquired. We examine a data-driven method to determine global fMRI thresholds using bridge voxels from intensity and correlation thresholded ROI.

Pre-defined intensity and correlation thresholds are used to generate initial ROI which are subsequently eroded using a mathematical morphology erosion operator. Successive erosions either decrease the ROI area or shatter the ROI into proximal sub-regions. When an ROI shatters, a bridge voxel is said to have been eroded. That is, bridge voxels are the structurally significant voxels that connect the erosion-susceptible components of the initial ROI. Erosion continues until all ROI are completely eroded or shattered. At this stage, the intensity and correlation values of the TC associated with the bridge voxels are computed. These derived intensity and correlation values provide a data-driven global threshold with which to reveal intrinsic ROI in the dataset. Initial ROI are generated by a pre-defined set of of intensity and correlation thresholds. Each ROI is then eroded using a 3 by 3 structuring element and the existence of any bridge voxels are recorded. The mean intensities of the bridge voxels, and their correlation to the stimulus paradigm, are used to generate a set of so-called data-driven thresholds. This second set of thresholds is then used to generate the data-driven ROI. The two sets of ROI may be compared in terms of voxel intersection, and average intensity and correlation values.

Table 40. Bridge Voxel Synopsis

Experiment Component	*Description*
Index: 10	Bridge Voxels
Comment	What is the relationship of proximal regions as determined by a FCM analysis of activated regions? When these regions are merged by alternate methods, what *TC*s are added to the activated region and what are their characteristics?
Independent Variables	CHAMELEON, FCM, FCMP parameters.
Dependent Variables	Voxels associated with the activated region.
Hypothesis	The voxels that will be added by CHAMELEON and FCMP will exhibit high temporal similarity. It is not obvious that voxels in the direct line of path between the activated regions centre of mass will be included (this reflects the nature of sulci and the contorted gray-white matter interface in the brain).
Main algorithm	FCMP
Benchmarks	EvIdent
Datasets	*in vivo*
Validation	A description of the voxels added to the activated region.

Executive Summary An investigation into the data-driven derivation of global thresholds for fMRI has been described and initial results provided. The bridge voxel thresholds were consistently higher than the pre-defined thresholds and the existence of holes in the bridge voxel ROI was unexpected.

Contribution Definition and demonstration of data-driven ROI generation for fMRI.

Overview Analysis of functional magnetic resonance imaging data is challenging since both inclusive and exclusive modes of analysis are being used. Patterns are being matched while an attempt is also made to discover unknown novelties. Default regions of interest are often defined by analyst-directed thresholding of intensity and correlation

values. This experiment examines a method by which thresholding is directed by the intensity and correlation values of voxels which are structurally significant to the default regions of interest.

Validation Correspondence of ROI with respect to gold standards and region metrics.

Discussion

S05 Results - ROI Composition The experiment consists of the following steps: for each intensity and correlation threshold,

- Pre-defined initial ROI are generated based on the threshold.
- The average intensity, mean square error (MSE) and correlation of the initial ROI are computed using a gold standard activation paradigm.
- Bridge voxels of the initial ROI are computed.
- Data-driven ROI are determined given the intensity and correlation values of the bridge voxels.
- The average intensity, MSE and correlation of the data-driven ROI are computed.

To determine the initial ROI, intensity thresholds of {0.6, 0.7, 0.8, 0.85, 0.9} of the maximum intensity value were used as were correlation thresholds of {0.05, 0.1, 0.15, 0.2, 0.25}. For each threshold, only the largest four-connected region was kept and the bridge voxels discovered by the noted erosion process.

S05 Results The initial ROI generated by intensity and correlation thresholds are seen in

Figs. 73-76. Figures 73 and 74 show typical ROI when a Pearson correlation coefficient of 0.05 and 0.15 is used as a threshold. Figures 75 and 76 show typical ROI when an intensity threshold of 60% and 70% of the maximum intensity value is used. The initial ROI exhibit irregular structure, roughly corresponding to anatomical structure of the visual cortex, and often contain holes of low average intensity or low correlation. A previous study [Alex5] has shown that similarity of intensity values, meaning the temporal features, are not always a good indicator of spatial proximity between voxels. As expected, the erosion process leaves remnants that are much smaller, and that have fewer holes, than the initial ROI.

The following heuristic was found useful in defining bridge voxels in the erosion process: the eroded voxels in the image are designated as candidates for the bridge voxel. The largest two subregions of the shattered initial ROI define a directed line segment, terminating in the centre of mass of each subregion. This line segment is dilated by a structuring element. Finally, the voxels in the intersection of the dilated line segment and the eroded voxels are selected. When more than one voxel is selected, the average intensity or correlation value of the group is used. Groups of bridge voxels were generated in this manner and tended to be small in size and fairly compact (hole-free).

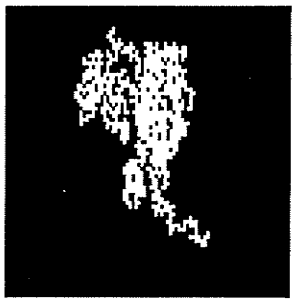


Figure 73. Initial ROI based on correlation value of 0.05.

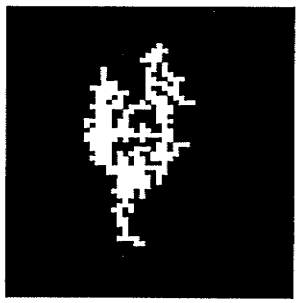


Figure 74. Initial ROI based on correlation value 0.15.

Figure 75. Initial ROI based on intensity threshold at 60% of maximum.

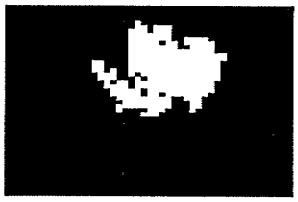


Figure 76. Initial ROI based on intensity threshold at 70% of maximum.

It was noted that some ROI derived from correlation thresholds eroded without leaving any bridge voxels, that is, they eroded without shattering. When this occurred, the centre of mass of the ROI was substituted for the bridge voxel.

Using the bridge voxels, or the centre of mass substitutes, to generate data driven ROI resulted in ROI that were substantially smaller in area than the ROI used to initiate the process. Reconstructed ROI areas were typically 25% or less of the original areas. Also, although some of the reconstructed areas were fairly small, consisting of 7-29 voxels, the regions often exhibited one or more holes (Euler numbers of 0 or more). Since the global threshold derived from the bridge voxels is generally higher than the initial

threshold, the reconstructed ROI are voxel subsets of the initial ROI.

7.11 Overall Results

The results from individual experiments are briefly re-iterated and conclusions drawn.

The experiments noted that:

- For high SNR values, the feature partition weight contributed little additional enhancement in terms of MSE and correlation. As the SNR decreased, there is an optimal weighting pair for the spatial and temporal feature partitions.
- An optimal partition weight was discovered for the dataset Synthetic1.
- FCMP generated ROI that compare well with those generated by FSRG.
- FCMP is shown to be able to detect novel small signals at low SNR levels. The performance of FCMP surpasses that of FCM and EvIdent.
- FCMP shows an intermediate performance (between FCM and CHAMELEON) in characterizing the idiosyncratic nature of individual clusters in a dataset.

It is apparent that FCMP has relevance as an analysis tool at low SNR levels where features exist in distinct conceptual groups that exhibit *cross-information* (meaning that one partition can be used to inform the organizational structure in the other (noisier) partition). A fine sampling in the parameter space may bring out beneficial weight combinations for feature partitions that improve global measurements of accuracy such as MSE and correlation. FCMP is effective in dealing with region growing and small signal detection when the paradigm of the small signal is known in advance, or when its location in time is known in advance. FCMP is partially effective in introducing a

hierarchy of idiosyncratic clusters. This completes the experiment review. A concluding chapter summarizes the thesis and details possible extensions to each experiment.

8. Conclusion

Reasoning draws a conclusion but it does not make the conclusion certain, unless the mind discovers it by the path of experience.
Roger Bacon

In order to appreciate the significance of FCMP, the challenges to fMRI data analysis, introduced in Chapter 1, are reviewed. To address these challenges, a summary of experiment results is provided and other contributions associated with this thesis are mentioned. Finally, the benefits of an abstracted clustering equation are listed and an explanation is provided regarding the incumbent duties associated with algorithm development. It is important to keep in mind that when FCMP was developed, traditional approaches to fMRI clustering analysis used only temporal intensities.

Challenges in fMRI Analysis

Challenges in fMRI data analysis revolve around the detection, extraction, and transformation of information from a large set of complex data. FCMP, or the use of bridge voxel derived thresholds, addressed these challenges. More specifically, analysis enhancement required:

- Incorporating spatial information, or spatial context, into the temporal fuzzy clustering process. FCMP addressed this by integrating distance measures in two feature domains, spatial and temporal. Further, the feature partition

mechanism in FCMP can be generalized to an arbitrary number of feature partitions with partition relationships that can be ordered or weighted and applied to fields outside of fMRI analysis.

- Discovering novel TCs to trigger further analytic investigation. Novelty elicitation is related to minimizing the mathematical data model and requires an inclusive balance to outlier rejection preprocessing. FCMP uses the minimal mathematical model of FCM and shares the sample partitioning power of FCM on general data.
- Reducing the impact of outlier samples on ROI definition or reducing the false positive rate. FCMP combines spatial and temporal distances to include in the ROI regions only TCs that have both high temporal similarity and spatial proximity. This was shown to significantly reduce the false positive rate.
- Ensure that ROI-proximal voxels with correlated TCs be included in the ROI, reduce the false negative rate. The integration of both spatial and temporal distances, as mentioned above, addressed this problem.
- Providing a visual means to determine intrinsic structure in fMRI datasets. The use of the VCV matrix, originally developed for use with FCM, allows visual inspection of TC cluster size and the degree to which samples are members of

multiple clusters.

- It was seen as significant to set fMRI fuzzy clustering analysis within the larger context of fuzzy clustering, including the use and relevance of validation indices. This thesis contains experiments that compare the validation indices commonly used in FCM analysis and show a preference for the Xie-Beni index.
- It is important to have a data-driven method to define ROI in a dataset without a gold standard, or to have an independent standard to complement, or burnish, a tarnished gold standard. The use of bridge voxels provides an independent means to generate intensity and correlation thresholds, which in turn, define data-driven ROI in the dataset.
- To develop an algorithm capable of discerning novelty among a small set of TCs, even when most TC in the dataset are noise or are correlated to the stimulation paradigm. FCMP captures novel intensity signatures by adjusting the distance metrics and weights in the feature partitions. Using a signature probe to trigger use of a different metric facilitates novelty discovery among a small number of samples while still capturing dataset structure. A priori knowledge of spectral properties, occurrence in activated (unactivated) epochs,

or spatio-temporal heuristics can all be integrated in the FCMP objective function using feature partitions.

- Making robust statistical statements about the data. This is important since hemodynamic response models are developing and can be expected to be idiosyncratic (unique for each individual).

