

Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification

Richard A. Derrig
Senior Vice President
Automobile Insurers Bureau
101 Arch Street
Boston, MA 02110
U.S.A.
Telephone: 617-439-4542
Fax: 617-439-6789

Krzysztof M. Ostaszewski
Asset Liability Modeling Director
ITT Hartford
Hartford Life Insurance Companies
P. O. Box 2999
Hartford, CT 06104-3801
Telephone: 203-843-3801
Fax: 203-843-7419

Summary

Applications of fuzzy set theory (FST) to property casualty and life insurance have emerged in the last few years through the work of Lemaire (1990), Cummins and Derrig (1991, 1993) and Ostaszewski (1993). This paper continues that line of research by providing an overview of fuzzy pattern recognition techniques. We utilize them in clustering for risk and claim classification. The classic clustering problem of grouping towns into rating territories (DuMouchel (1983), Conger (1987)) is revisited using these fuzzy methods and 1988-1991 Massachusetts automobile insurance data. The new problem of classifying claims in terms of suspected fraud is also addressed using these same fuzzy methods and data drawn from a study of 1989 bodily injury liability claims in Massachusetts.

Techniques floues de reconnaissance des modèles dans la classification des risques et des demandes d'indemnisation

Richard A. Derrig
Premier vice-président
Bureau des assureurs automobiles
101 Arch Street
Boston, Massachusetts 02110
Etats-Unis
Téléphone : 1 (617) 439-4542
Télécopie : 1 (617) 439-6789

Krzysztof M. Ostaszewski
Directeur de la modélisation actif/passif
ITT Hartford
Groupe des compagnies d'assurances Hartford Life
P.O. Box 2999
Hartford, Connecticut 06104-3801
Téléphone : (203) 843-3801
Télécopie : (203) 843-7419

Résumé

Les applications de la théorie des ensembles flous (*fuzzy set theory*, FST) à l'assurance-vie ont été mises en lumière au cours des dernières années par les travaux de Lemaire (1990), Cummins et Derrig (1991, 1993) et Ostaszewski (1993). Cet article continue ces recherches en offrant un panorama des techniques de reconnaissance des modèles flous. Nous les utilisons pour l'agrégation des classifications de risques et les demandes d'indemnisation. Le problème classique d'agrégation posé par le regroupement de villes en territoires de tarification (DuMouchel (1983), Conger (1987)) est réexaminé à la lumière de ces méthodes floues et de données concernant l'automobile et provenant du Massachusetts, pour la période 1988-1991. Le nouveau problème que pose la classification des demandes d'indemnisation, eu égard à la fraude à l'assurance, est également évoqué à l'aide de ces mêmes méthodes, et de données extraites d'une étude de demandes d'indemnisation concernant des blessures corporelles dans l'Etat du Massachusetts.

1. Introduction

Fuzzy set theory (FST) has been developed since 1965 as an alternative to probability theory in modeling uncertainty. Pattern recognition, or the search for structure in data, provided the early impetus for developing FST because of the fundamental involvement of human perception (Dubois and Prade, 1980) and the inadequacy