FCMP was developed with these challenges in mind and the popular FCM approach was extended. FCMP incorporates spatial information by defining multiple feature partitions. For fMRI analysis, it is common to use one partition for temporal features, intensities and another for spatial features. As a generalization of FCM, advantages and experience accumulated from FCM and its many variants can be extended to FCMP.

FCMP discovers novel TCs for further analysis by calibrating the contribution to objective function optimization from the different feature partitions. Extreme weight values correspond to considering only a single feature partition. Thus the original FCM analysis is preserved as an option in all FCMP analysis. Novelty was defined in terms of results with current industry standard cluster analysis algorithms. The novel TCs discovered by FCMP had significant temporal correlation as well as spatial proximity to regions of interest (e.g. visual cortex).

FCMP makes robust statements about fMRI data, for example: FCMP had a lower false

positive rate than an industry standard for the determination of correlated voxel regions (spatial outliers were rejected; FCMP returned more stable spatial regions of interest than the vanilla FCM for fMRI data. This was true with respect to both changes in SNR values for temporal features and distance for spatial features.

Experiment Result Summary

Experiments were conducted in order to examine conditions for general use of FCMP, evaluate cases of FCMP specialization, quantify FCMP and benchmark comparisons, quantify the utility of FCM cluster validation indices with respect to FCMP, determine the efficacy of preprocessing methods when used in conjunction with FCMP, determine the required resolution of FCMP parameter-space sampling to achieve an optimal performance with respect to a specific dataset. From the series of experiments conducted, it has been shown that:

- A generalization of FCM facilitates adaptation of a popular EDA clustering technique to particular datasets.
- An optimal spatio-temporal weighting exists for synthetic fMRI signals.
- FCMP reduces the false positive rate in the identification of visual cortex regions.
- FCMP maintains critical features of an exploratory data analysis (EDA) technique such as FCM, namely: maintaining implicit inter-sample relations through

membership matrices; the ability to detect novel structure; the ability to represent groups of objects without masking individuality.

Summaries of the knowledge gained from the ten experiments in Chapter 7 are listed in Table 41.

Table 41: Experiment Knowledge Contributions

Index	*Knowledge Contribution*
1	**Concept Partition:** FCMP provides a means for the fMRI analyst to integrate temporal features information with spatial feature information. This is of benefit when one feature domain has been contaminated by noise or side-information about feature inter-relations is available.
2	**Validation Indices:** The XB index is preferable for fMRI data and is able to be used with FCMP.
3	**VCV:** The VCV matrix may be successfully used to make accurate statements about intrinsic data structure in fMRI datasets.
4	**Optimal Partitions:** A set of feature partitions, and a range of feature partition weights, were evaluated that demonstrate a superior ROI with respect to a dataset gold standard.
5	**Induced Hierarchies:** Optimal partitions for FCMP generate accurate sample-hierarchies for fMRI datasets.
6	**ROI:** Optimal partitions FCMP reduce the number of outliers presented to the fMRI analyst and increase the size and spatial continuity of ROI.
7	**Region Growing:** FCMP can be combined with region growing methods to enhance the spatial-temporal associations in the fMRI datasets.
8	**Activated Epochs:** FCMP can be tuned to various temporal and spatial subsets in the data to enhance sensitivity to a pattern critical for overall sample grouping.

Index	*Knowledge Contribution*
9	**Novelty Detection:** FCMP can be tuned to simultaneously pattern matched and discover novelty.
10	**Bridge Voxels:** Erosion of default ROI generates data-driven intensity and correlation thresholds for fMRI datasets that correspond well to gold standard ROI.

Benefits of an Abstracted Clustering Equation

An important contribution of this thesis is the facility of the developed terminology and equations to express a multitude of specializations, or adaptations, of FCM. Several mappings (or specializations) from the generalized cluster analysis formula have been shown which include preprocessing operations (ICA and PCA), express robust metrics, and combine multiple criteria for clustering (detection of small signals during only activated epochs). This algebraic flexibility suggests a new taxonomy for clustering algorithms based on algorithmic degrees of freedom. Such a taxonomy provides mechanisms to explore the degrees of freedom of an algorithm and can be used to suggest future development, to quantify the degree of adaptation of an algorithm, and to define regions of specialty for algorithms. Regions of specialty are datasets for which algorithms are particularly suited.

Contributions and Publications

Original contributions by the author include:

- Generation of synthetic data for investigating salient features of the FCMP algorithm.
- Theoretical derivation of formulas, terminology and investigation of use for the FCMP algorithm.
- Experiments executed with the purpose of investigating and comparing the efficacy of the FCMP algorithm on synthetic and *in vivo* datasets.
- Comparison of a novel clustering algorithm with an industry standard.

Public access to the synthetic and in vivo datasets, as well as the FCMP source code, will be made available at the website www.scopira.org/~alexiuk/fcmp.

Partial results and a discussion of cluster analysis concepts and development of the FCMP algorithm have been presented at various conferences, in particular: the North American Fuzzy Information Processing Society (NAFIPS) in 2003, 2004, 2005, and at the University of Manitoba Graduate Conference (GradCon) 2002, 2003, 2005, Human Centric Computing 2004, Canadian Applied and Industrial Mathematics Society (CAIMS) in 2005, and the Canadian Medical and Biological Engineering Society (CMBES) in 2006. A journal publication on FCMP occurred in Pattern Recognition Letters in 2005. Finally, a book chapter, co-authored with Dr. N. Pizzi, on advances in fuzzy clustering is in press.

8.1 Recommendations

It is recommended that a new taxonomy for fuzzy clustering analysis be based on the

realized possible specializations of the more general FCMP algorithm. Both degrees and types of specialization can be considered. FCMP should be used more extensively as an EDA technique and specifically as an analysis tool for fMRI. As a recent publication [Chua] has shown, FCMP is also valuable as a segmentation tool for biomedical data analysis.

8.2 Future Work

Many opportunities exist for this work to be extended. They include comparing FCMP against more of the many variants of FCM, developing a set of heuristics to initialize the feature partitions (feature indices, weight and metric triple), develop alternate robust metrics for use with topological and topographic properties such as sulci and gyri regions, incorporating knowledge of tissue classes in analysis, expanding the set of feature partitions considered, and extending the class of transformations used in auxiliary data processing.

Extending Proof of Concept: Determine the are optimal weight combinations for more than two partitions with respect to synthetic and *in vivo* fMRI datasets.

Extending Validation Indices: Generate a FCMP-specific validation index.

Extending Visual Cluster Validity: A heuristic method by which the *VCV* matrix may be interpreted as an indicator for the number of inherent clusters in the dataset is

provided. Define a dis-similarity threshold T and use it to convert the *VCV* matrix into a binary matrix where dis-similarity values less than T are set high (1). Denote this matrix *bVCV* (binarized *VCV*). The number of clusters inherent in the data is the minimum number of different sized squares that, when located along the diagonal of *bVCV*, cover all the high elements in *bVCV*. It is to be expected that these covering squares will cover holes (0) in *bVCV* and that singleton high values will exist (causing perhaps unjustified extension of the covering block sizes). However, a process of mathematical morphological closing operations (erosion and dilation) should address this. Covering blocks are computed starting with a 2×2 block template. The template is moved along the *bVCV* diagonal until the template captures all high values in the rows and columns it currently occupies. A covering block is defined at each such location and the high values set low. Repeat with templates of increasing $n\times n$ size until all values are set low. The number of defined blocks is the number of inherent clusters. Note that size information is gained as well.

Extending Optimal Partitions: Test other datasets for the existence of optimal partition weights.

Extending Induced Hierarchy: A metric for comparing hierarchies, and permutations in

a hierarchy, may be developed specifically for fMRI. The metric may consider whether or not the TC in question is part of the gold standard ROI and its temporal similarity. Exploring minimum spanning trees of TC, using spatial distance, temporal correlation, and combinations, may also prove fruitful.

Extending Region Definition: The use of a cluster membership map can be explored with respect to region growing. That is, can cluster assignment inform growth directions, say in terms of replacing correlation values with membership values ? Good measures for the spatial distribution of TCs in the spatial plane need to be determined. In addition, a data-driven approach to determine paradigms in activation studies is required, possibly to burnish tarnished gold standards. Thresholded distance matrix, or the use of ε-insensitive metrics, needs to be explored.

Extending Region Growing: A consideration of the dynamic qualities of region growth should be considered. For example, if the region growing process is prematurely terminated in some use-case, can methods be adapted to include the most critical TC for that case analysis?

Extending Activated Epochs: Determining a advantageous weightings between activated and unactivated epochs, especially in a general stimulation sense, would

increase the robustness of FCMP analysis.

Extending Novelty Detection: Different types of novelty should be investigated and a GMDH type of infrastructure devised to discover the presence of novelties.

Extending Bridge Voxels: Of interest is the case where a threshold derived from bridge voxels is less than that of the initial threshold, as this has been the pattern to date. Also of interest extending the bridge voxel approach to incorporate a spatio-temporal region growing approach [Viva]. Finally, an examination of the interdependence between intensity value and correlation in the bridge voxel process is needed. That is, what effect do predefined intensity thresholds have on the correlation values of the data-driven ROI, and vice versa.

General Extensions

Future work includes the examination of an expanded set of *in vivo* fMRI datasets. A method to deal with integrating activation patterns in different z-planes, for example, which corresponds to integrating feature relations would be beneficial since for some datasets such as Halx, each of the sixteen planes contributes different levels and types of information to the clustering task at hand. Knowledge elucidation or collaborative clustering [Pedr6] is also a field where FCMP can contribute. These studies consider

independent parties pooling information at a high abstract level (to maintain confidentiality regarding individuals or proprietary methods). [Pedr]. Use of FCMP is advantageous in that the trust and feature associated with each information repository, or database, is different and should be integrated in an optimal fashion. Other possible investigations using FCMP include the following preprocessing operations and approaches: principal component analysis (PCA), independent component analysis (ICA), fuzzy inter-quantile, encoding, entropy-based label adjustment, robust clustering, and alternate group or centroid representations.

Glossary

Algorithmic degrees of freedom are the manners in which a basic algorithm can be adapted while remaining fit to execute the purpose of the original algorithm. Limiting an algorithmic degree of freedom results in a modified algorithm which remains faithful to the objective of the original.

Analysis is the resolution of obtained data back to first principles.

An **anomaly** is an irregularity or deviation from the rule.

A **basis** is a set of vectors which span the space.

A **centroid-label** is the index of the cluster to which a sample has maximum membership. See also cluster-homogeneity.

A **class-label** is the true category of a sample.

Cluster analysis is the result of any data categorization algorithm which produces a hard or soft partition of the samples. Partitions of the data into groups corresponds to a clustering of samples to representative proto-types.

Cluster-homogeneity is a measure of diversity in the class labels of samples associated with a single cluster. A cluster that has samples with only a single same class-label will have the highest cluster-homogeneity. If a cluster has only (or mostly) samples of a single class-label, the centroid-label will correspond to a single class label and will be called a (fairly) homogeneous cluster.

The **convergence point** of a clustering algorithm is the membership-centroid pair $\{U^*,V^*\}$ to which successive iterations membership-centroid pairs $\{U_n,V_n\}$ approach. Convergence proceeds as iterations increase. The $\{U^*,V^*\}$ pair is dependent on initialization and is often never determined in practice. Termination criteria are used to approximate $\{U^*,V^*\}$ by $\{U_n,V_n\}$ for sufficiently large n since the convergence process is asymptotic.

In reference to an algorithm, **data-driven** means that algorithm parameters are preferentially determined by the data using intrinsic, usually statistical, properties of the dataset. Extrinsic parameter values from an analyst or mathematical model are deprecated, if they are used at all. The data is explored; residuals with respect to a mathematical model are not deemed significant, since the model itself is in question.

A cluster v_j is said to have the **exclusive membership** of a sample x if x has a membership value of unity for v_j and a membership value of zero v_i, for all $i \neq j$.

Exploratory data analysis is a mode of data analysis which limits the use of mathematical models in order to elicit the data structure most justified by the data itself.

Features are qualitative or quantitative characteristics of an object or sample. Samples, considered as a set of features, order the features to faciliate comparison between samples. In supervised learning, discrimination between samples from different classes is achieved through characteristics of a feature or groups of features. In unsupervised learning, feature characteristics are examined for overall structure between the samples.

A **feature partition** is a triple composed of a set of features indices, a metric associated with the features, and a weighting. A **strict feature partition** is a feature partition in which a feature index may be present in only one set of features indices.

The **feature-space** is the vector space containing all possible combinations of features.

A **framed hypothesis** is a hypothesis regarding a dataset which has been made explicit by an analyst and which is undergoing verification by experiment.

The **generalized recognition ability** of a classifier is its ability to classify previously unseen samples correctly. The classification rate of new samples is used as a predictive error rate for the classifier. The actual value depends on the degree to which the the training and test sets are representative of the actual sample distribution.

A **measure** is a quantity determined by comparison to a standard. It may be a metric or a measure of similarity. The standard may be explicitly or implicitly defined.

The **membership partition** of a dataset with n samples in C clusters is the $n \times C$ matrix that quantifies the degree of membership of each sample in each cluster. Element $U(i,j)$ is the degree to which sample x_i belongs to cluster v_j.

For a non-empty set X, a **metric** is a real function of ordered pairs of elements of X which satisfies [Simm]:

1. $d(x,y) \geq 0$ and $d(x,y) = 0 \Leftrightarrow x = y$
2. $d(x,y) = d(y,x)$ (transitivity)
3. $d(x,y) \leq d(x,z) + d(z,y)$ (triangle inequality)

A system is **modular** if it is constructed with standardized units which facilitate flexibility and variety in use [Merr]. Modular systems contribute to component re-use.

A basic algorithm is, or becomes, **monolithic** when it has become so encumbered with pre-processing and other conceptually distinct functions that the algorithm can no longer be considered modular or basic. Monolithic algorithms resist decomposition and often produces software known as legacy code.

Noise is meaningless interference in a signal transmission or record. The information content of noise is zero. When statistical attributes of the noise are known, actions may be taken to remove or reduce noise effects on the signal of interest. Use of the data for analysis may be precluded entirely by sufficiently high levels of noise.

The **non-exclusive membership** of a sample in a cluster indicates that the sample is a partial membership in one or more clusters. The sample is a member of multiple clusters to a specified degree between zero and unity. The sum of all partital memberships of a sample is unity.

A **norm** is a real-valued non-negative function defined on a vector space and where the function is zero if and only if the vector is zero, the function of the product of a scalar and a vector is equal to the product of the absolute value of the scalar and the function of the vector, and the function of the sum of two vectors is less than or equal to the sum of the functions of the two vectors; *specifically* : the square root of the sum of the squares of the absolute values of the elements of a matrix or of the components of a vector [Merr].

A **novelty** is an unusual appearance. A new, strange or different sample in a collection would constitute a novelty.

Ordering a collection of objects is the process by which each element is assigned a unique index.

Pattern recognition is the body of knowledge dealing with the automated characterization, categorization, and subsequent classification of a collection of samples.

A **poset** is a partially ordered set. A poset is defined by the following properties [Roit]:
$\forall x, y, z \in X$

1. $x \leq x$ (reflexive)
2. if $x \leq y$ and $y \leq x$ then $x = y$ (antisymmetric)
3. if $x \leq y$ and $y \leq z$ then $x \leq z$ (transitive)

A n×n matrix A is **positive definite** if, for any vector x of length n, the vector-matrix product xAx is non-negative.

The **primary membership** of a sample in a cluster is the cluster to which a sample most belongs. It is the maximum membership value for a sample over all the clusters.

Proximal refers to a location near the centre of, or closer to the origin, of a body or structure.

Samples are a collection of objects that are comparable and undergo analysis.

The **sample-space** is the vector span of all the samples in the dataset. The sample-space may be a sub-space of the feature space since not all combinations of features may be realized in the dataset.

Supervised learning is the collection of discrimination techniques used to distinguish between groups of objects when the classes of the objects are known a priori.

A **tarnished gold standard** is an authoritative set of class labels which has become corrupt.

Unsupervised learning is the collection of discrimination techniques used to distinguish between groups of objects when the classes of the objects are not known a priori.

A **voxel** is a volume element.

Acronyms

Acronym	*Expansion*
AFNI	Software for analysis and display of fMRI data.
EDA	Exploratory data analysis.
EROICA	Exploring Regions of Interest with Cluster Analysis.
EvIdent®	EVent IDENTification.
FCM	Fuzzy c-means.
FCMP	FCM with Feature Partitions.
FMRIB	Oxford Centre for Functional Magnetic resonance Imaging of the Brain.
FSL	FMRIB Software Library.
FSRG	Fuzzy Seeded Region Growing.
GA	Genetic algorithm.
GPL	GNU Public Licence.
HCM	Hard c-means.
IBD	Institute for Biodiagnostics (National Research Council).
ICA	Independent component analysis.
LOO	Leave one out.
MedX	Medical Image Processing Application.
MRI	Magnetic resonance imaging.
MSE	Mean square error.
NMR	Nuclear magnetic resonance.
NRC-CNRC	National Research Council.
PCA	Principal component analysis.
PCM	Possibilistic c-means.
rf	Radio-frequency electromagnetic radiation (1-500 Mhz).
ROI	Region of interest.
SNR	Signal to noise ratio [dB].
TC	Time course.
VCV	Visual cluster validity.

Symbols

The following table lists symbols that appear in the text.

Symbol	*Definition*
D	A generalized metric.
K	The number of clusters; the number of classes.
d	A distance metric for a single feature partition.
L_q	The Minkowski norm.
A	A general n×n matrix used in metric definitions.
x,y∈X	Samples in a dataset.
S	Similarity measure. A fuzzy S-norm
T	A fuzzy T-norm
ρ	Pearson correlation
λ	Lagrange multiplier; eigenvector.
N_{*CM}	Membership values classes for fuzzy clustering
M_{*CM}	Membership matrix classes for fuzzy clustering
u_{ij}, U	Membership of a sample $x_i$in cluster v_j; the matrix of all memberships.
v_j∈V	A particularity centroid; the set of centroids.

Appendices

The following topics are covered briefly:

A: Metrics and Measures

B: FCMP Algorithm and Derivation

C: Fuzzy Sets

D: Clustering Algorithm Comparisons

A: Metrics and Measures

Quantification of the distance (similarity) between two objects, $x,y \in X$, is defined by a metric (measure). All metrics have the following properties:

1. A metric is non-negative and has a self-distance of zero: $d(x,y) \geq 0 \wedge d(x,y)=0$ iff $x \equiv y$.
2. A metric exhibits symmetry: $d(x,y)=d(y,x)$.
3. A metric obeys the triangle inequality: $d(x,y)+d(y,z) \geq d(x,z)$.

Similarity measure have the properties:

1. A similarity measure has self-similarity of unity: $S(x,x)=1$.
2. A similarity measure is transitive: $S(x,y)=S(y,x)$
3. A similarity measure obeys the similarity translation for some function *g:*

$$S(x,y),\ S(y,z)>T \rightarrow S(x,z)>g(T).$$

Since a similarity measure is a function of a metric d, $S=f(d)$, the triangle inequality also follows: $S(x,y) + S(y,z) \geq S(x,z)$. Several metrics and similarity measures are discussed.

Metrics

Metrics [Simm], also known as distance functions, map a relation between two samples into $\mathbb{R}_+$. The Minkowski metric, L_q, is commonly used since it is parameterized on q and different values of q have wide applications.

$$(47) \qquad L_q = \left[\sum_{i=1}^{p} (x_i - y_i)^q \right]^{1/q}$$

For q=1, the Minkowski metric is known as the L_1 metric, or Manhattan (city-block) distance,

$$(48) \qquad L_1=\sum_{i=1}^{p}|x_i-y_i|$$

For q=2, the Minkowski metric is known as the L_2 or Euclidean distance.

$$(49) \qquad L_2=\left[\sum_{i=1}^{p}(x_i-y_i)^2\right]^{1/2}$$

In the limit as $q\rightarrow\infty$, the Minkowski equation is known as the Chebychev metric and the metric has enhanced sensitivity to outliers.

$$(50) \qquad \lim_{q\rightarrow\infty} L_q=\lim_{q\rightarrow\infty}\sum_{i=1}^{p}(x_i-y_i)^{1/q}=\mathrm{Max}_i|x_i-y_i|$$

A generalized metric considers the samples in light of additional factors. One type of generalized metric scales the features through multiplication by a *positive definite* p×p matrix A. Such a scaling constitutes a transformation and may in practice be a PCA or ICA transform. The transformation introduces p^2 additional parameters or degrees of freedom. The metric is denoted

$$(51) \qquad \|x\|_A=\sqrt{\langle x,x\rangle_A}=\sqrt{x^T Ax}$$

When the matrix A is the covariance matrix of the dataset, $A=M^{-1}$, the metric is called

the Mahalanobis metric

$$\|x-v\|_{M^{-1}}=\sqrt{(x-v)^T M^{-1}(x-v)} \tag{52}$$

When A is required to be diagonal, A=D, the number of parameters in the classification system is reduced from p^2 to p. Note that only the trace of A is non-zero.

$$\|x-v\|_{D^{-1}}=\sqrt{(x-v)^T D^{-1}(x-v)} \tag{53}$$

A final example metric is the Canberra, which is sensitive to small changes around 0,

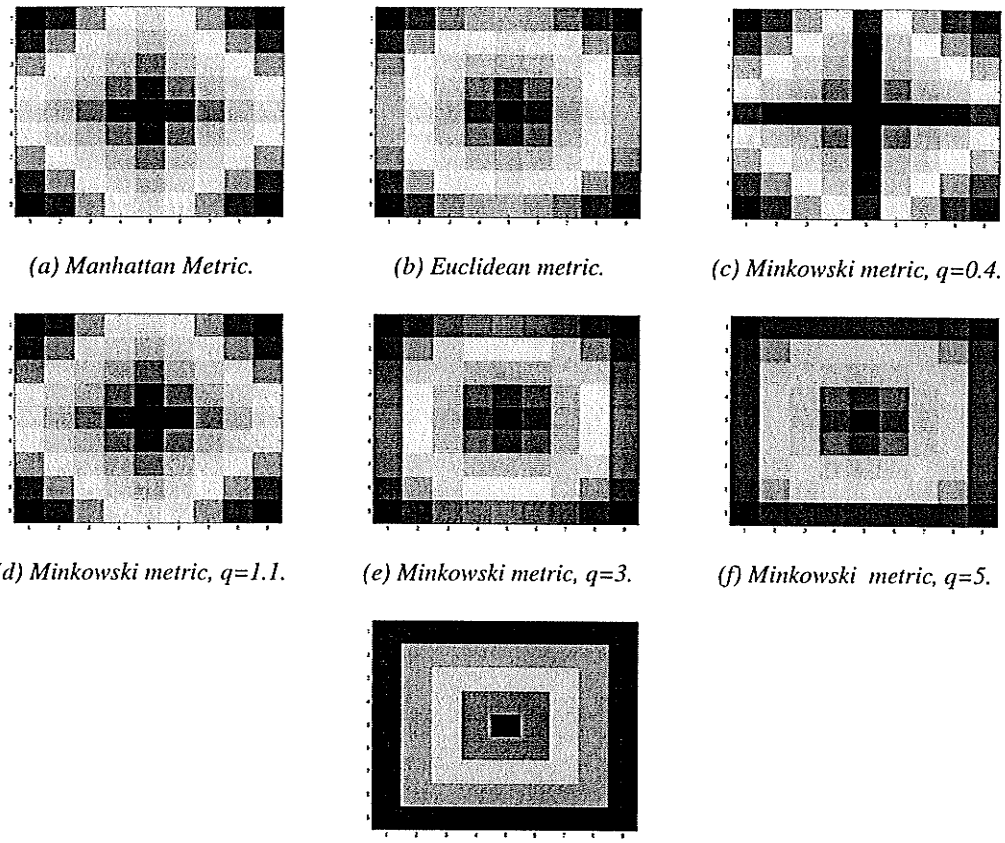
$$\|x\|=\frac{\sum |x_{ki}-x_{kj}|}{|x_{ki}+x_{kj}|} \tag{54}$$

Fig. 77 displays how different metrics compute distance from the origin over a small grid.

Measures

When the definition of similarity is based on a metric, similarity measures induce a partially ordered set (*poset*) [Roit] on the samples. For $x,y,z \in X$, X is a poset if the relation $\leq$ is

1. Reflexive; $x \leq x$.
2. Anti-symmetric; $x \leq y$, $y \leq x$, then $x=y$.
3. Obeys the triangle inequality; $x \leq y$, $y \leq z$, then $x \leq z$.



(a) Manhattan Metric. *(b) Euclidean metric.* *(c) Minkowski metric, q=0.4.*

(d) Minkowski metric, q=1.1. *(e) Minkowski metric, q=3.* *(f) Minkowski metric, q=5.*

(g) Infinity metric.

Figure 77. Distance from the origin, using different metrics.

One similarity measure of interest is the Pearson Correlation ρ for time series

$x = [x_1, ..., x_n]$ and $y = [y_1, ..., y_n]$.

$$\rho(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{55}$$

where $\bar{x} = \frac{1}{n}\sum x_i$ and $\bar{y} = \frac{1}{n}\sum y_i$.

A.1 Binary Similarity Measures

Similarity measures for binary data exhibit a variety of measures for data structure and relations. Binary similarity measures use the cross tabulation, or joint occurrence matrix,

$$(56) \qquad JO = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

where, for sets X and Y, the coefficients represents the number of: items held by both X and Y (a), items held only by X (b), items held only by Y (c), and items held by neither X or Y (d).

Table 42. Binary Metrics

Anderberg $\frac{\frac{a}{a+b}+\frac{a}{a+c}+\frac{d}{c+d}+\frac{d}{b+d}}{4}$	**Anti-Dice** $\frac{a}{a+2(b+c)}$
Gower $\frac{ad}{\sqrt{((a+b)(a+c)(d+b)(d+c))}}$	**Hamann** $\frac{(a+d)-(b+c)}{(a+b+c+d)}$
Kulczynski $\frac{1}{2}\left(\frac{a}{a+b}+\frac{a}{a+c}\right)$	**Matching** $\frac{a+d}{a+b+c+d}$

Pearson binary $\frac{(ad-bc)}{\sqrt{((a+b)(a+c)(d+b)(d+c))}}$	**Rogers and Tanimoto** $\frac{(a+d)}{(a+d)+2(b+c)}$
Sneath and Sokel $\frac{2(a+d)}{2(a+d)+b+c}$	**Yule** $\frac{(ad-bc)}{ad+bc}$
Dice $\frac{2a}{2a+b+c}$	**Jaccard** $\frac{a}{a+b+c}$
Russell and Rao $\frac{a}{a+b+c+d}$	**Ochiai** $\frac{a}{\sqrt{((a+b)(a+c))}}$

B: Derivation of FCMP Equations

The FCM objective function measures the weighted sum of sample-centroid distances over all sample-centroid combinations. It maximizes inter-centroid distances and minimizes intra-cluster variance. The FCM objective function is partially differentiable and the constrained optimization problem is converted to a corresponding unconstrained optimization problem using Lagrange multipliers. Partial derivatives of the objective function define equations which, when algebraically manipulated, relate the centroids, V_t, and memberships, U_t, at successive time steps t. Iteration of the algorithm proceeds until convergence, which is defined as a change in memberships or centroids below a specified threshold. Both initialization of the centroids and the choice of metric can have consequences on FCM convergence[16]. The fuzzy set requirement on the memberships constrains the memberships of a sample in all centroids to sum to unity. This requirement appears in the unconstrained objective functions with the Lagrange multiplier λ. Formally, for each sample x_i, $1 \le i \le N$

$$J_i = \sum_{k=1}^{C} u_{ik}^{m} D_{ik}^{2} - \lambda \sum_{k=1}^{C} \left(u_{ik} - 1 \right) \tag{57}$$

Taking partial derivatives of J_i with respect to u_{ik} and λ and setting the resulting

16 FCM is often executed multiple times to avoid convergence to objective function saddle points, which can be caused by an inauspicious random initialization. Exchanging the Euclidean distance for, say the L_∞, does not guarantee convergence.

equations to 0 generates two equations in two unknowns. This system of equations is then solved to determine the update equations used iteratively in the algorithm.

$$\frac{\partial J_i}{\partial \lambda}=0=\sum_{k=1}^{C} u_{ik}-1 \tag{58}$$

$$\frac{\partial J_i}{\partial u_{ik}}=0=m\,u_{ik}^{(m-1)}D_{ik}^2-\lambda \tag{59}$$

Isolate u_{ik} as a preparatory step. Also, it is common to set m=2 to simplify the algebra.

$$u_{ik}=\left(\frac{\lambda}{mD_{ik}^2}\right)^{\frac{1}{m-1}} \tag{60}$$

Summing u_{ik} over all clusters replaces an unknown value with a known value, unity,

$$\sum_{k=1}^{C} u_{ik}=1=\sum_{k=1}^{C}\frac{\lambda}{2D_{ik}^2} \tag{61}$$

In order to substitute a known value for $\lambda/2$, isolate $\lambda/2$ in the above equation

$$\frac{\lambda}{2}=\frac{1}{\sum_{k=1}^{C}\frac{1}{D_{ik}^2}} \tag{62}$$

and substitute the right hand side into the previous equation for u_{ik}.

$$u_{jk}=\frac{1}{D_{jk}^2}*\frac{1}{\sum_{k=1}^{C}\frac{1}{D_{ik}^2}}=\left[\sum_{k=1}^{C}\frac{D_{jk}^2}{D_{ik}^2}\right]^{-1} \tag{63}$$

The centroid equation that results is a normalized, weighted sum of the samples.

$$v_k = \frac{\sum_{i=1}^{N} u_{ik}^m x_i}{\sum_{i=1}^{N} u_{ik}^m} \tag{64}$$

The possibilistic and various spatio-temporal equations are all derived in a similar manner.

Derivation of FCMP

Given a dataset X, with n samples $x \in X$, and a set of C centroids $k \in K$, where samples indices run from i=1...n, and centroid indices run from j=k...C. For q partitions Ψ, the feature partition weights for a vector $\nu = [\nu_1 \nu_2 \ldots \nu_q]$ which sums to unity

$$\sum_{\psi \in \Psi} \nu_\psi = 1, 0 \le \nu_\psi \le 1, \forall \psi \in \Psi \tag{65}$$

The generalized distance is composed as the sum of weighted distance functions on the partitions

$$D^2(x, v) = \sum_{\psi \in \Psi} \nu_\psi d_\psi^2(x_\psi, v_\psi) \tag{66}$$

where the x_ψ indicates that only the feature indices in the feature partition ψ are considered in the distance calculation.

FCMP Equations Applied to fMRI

When adapting the general FCMP model to a two-partition model, in particular one that uses spatial and temporal domain, the weights may be referred to as ν_S and ν_T respectively. The objective function of this algorithm can now be differentiated and the necessary update equations, for sample memberships and centroids, derived. The FCMP objective function, for spatial and temporal domains using distances ${}^{S}d_{ik}$ and ${}^{T}d_{ik}$ respectively, is:

$$J=\nu_S\sum_{i=1}^{N}\sum_{k=1}^{C}u_{ik}^{mS}d_{ik}^2+\nu_T\sum_{i=1}^{N}\sum_{k=1}^{C}u_{ik}^{mT}d_{ik}^2-\lambda\sum_{i=1}^{N}\sum_{k=1}^{C}\left(u_{ik}-1\right) \tag{67}$$

And for a particular sample x_i, the objective function is:

$$J_i=\nu_S\sum_{k=1}^{C}u_{ik}^{mS}d_{ik}^2+\nu_T\sum_{k=1}^{C}u_{ik}^{mT}d_{ik}^2-\lambda\sum_{k=1}^{C}\left(u_{ik}-1\right) \tag{68}$$

Again, partial derivatives of the objective function generate a system of equations which provide the centroid and membership update equations. These equations are iteratively solved and updated until the algorithm converges. Differentiating the objective function with respect to the constraint on the membership values to sum to unity generates

$$\frac{\partial J_i}{\partial\lambda}=\sum_{k=1}^{C}\left(u_{ik}-1\right)=0 \tag{69}$$

and with respect to the membership value u_{sw} generates

$$\frac{\partial J_i}{\partial u_{sw}} = m \nu_S u_{sw}^{m-1\,S} d_{sw}^2 + m \nu_T u_{sw}^{m-1\,T} d_{sw}^2 - \lambda = 0 \tag{70}$$

Rearranging,

$$u_{sw} = \left[\frac{\lambda}{m \nu_S {}^{S}d_{sw}^2 + m \nu_T {}^{T}d_{sw}^2} \right]^{\frac{1}{m-1}} \tag{71}$$

where m, ν_S, ν_T, ${}^{T}d_{ik}$, ${}^{S}d_{ik}$ are constant with respect to λ and u_{sw}. Factor out λ/m,

$$u_{sw} = \left[\frac{\lambda}{m} \frac{1}{\nu_S {}^{S}d_{sw}^2 + \nu_T {}^{T}d_{sw}^2} \right]^{\frac{1}{m-1}} \tag{72}$$

Let m=2, for simplicity and sum over all clusters

$$\sum_{s=1}^{C} u_{sw} = 1 = \sum_{s=1}^{C} \frac{\lambda}{m} \frac{1}{\nu_S {}^{S}d_{sw}^2 + \nu_T {}^{T}d_{sw}^2} \tag{73}$$

Solve the above for λ/m.

$$\frac{\lambda}{m} = \frac{1}{\sum_{s=1}^{C} \frac{1}{\nu_S {}^{S}d_{sw}^2 + \nu_T {}^{T}d_{sw}^2}} \tag{74}$$

Note that the sample / centroid indices in the above are arbitrary and should be changed before we substitute the solved equation for λ/m into the update equation. Replace indices **sw** with **iw**. Now, substitute our equation for λ/m into the original formula for u_{sw}

$$(75) \qquad u_{sw} = \left[\frac{1}{\sum_{i=1}^{C} \frac{1}{{\nu_S}^S d_{iw}^2 + {\nu_T}^T d_{iw}^2}} \frac{1}{{\nu_S}^S d_{sw}^2 + {\nu_T}^T d_{sw}^2} \right]^{\left(\frac{1}{m-1}\right)}$$

And the final membership update equation is:

$$(76) \qquad u_{sw} = \left[\frac{1}{\sum_{i=1}^{C} \frac{{\nu_S}^S d_{sw}^2 + {\nu_T}^T d_{sw}^2}{{\nu_S}^S d_{iw}^2 + {\nu_T}^T d_{iw}^2}} \right]^{\left(\frac{1}{m-1}\right)}$$

Applicability

FMCP is applicable to the same set of problems as FCM; collections of (partially) labeled or unlabeled data where hypotheses about inherent (or justifiably expedient) global structure is examined. By *justifiably expedient* it is understood that it is not critical that class labels are homogeneous in different clusters but simply that an administrative order is being imposed on the data and that it is to some extent data-driven. FCMP extends this problem domain by adding those problem sets where features are known a priori to have distinct relations and hypotheses about the priority or weighting are being tested in addition to inherent global structure.

FCMP differ from its original form, FCM, in its definition and use of feature partitions (FCM has only one), weightings that rank the importance of feature subsets (FCM has none), and the expectancy that different metrics may be appropriate for various feature

partitions (FCM most often uses the Euclidean metric). FCMP can be modified for future use in several ways. The equations can be extended to take into account covariance matrices for different feature partitions. The algorithmic equations can be expanded to take into account pre-processing operations typical in practice in the particular problem domain. Auxiliary rules of thumb can be developed to optimize FCMP in certain manners to specific types of datasets. Such rules would relate to qualities of the parameter space (the number and types of feature partitions, the relative weights between the partitions, the types of metrics, effects of combining metrics...).

Implementation

FCMP is constrained in its implementation in that it is iterative. Since FCM can be optimized and the core update equations in FCMP are algebraically the same as those in FCM, it follows that FCMP can also be optimized. (Such optimization removes the requirement to update the membership matrix.) This formula needs to be developed for FCMP.

For FCMP, once the partitions have been determined (the number of sets of feature indices and their particular *configuration* (a feature partition *configuration* lists which features are in which partitions)), the parameter space may initially be evaluated over a

rough set of weights for the different partitions. For two partitions, these weights form a pair of intersecting lines in weight space (intersecting where both partitions have value 0.5). For multiple partitions, a grid in weight space is evaluated. These sampling points may then be expanded in regions in the parameter space where the objective function indicates inflection points and optima. Procedures on increasing sampling in these parameter space regions are common in texts on optimization.

FCMP parameters that can be determined a priori are mainly the feature partitions and the relative weights between the partitions, both being determined by conceptual analysis of the problem or expert intuition. Analysis that leads to the initial FCMP parameters and the initial sampling points in the parameter space should be formulated as hypotheses and tested in the accompanying experiments.

FCMP can use data-driven statistics to determine parameters such as the feature partitions (using variance as in the PCA approach, using the mixing matrix for independent components as in the ICA approach). An experiment using FCMP should evaluate FCM on each of the single partitions as well. This incorporates a classical approach to the problem and provides an important parameter-space evaluation for the experiment (the point where all partitions except one have a partition weight of 0). A valid benchmark for FCMP is FCM, since FCMP generalizes FCM. Optimal parameter values for FCMP

are dataset specific, since they exploit feature relationships, though further analysis may define problem sets where valid parameter rules of thumb exist. Current rules of thumb for FCM may be extended for FCMP parameters.

Analysis

FCMP results be interpreted as:

- Membership assignment maps in the different feature-partition spaces. As an example, consider the voxel assignment maps for fMRI data. This considers only the spatial feature partition.
- Trends in a higher dimensional parameter space. Since FCMP has more algorithmic degrees of freedom than FCM, trends in the objective function are evaluated along more axes. Such flexibility may lead to enhanced performance when compared to FCM. At worst, it return a result with the same optimality as FCM. Note that increasing degrees of freedom is not sufficient to provide enhanced performance. Rather, it is the increase of degrees of freedom that mirror feature relationships that provide enhanced analysis. Thus, having more feature partitions is a more significant axes of freedom than having another parameter corresponding, say, to a fuzzy exponent.

A study to discover problematic datasets for FCMP has not yet been completed. Since attempts to optimize the objective function may lead to saddle points, repeated FCMP runs that converge to the saddle points may constitute an anomaly in a collection of FCMP results. Detection of such an anomaly has yet to be examined.

C: Fuzzy Sets

Fuzzy sets can quantify degrees of membership of samples in subsets (clusters). In 1965, Zadeh [Zadeh] published a method of modeling imprecision in mathematical equations. The *fuzzy* prefix is associated with the mathematics of imprecision and is a common thread through fuzzy sets, fuzzy logic, and fuzzy reasoning. Membership of objects in a certain class is considered with the novelty that membership could be partial and non-exclusive. Hard, or classical set membership, has the membership, b, of x in set A as $b_A(x) \in \{0,1\}$, while fuzzy membership has $b_A(x) \in [0,1]$. For example, a single object could be hot to some degree while simultaneously being cold to another degree. In this case *hot* and *cold* are simply two sets to which objects may partially belong. Fuzzy sets have become widely used in analysis, industry and research due to the facility of fuzzy sets to capture plain-language concepts in its mathematics. Fuzzy sets have well defined operations for addition/subtraction, multiplication/division. However, once a fuzzy operation has taken place, it is necessary to convert the fuzzy degrees of membership into a real world quantity before the computation can be acted on. De-fuzzification refers to the mapping from a fuzzy set back into a world where actions must be specified precisely. S-norms and T-norms define the properties of fundamental fuzzy set

operations. An S-norm (or T-conorm) operator is a binary mapping S(°,°) satisfying conditions:

boundary:	S(1,1) = 1, S(a,0) = S(0,a) = a
monotonicity:	S(a,b)≤S(c,d) iff (a≤c) ⋀ (b≤d)
commutativity:	S(a,b) = S(b,a)
associativity:	S(a,S(b,c)) = S(S(a,b),c)

A T-norm (or co-norm) operator is a binary mapping $T(\circ,\circ)$ satisfying conditions:

boundary:	T(0,0)=0, T(a,1) = T(1,a) =a
monotonicity:	T(a,b)≤T(c,d) iff (a≤c) ⋀ (b≤d)
commutativity:	T(a,b) = T(b,a)
associativity:	T(a,T(b,c)) = T(T(a,b),c)

Figures 78-77 show examples of an s-norm, or maximum, intersection, and a t-norm, or minimum.

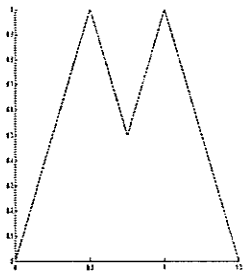


Figure 78. Fuzzy Set Max.

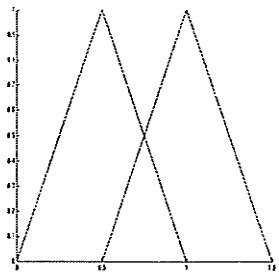


Figure 79. Fuzzy Set Intersection.

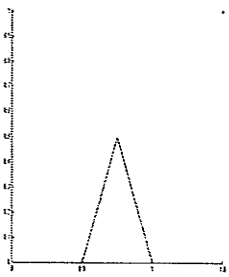


Figure 80. Fuzzy Set Min.

D: Clustering Algorithm Comparisons

The following discussion examines FCM but includes accounts of two closely related algorithms: hard C-means (HCM) and possibilistic C-means (PCM). Both HCM and PCM are variations on FCM incorporating modified definitions of membership. Examining the trio of clustering algorithms, HCM, FCM and PCM, illuminates the novelty of FCMP.

Table 43. HCM, FCM and PCM Objective Functions

Algorithm	*Objective Functions*
HCM	$J(V;X)=\sum_{v\in V}\sum_{x\in X} D_{xv}^2$
FCM	$J_m(U,V;X)=\sum_{v\in V}\sum_{x\in X} u_{xv}^m D_{xv}^2$
PCM	$J_{m,\beta}(U,V;X)=\sum_{v\in V}\sum_{x\in X} u_{xv}^m \|x-v\|_A^2+\sum_{v\in V}\beta_c\sum_{x\in X}(1-u_{xv})^m$

Note that in Table 43 a generalized distance function is used for PCM: $D_{ik}^2=\|x_k-v_i\|_A^2$ where matrix A denotes the covariance matrix. Update equations are listed in Table 44. The equations describing the family of membership matrices associated with the various clustering algorithms are shown in Table 45.

Table 44. Clustering Algorithms Update Equations

Cluster Method	*Membership Update Equation*	*Centroid Update Equation*
HCM	$u_{ik}=1; D_{ik}\le D_{ij}, j\neq i \forall i,k$ $u_{ik}=0;\text{otherwise} \forall i,k$	$v_i=\frac{\sum_{k=1}^{n} u_{ik} x_k}{\sum_{k=1}^{n} u_{ik}}=\frac{\sum_{x_k\in X_i} x_k}{n_i}=\bar{v}_i \forall i$
FCM	$u_{ik}=\left[\sum_{c=1}^{C}\left(\frac{D_{niA}}{D_{ncA}}\right)^{\frac{2}{(m-1)}}\right]^{-1}$	$v_c=\frac{\sum_{n=1}^{N}(u_{nc})^m x_n}{\sum(u_{nc})^m} \forall i$
PCM	$u_{ik}=\left[1+\left(\frac{D_{ik}^2}{w_i}\right)^{\frac{1}{m-1}}\right]^{-1} \forall i,k$	$v_i=\frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m} \forall i$

Table 45. Clustering Algorithms Membership Matrices

Algorithm	*Membership Matrices*
HCM	$M_{HCM}=\{U\in M_{fcn}: U_k\in N_{hc} \forall k\}$
FCM	$M_{FCM}=\{U\in M_{pcn}: U_k\in N_{fc} \forall k\}$
PCM	$M_{PCM}=\{U\in \mathbb{R}^{cn}: U_k\in N_{pc} \forall k; 0<\sum_{k=1}^{n} u_{ik} \forall i\}$

HCM, when augmented by specific merge and split heuristics, is known as ISODATA. PCM has an additional parameter with respect to FCM. The bandwidth, or resolution, parameter β, $\beta_i>0\forall i$, is estimated as

$$\beta_i = Q \frac{\sum_{c=1}^{C} u_{nc}^m \|x_n - v_c\|}{\sum_{n=1}^{N} u_{nc}^m} \qquad (77)$$

where Q is a scaling factor.

The membership values that result from these clustering algorithms are listed in Table 46

Table 46. Clustering Algorithms Membership Values

Algorithm	*Membership Values*	*Notes*
HCM	$N_{HCM} = \{\bar{u} \in N_{fc} : u_i \in [0,1] \forall i\}$ $= \{\bar{e}_1, \bar{e}_2 \ldots \bar{e}_c\}$	Membership in one cluster only.
FCM	$N_{FCM} = \{\bar{u} \in N_{pc} : \sum_{i=1}^{C} u_i = 1\}$	Sum of centroid memberships for each sample is unity.
PCM	$N_{PCM} = \{\bar{u} \in \mathbb{R}^C : y_i \in [0,1] \forall i, u_i > 0 \exists i\}$ $= [0,1]^C - \bar{0}$	Label vector is non-zero for each cluster centre.

Electronic Resources

URL	*Description*
math.york.ca/SCS/Gallery/accent.html	ACCENT
afni.minh.nih.gov/afni	AFNI
medx.sensor.com/products/medx/documentation.html	Medx
scopira.org	Scopira
fil.ion.ucl.ac.uk/spl	SPM
fmrib.ox.ac.uk/fsl/melodic2.html	FSL and Melodic
www.scopira.org/~alexiuk/fcmp	FCMP
www.nrc-cnrc.ca	National Research Council

References

[Alex1] Alexiuk M.D., N.J. Pizzi, "Multilayer perceptron discrimination of software quality in a biomedical data analysis system," in *Proc IEEE Canadian Conference on Electrical and Computer Engineering*, Winnipeg, Canada, May 12-15, 770-775, 2002.

[Alex2] Alexiuk M.D., *Pattern Recognition Techniques as applied to the Classification of Convective Storm Cells*, MSc thesis, University of Manitoba, 1999.

[Alex3] Alexiuk, M.D., N.J. Pizzi, "Robust centroid determination of noisy data using FCM and domain specific partitioning," in *Proc 22nd Intl Conf North American Fuzzy Information Processing Society (NAFIPS)*, July 24-26, Chicago, 233-238, 2003.

[Alex4] Alexiuk M.D., N.J. Pizzi, "Cluster validation indices for fMRI data: Fuzzy c-means with feature partitions versus cluster merging strategies," in *Proceedings of the North American Fuzzy Information Processing Society*, June 27-30, Banff, Canada, 298-301, 2004.

[Alex5] Alexiuk, M.D. N.J. Pizzi, "Fuzzy C-Means with Feature Partitions: A Spatio-temporal Approach to Clustering fMRI data," *Pattern Recognition Letters*, 26, 1039-1046, 2005.

[Alex6] Alexander, M, R. Baumgartner, A.R. Summers, C. Windischberger, M. Klarhoefer, E. Moser, R.L. Somorjai, "A Wavelet-based Method for Improving Signal to Noise Ratio and Contrast in MR Images," *MRI* 18, 169-180, 2000.

[Amari] Amari S., A. Cichocki, H.H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems, 8,* MIT Press, Cambridge, 1996.

[Arei] Areibi S., "The Effect of Clustering and Local Search on Genetic Algorithms," *Recent Advances in Soft Computing*, Leicester, 172-177, 1999.

[Band] Bandyopadhyay S., U. Maulik, "Genetic Clustering for Automatic Evolution of Clusters and Application to Image Classification," *Pattern Recognition* 35(6),1197-1208, 2002.

[Barr] Barra V., J. Boire, "Segmentation of Fat and Muscle From MR Images of the Thigh By a Possibilistic Clustering Algorithm," *Computer Methods and Programs in Biomedicine* 68(3),185-193, 2002.

[Baum1] Baumgartner R., C. Windischberger, E. Moser, "Quanitification in Magnetic Resonance Imaging: Fuzzy clustering vs. Correlation Analysis," *Magnetic Resonance Imaging* 16(2), 115-125, 1998.

[Baum2] Baumgartner R., L. Ryner, W. Richter, R. Summers, M. Jarmasz, R.L. Somorjai, "Comparison of Two Exploratory Data Analysis Methods for fMRI: Fuzzy Clustering vs. Principal Component Analysis," *Magnetic Resonance Imaging* 18, 89-94, 2000.

[Baum3] Baumgartner R., R.L. Somorjai, R. Summers, W. Richter, L. Ryner, "Novelty Indices: Identifiers of Potentially Interesting Time-Courses in Functional MRI Data", *Magnetic Resonance Imaging*, 18, 845-850, 2000.

[Baum4] Baumgartner R., R.L. Somorjai, R. Summers, W. Richter, L. Ryner, "Correlator Beware: Correlation has Limited Selectivity for fMRI Data Analysis", *NeuroImage,* 12, 240-243, 2000.

[Ben] Ben-Ari M., *Principles of Concurrent and Distributed Programming*, New York, Prentice Hall, 1990.

[Berg] Berger G.E., S.J. Wood, C. Pantelis, D. Velakoulis, R.M. Wellard, P.D. McGorry, "Implications of lipid biology for the pathogenesis of schizophrenia," *Australian and New Zealand Journal of Psychiatry,* 35, 355-366, 2002.

[Bezd1] Bezdek J.C., "A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms," *IEEE Trans PAMI,* 2(1), 1-8, 1980.

[Bezd2] Bezdek J.C., R. Hathaway, M.J. Sabin, W.T. Tucker, "Convergence Theory for Fuzzy C-Means: Counterexamples and Repairs," *IEEE Trans SMC,* 17(5), 873-877, 1987.

[Bezd3] Bezdek J.C., J. Keller, R. Krishnapuram, N.R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Kluwer Academic Publishers, Norwell, 1999.

[Bish] Bishop, C.M. (Editor), *Neural Networks and Machine Learning*, Springer-Verlag, Berlin, 1998.

[Bitt] Bittar R.G., J.V. Rosenfeld, G.J. Klug, I.J. Hopkins, A.S. Harvey, "Resective surgery in infants and young children with intractable epilepsy," *J Clin Neurosci,* 9(2), 142-146, 2002.

[Boak] Boakye M., S.C. Huckins, N.M. Szeverenyi, B.I. Taskey, and C.J. Hodge, "*Functional magnetic resonance imaging of somatosensory cortex activity by electrical stimulation of the median nerve or tactile stimulation of the index finger,*" J Neurosurg, 93,774-783, 2000.

[Bobr] Bobrowski L., J.C. Bezdek, "C-Means Clustering with the L-1 and L-infinity Norms", *IEEE Trans SMC*, 545-554, 1991.

[Box] Box, G.E.P., G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Fransisco, 1970.

[Bozi] Bozinov D., J. Rahnenfuhrer, "Unsupervised Technique for Robust Target Separation and Analysis of DNA Microarray Spots Through Adaptive Pixel Clustering," *Bioinformatics*, 18, 5, 747-56, 2002.

[Bran] Brant-Zawadzki M., D. Norman, *MRI of the Brain II: Non-Neoplastic Disease*, Lippincott-Raven Publishers, Philadelphia, 1991.

[Brie1] Briellmann R.S., S.F. Berkovic, A. Syngeniotis, M.A. King, G.D. Jackson, "Seizure-associated hippocampal volume loss: a longitudinal MR-study of temporal lobe epilepsy," *Ann Neurol*, 51, 641-644, 2002.

[Brie2] Briellmann R.S., R.M. Kalnins, S.F. Berkovic, G.D. Jackson, "Hippocampal pathology in refractory TLE: T2-weighted signal change reflects dentate gliosis," *Neurology,* 58, 265-271, 2002.

[Brow] Brown M.A., R.C. Semelka, *MRI: Basic Principles and Applications*, Wiley-Liss, Toronto, 1995.

[Bruz] Bruzzone L., D. Fernandez Prieto, "A Partially unsupervised Cascade Classifier for the Analysis of Multitemporal Remote Sensing Images," *Pattern Recognition Letters*, 23(9), 1063-1071, 2002.

[Bryl] Bryll R., R. Guitierrez-Osuna, F. Quek, "Attribute Bagging: Improving Accuracy of Classifier Ensembles by Using Random Feature Subsets," *Pattern Recognition*, 36(6), 1291-1302, 2003.

[Burn] Burn D. A. , "Designing Effective Statistical Graphs," in C. R. Rao, ed., *Handbook of Statistics*, 9, Chapter 22, Amsterdam, 1993.

[Chat] Chatfield C. *The Analysis of Time Series: An Introduction*, Chapman and Hall, New York, 1996.

[Chen] Cheng L.," TROI - A Training and Research System For Tracing Regions of Interest in Brain Images", Master's thesis, Dartmouth College Dept. of Computer Science, June 2000.

[Chua] Chuang, K-S, H-L Tzeng, S. Chen, J. Wu, T-J Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics* 30, 9-15, 2006.

[Clare] Clare, S., "Functional MRI : Methods and Applications", Ph.D. Thesis, University of Nottingham, 1997.

[Cous] Couso I., S. Montes, P. Gil, "Stochastic Convergence, Uniform Integrability and Convergence in Mean on Fuzzy Measure Spaces," *Fuzzy Sets and Systems,*129(1), 95-104, 2002.

[Cove] Cover, T.M., J.A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

[Cox1] Cox, D.R., H.D. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, London, 1968.

[Cox2] Cox, R.W., "AFNI: Software for the Analysis and Visualization of Functional Magnetic Resonance Neuroimages," *Computers and Biomedical Research* 29, 162-173, 1996.

[Croux] C. Croux, C. Dehon, "Robust Linear Discriminant Analysis using S-estimators," *The Canadian Journal of Statistics*, 29, 473-492, 2001.

[Demi] Demirli K., P. Muthukumaran, "Higher Order Fuzzy System Identification using Subtractive Clustering," *J of Intelligent and Fuzzy Systems - Applications in Engineering and Technology*, 9, 3-4, 129-158, 2000.

[Demk] Demko, A., N.J. Pizzi, R.L. Somorjai, "Scopira: A system for the analysis of biomedical data," in *Proceedings of the 2002 IEEE Canadian Conference on Electrical & Computer Engineering, (CCECE 2002), Winnipeg, Canada, 12 - 15 May 2002* , Vol. 2 (Kinsner, W., Sebak, A. and Ferens, A., Eds.) Piscataway, IEEE, 1093-1098, 2002.

[Devi] Devillez A., P. Billaudel, G.V. Lecolier, "A Fuzzy Hybrid Hierarchical Clustering Method with a New Criterion Able to Find the Optimal Partition," *Fuzzy Sets and Systems,* 128(3), 323-338, 2002.

[Digg] Diggle P.J. *Time Series: A Biostatistical Introduction*, Oxford University Press, London, 1990.

[Dimi] Dimitriadou E., A. Weingessel, K. Hornik, "A Combination Scheme for Fuzzy Clustering," *International Journal of Pattern Recognition and Artificial Intelligence*, 16(7), 901-912, 2002.

[Drebin] Drebin, R.A., D.R. Ney, E.K. Fishman, D. Magid, "Volumetric Rendering of Computed Tomography and Data: Principles and Techniques," *IEEE Computer Graphics and Applications*, 10(2), 24-31, 1990.

[Duann] Duann J.R., T.P. Jung, W.J. Kuo, T.C. Yeh, S. Makeig, J.C. Hsieh, T.J. Sejnowski. Measuring variability of hemodynamics responses in event-related BOLD signals", *3rd International Conference on Independent Component Analysis and Blind Signal Separation*, December 9-12, San Diego, *528-533, 2001.*

[Edwa] Edwards, A. L., *An Introduction to Linear Regression and Correlation.* San Francisco, W. H. Freeman, 1976.

[Elli] Elliott D.F., K.R. Rao, *Fast Transforms, Algorithms, Analyses, Applications*, Academic Press, New York, 1982.

[Ermi] Erminio D., F. Guerrisi, "A Fuzzy Clustering Algorithm Based on the Analogy with Mechanical Physics: MeccFuzz," *Quality and Quantity,* 36, 239-257, 2002.

[Fan] Fan, J., W. Zhen, W. Xie, "Suppressed Fuzzy C-Means Clustering Algorithm," *Pattern Recognition Letters*, 24(9), 1607-1612, 2003.

[Farl] Farlow S.J. (Ed) *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, Marcel Dekker Inc, New York, 1984.

[Ford] Ford J., F. Makedon, T. Steinberg, C.B. Owen, S. Johnson, A.J. Saykin. "Stimulus tracking in functional magnetic resonance imaging (fMRI)", in *ACM Multimedia '98*, 445–454, 1998.

[Fort] Fortuna A. L. Ferrante, P. Lunardi, *Essential Illustrated Neurosurgery*, Springer-Verlag Italia, Milano, 2001.

[Fran] Frances P.H., P. De Bruin, "On Data Transformations and Evidence of Non-linearity," *Computational Statistics and Data Analysis*, 40(3), 621-632, 2002.

[Frist] Friston K.J., A.P. Holmes, K.J. Worsley, J.P. Poline, C.D. Frith, R.S.J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," in *Human Brain Mapping*, 2, 189-210, 1995.

[Fuji] Fujikawa Y., T. Ho, "Outliers, Missing Data and Causation - Cluster Based Algorithms for Dealing with Missing Values," *Lecture Notes in Computer Science,* 2336, 549-554, 2002.

[Gath] Gath I., A.B. Geva, "Unsupervised Optimal Fuzzy Clustering," *IEEE Trans on Pattern Analysis and Machine Intelligence*, 11(7), 773-781, 1989.

[Geye] Geyer A., A. Geyer-Schulz, A. Taudes, "A Fuzzy Time Serise Analyser," in Janko, W.H., M. Roubens, H.-J. Zimmerman (Eds), *Progress in Fuzzy Sets and Systems*, Kluwer Academic Publishers, London, 63-74, 1990.

[Gnit] Gnitecki J. and Z. Moussavi, "The fractality of lung sounds: A comparison of three waveform fractal dimension algorithms," *Journal of Chaos, Solitons and Fractals,* 26 (4), 1065-1072, 2005.

[Gokc] Gokcay E., J.C. Principe, "Information Theoretic Clustering," *IEEE Trans PAMI*, 24, 158-171, 2002.

[Gola] Golay X., S. Kollias, G. Stoll, D. Meier, A. Valavanis, "A New Correlation-Based Fuzzy Logic Clustering Algorithm for FMRI," *Magn Reson Med*, 40(2), 249-60, 1998.

[Gonz] Gonzalez, R.C., R.E. Woods, *Digital Image Processing*, Prentice Hall, Upper Saddle River, 2002.

[Greb] Greblicki, W., M. Pawlak, "Learning to recognize patterns with a probabilistic teacher," *Pattern Recognition*, 12, 159-164, 1980.

[Gust] Gustafson E.E., W.C. Kessel, "Fuzzy clustering with a covariance matrix," in *Proc. IEEE Conf. Decision Contr.*, San Diego, 761-766, 1979.

[Halk] Halkidi M., Y. Batistakis, M. Vazirgiannis, "Cluster Validity Methods: Part 1," *Special Interest Group on Management Data* (SIGMOD), 31(2), 40-45, 2002.

[Hara] Haralick, R. M., K. Shanmugam, I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, 3(1), 610-621, 1973.

[Hark] Harkin L.A., D.N. Bowser, L.M. Dibbens, R. Singh, F. Phillips, R.H. Wallace et al., "Truncation of the GABA(A)-receptor gamma2 subunit in a family with generalized epilepsy with febrile seizures plus," *Am J Hum Genet,* 70(2), 530-6, 2002.

[Hath1] Hathaway R.J., J.C. Bezdek, "Recent Convergence Results for the Fuzzy C-Means Clustering Algorithms," *Journal of Classification*, 237-247, 1988.

[Hath2] Hathaway R.J., J.C. Bezdek, "Visual Cluster Validity for Prototype Generator Clustering Models," *Pattern Recognition Letters*, 24(9), 1573-1569, 2003.

[Hond] Honda K., N. Sugiura, S. Ichihashi, "Web Agents- Collaborative Filtering Using Principal Component Analysis and Fuzzy Clustering," *Lecture Notes in Computer Science* 2198,394-402, 2001.

[Höpp] Höppner F., F. Klawonn, R. Kruse, T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, John Wiley and Sons, Ltd., New York, 1999.

[Hoth] Hothorn T., B. Lausen, "Double-bagging: Combining Classifiers by BootStrap Aggreation," *Pattern Recognition*, 36(6), 1303-1309, 2003.

[Hube] Huber, P. J. "Robust smoothing", in R. L. Launer and G. N. Wilkinson (Eds.), *Robustness in Statistics*, 33-48, New York, Academic Press, 1979.

[Hyva] Hyvarinen A, *Independent Component Analysis: A Neural Networks Approach*, Espoo : Finnish Academy of Technology, 1997.

[Jack1] Jackson GD, W. VanPaesschen, "Hippocampal sclerosis in the MR era," *Epilepsia,* 43 (Suppl 1), 4-10, 2002.

[Jack2] Jackway P.T., "On the Scale-Space Theorem of Chen and Yan," *IEEE Trans Pattern Analysis and Machine Intelligence*, 20(3), 351-352, 1998.

[Jain] Jain, A.K., R.P.W. Duin, J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Trans Pattern Recognition and Machine Intelligence*, 22(1), 4-37, 2000.

[Jarm] Jarmasz, M., R.L. Somorjai, "EROICA: exploring regions of interest with cluster analysis in large functional magnetic resonance imaging data sets," *Concepts in Magnetic Resonance*, 16A(1), 50-62, 2003.

[Just] Justusson B.I., "Median Filtering: Statistical Properties," in *Topics in Applied Physics,* 43, 1981.

[Kary] Karypis, G, E.H. Han, V. Kumar, "CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling," *IEEE Computer*, 32, 8, 68-75, 1999.

[Kim] Kim, T., J.C. Bezdek, R. Hathaway, "Optimality Tests for Fixed Points of the Fuzzy C-Means Algorithm," *Pattern Recognition*, 651-663, 1988.

[Kirb] Kirby M., *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*, John Wiley and Sons, New York, 2001.

[Kita] Kitano H., V. Kumar, C.B. Suttner (Ed.), *Parallel Processing for Artificial Intelligence 2,* Elsevier Science B.V., Amsterdam, 1994.

[Koho] Kohonen, T. "*The self-organizing map*," Proc. of the IEEE, 78(9), 1464-1480, 1990.

[Kole] Kolen J.F., T. Hutcheson, "Reducing the Time Complexity of the Fuzzy C-Means Algorithm," *IEEE Trans Fuzzy Systems,* 10(2), 263-267, 2002.

[Kris] Krishnapuram R., H. Frigui, O. Nasraoui, "Fuzzy and Possibilistic Shell Clustering Algorithms and Their Application to Boundary Detection and Surface Approximation - Part 1," *IEEE Trans Fuzzy Systems,* 3(1), 29-43, 1995.

[Kryza] Kryzanowski W.J., *Principles of Multivariate Analysis: a user's perspective*, Oxford University Press, New York, 1988.

[Kupe] Kuperman V., *Magnetic Resonance Imaging: Physical Principles and Applications*, Academic Press, New York, 2000.

[Laza] Lazarevic A., D. Pokrajac, V. Megalooikonomou, Z. Obradovic, "Distinguishing Among 3-D Distributions for Brain Image Data Classification," in *Proceedings of the 4th International Conference on Neural Networks and Expert Systems in Medicine and Healthcare*, Milos Island, 389-396, 2001.

[Lesk] Leski, J., "Towards a Robust Fuzzy Clustering," *Fuzzy Sets and Systems*, 137(2), 215-233, 2003.

[Li] Li, B., E.B. Martin, A.J. Morris, "On Principal Component Analysis in L1," *Computational Statistics and Data Analysis*, 40(3), 471-474, 2002.

[Liu] Liu, S., J. Lin, "Vector Quantization in DCT Domain Using Fuzzy Possibilistic C-Means Based on Penalized and Compensated Constraints," *Pattern Recognition,* 35(10), 2201-2211, 2002.

[Lowe] Lowe A., R.W. Jones, M.J. Harrison, "Temporal Pattern Matching using Fuzzy Templates," *Journal of Intelligent Information Systems*, 13(1), 27-45, 1999.

[Mart] Martin, R.D., "Robust estimation for time series autoregressions," in *Robustness in Statistics*, R.L. Launer and G. Wilkinson (eds.), New York, Academic Press, 147-176, 1979.

[Mega1] Megalooikonomou V., C. Davatzikos, E. H. Herskovits. "Mining Lesion-Deficit Associations in a Brain Image Database," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, Aug. 15-18, 1999.

[Mega2] Megalooikonomou V., "Evaluating the performance of association mining methods in 3-D medical image databases," *Proceedings of the 2nd SIAM International Conference on Data Mining*, Arlington, VA, 474-494, April 11-13, 2002.

[Mega3] V. Megalooikonomou, H. Dutta, D. Kontos, "Fast and Effective Characterization of 3D Region Data," Proceedings of the *IEEE International Conference on Image Processing (ICIP) 2002*, Rochester, 421-424, September 22-25, 2002.

[Melg] Melgani F., S.B. Serpico, "A Statistical Approach to the Fusion of Spectral and Spatio-temporal Contextual Information for the classification of Remote Sensing Images," *Pattern Recognition Letters*, 23(9), 1053-1061, 2002.

[Miln] Milne RD, A. Syngeniotis, C.D. Jackson, M.C. Corballis, "Mixed lateralization of phonological assembly in dyslexia," *Neurocase,* 8(3), 205-209, 2002.

[Mose] Moser E., R. Baumgartner, M. Barth, C. Windischberger, "Explorative signal processing in functional MR imaging," *Int. J. Imaging Syst. Technol.* 10, 166-176, 1999.

[Ng] Ng, M., J.M. Huang, "Fastmap: A Modified Fastmap Algorithm for Visual Cluster Validation in Data Mining," *Lecture Notes in Computer Science,* 2336, 224-236, 2002.

[Nikias] Nikias, C.L. and A.P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, New Jersey, Prentice-Hall, 1993.

[Park] Park S.H., I.D. Yun, S.U. Lee, "Color Image Segmentation Based on 3-D Clustering: Morphological Approach," *Pattern Recognition*, 1061-1076, 1998.

[Pedr1] Pedrycz W., "Collaborative Fuzzy Clustering," *Pattern Recognition Letters,* 23, 1675-1686, 2002.

[Pedr2] Pedrycz W., L.A. Zadeh, *Fuzzy Sets Engineering*, CRC Press, Inc. Boca Raton, 1995.

[Pedr3] Pedrycz W., G. Vukovich, "Fuzzy Clustering with Supervision," *Pattern Recognition, 37(7)*, 1339-1349, 2004.

[Pedr4] Pedrycz W., G. Vukovich, "Logic Oriented Fuzzy Clustering," *Pattern Recognition Letters,* 23(13), 1515-1527, 2002.

[Pedr5] Pedrycz W., J. Waletzky, "Fuzzy Clustering with Partial Supervision," *IEEE Trans SMC,* 27(5), 787-795, 1997.

[Pedr6] Pedrycz, W. *Knowledge-Based Clustering,* John Wiley, Hoboken, 2005.

[Pela] Pelagatti S., *Structured Development of Parallel Programs*, Taylor & Francis Inc., Bristol, 1998.

[Pizzi1] Pizzi N.J., "Pattern recognition using robust discrimination and fuzzy set theoretic preprocessing", PhD thesis, University of Manitoba, Canada, 1997.

[Pizz2] Pizzi N.J., A.R. Summers, W. Pedrycz, "Software quality prediction using median-adjusted class labels," *Proc Int'l Joint Conf Neural Networks IJCNN'02.* May 12-17. Honolulu, 2405-2409, 2002.

[Pizz3] Pizzi N.J., R. Vivanco, R.L. Somorjai, "EvIdent: A Java-based fMRI data analysis application," *Proc. SPIE,* 3808, 761-770, 1999.

[Pizzi4] Pizzi N.J., A. Demko, R. Vivanco, "Discrimination of software quality in a biomedical data analysis system," *Proc Joint 9th IFSA World Congress and 20th NAFIPS Intl Conf*, Vancouver, Canada, July 25-28, 1702-1707, 2001.

[Pizzi5] Pizzi,N.J., M. Alexiuk, W. Pedrycz, "Stochastic Feature Selection for the Discrimination of Biomedical Spectra," *Proc Int'l Joint Conf Neural Networks IJCNN'05,* July 31-August 4, Montréal, CA, 3029-3033, 2005.

[Pizzi6] Pizzi, N.J., L-P. Choo, J. Mansfield, M. Jackson, W.C. Halliday, H.H. Mantsch, R.L. Somorjai, "Neural Network Classification of infrared spectra of control and Alzheimer's diseased tissue," *Artificial Intelligence in Medicine,* 7, 67-79, 1995.

[Pizzi7] Pizzi, N.J., "Classification of biomedical spectra using stochastic feature selection," *Neural Network World*, 15, 257-268, 2005.

[Pres] Press W.H, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 2002.

[Quin] Quinlan, J. R. Decision trees and instance-based classifiers," in *CRC Handbook of Computer Science and Engineering*. A. B. Tucker, Ed., CRC Press, Boca Raton, 1996.

[Raud] Raudys S., A. Saudargiene, "Structures of the Covariance Matrices in the Classifier Design," *Advances in Pattern Recognition: Joint IAPR International Workshops SSPR'98* and SPR'98, Sydney, Australia, August 11-13, 1998.

[Rein] Reinsel G.C. *Elements of Multivariate Time Series Analysis*, Springer-Verlag, New York, 1997.

[Rieg] Rieger R.H., C.R. Weinberg, "Analysis of Clustered Binary Outcomes Using the Within-Cluster Paired Resampling," *Biometrics,* 58, 332-341, 2002.

[Roit] Roitman J. *Introduction to Modern Set Theory*, John Wiley and Sons, New York, 1990.

[Romes] Romesburg, H. C., *Cluster Analysis for Researchers*, Lulu Press, North Carolina, 2004.

[Rous] Rousseeuw P.J., S. Verboven, "Robust Estimation in Very Small Samples," *Computational Statistics and Data Analysis*, 40(4), 741-758, 2002.

[Royc] Roychowdhury S., W. Pedrycz, "Modeling Temporal Functions with Granular Regression and Fuzzy Rules," *Fuzzy Sets and Systems*, 126(3), 377-387, 2002.

[Royer] Royer J.S., J. Case, *Subrecursive Programming Systems: Complexity and Succinctness*, Birkhauser, Boston, 1994.

[Sark] Sarkar M., "Fuzzy K-means clustering with Missing Values," in S. Bakken (Ed.) Proceedings of American Medical Informatics Association Annual Symposium (AMIA), 588-592, Hanley & Balfus, Inc. Medical Publishers, Philadelphia, 2001.

[Sava] Savage, G.R., M.M. Saling, C.W. Davis, S.F. Berkovic, " Direct and indirect measures of verbal relational memory following anterior temporal lobectomy," *Neuropsychologia,* 40(3),302-316, 2002.

[Schol] Scholkopf B., A.J. Smola, *Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond,* MIT Press, Cambridge, 2001.

[Seli] Selim S.A., M.A. Ismail, "K-Means type algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality," *IEEE Trans PAMI*, 6, 81-87, 1984.

[Seul] Seul M., L. O'Gorman, M.J. Sammon, *Practical Algorithms for Image Analysis: Descriptions, Examples, and Code*, Cambridge University Press, New York, 2001.

[Serr] Serra, J., "Image Analysis and Mathematical Morphology," *Academic Press*, 2, 1988.

[Sgar] Sgarro A., "Possibilistic Information Theory: A Coding Theoretic Approach," *Fuzzy Sets and Systems*, 132(1), 11-32, 2002.

[Shen] Shen L., L. Cheng, F. Teng, F. Makedon, J. Ford, T. Steinberg A.J. Saykin, "A Multimedia System for Tracing and Studying Regions-of-Interest in Brain Images," *IEEE Multimedia Technology and Application Conference*, Irvine, CA, USA, November 7-9, 238-245, 2001.

[Simm] Simmons, G.F., *Introduction to Topology and Modern Analysis*, McGraw-Hill Book Co., New York, 1963.

[Sinc] Sincak P., J. Vascak (Ed.), *Quo Vadis Computational Intelligence ? New Trends and Approaches in Computational Intelligence*, Springer-Verlag, Heidelberg, 2000.

[Smith] Smith, S.M., M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M.De Luca, I. Drobnjak, D.E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews, *"Advances in functional and structural MR image analysis and implementation as FSL," NeuroImage*, 23(S1):208-219, 2004.

[Torr] Torra V. "Learning Weights for the Quasi-weighted Means," *IEEE Trans Fuzzy Systems*, 10(5), 653-666, 2002.

[Tous] Toussaint, G.T., "Bibliography on estimation of misclassification," *IEEE Trans. Inform. Theory*, vol. IT-20, 472 479, 1974.

[Tuke] Tukey J.W., *Exploratory Data Analysis*, Addison-Wesley Publishing Company, Reading, 1977.

[Tyan] Tyan. S.G., "Median Filtering: Deterministic Properties," in *Topics in Applied Physics*, 43, 1981.

[Van] Van Hulle M.M., *Faithful Representations and Topographic Maps: From Distortion- to Information-Based Self-Organization*, Wiley, New York, 2000.

[Viva] Vivanco R., N.J. Pizzi, "Fuzzy region growing of fMRI activation areas," *Medical Imaging 2003: Image Processing*, Sonka M., J.M. Fitzpatrick (eds.), Proc SPIE, 5032, February 15-20, San Diego, USA, 1442-1449, 2003.

[Wall] Wallace R.H., I.E. Scheffer, G. Parasivam, S. Barnett, G.E. Wallace, G.R. Sutherland, S.F. Berkovic, J.C. Mulley, " Generalised epilepsy with febrile seizures plus: mutation of the sodium channel subunit SCN1B," *Neurology*, 58(9), 1426-9, 2002.

[Wang1] Wang D. "Primitive Auditory Segregation Based on Oscillatory Correlation," *Cognitive Science*, 20, 409-456, 1996.

[Wang2] Wang S., H. Jiang, H.A. Lu, "A New Integrated Clustering Algorithm GFC and Switching Regressions", *International J of Pattern Recognition and Artificial Intelligence*, 16, 4, 433-446, 2002.

[Wein] Weintrob D.L., M.M. Saling, S.F. Berkovic, S.U. Berlangieri, D.C. Reutens, "Verbal memory in left temporal lobe epilepsy: Evidence for task-related localization," *Ann Neurol*, 51(4), 442-447, 2002.

[Wei] Wei, W., J.M. Mendel, "Optimality Tests for the Fuzzy C-Means Algorithm," *Pattern Recognition*, 27(11), 1567-1573, 1994.

[Wolf] Wolff K.E., "Concepts in Fuzzy Scaling Theory: Order and Granularity," *Fuzzy Sets and Systems*, 132(1), 63-75, 2002.

[Wool] Woolrich, M.W., T.E.J. Behrens, C.F. Beckmann, S.M. Smith, "Mixture models with adaptive spatial regularisation for segmentation with an application to FMRI data," *IEEE Trans. on Medical Imaging*, 24(1), 1-11, 2005.

[Wors] Worsley, K.J., C. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, A.C. Evans, "A general statistical analysis for fMRI data," *NeuroImage,* 15, 1-15, 2002.

[Wu1] Wu, J., T. Tsai, "Weighted Quasi-Likelihood Estimation Based on Fuzzy Clustering Analysis Method and Dimension Reducing Technique," *Fuzzy Sets and Systems,* 128(3), 353-364, 2002.

[Wu2] Wu, K., M. Yang, "Alternative C-Means Clustering Algorithms," *Pattern Recognition,* 35(10),2267-2278, 2002.

[Xie] L. X. Xie, G. Beni, "Validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(8), 841-847, 1991.

[Zadeh] Zadeh, L.A., "Fuzzy sets", *Inf. Control* 8, 338-353, 1965.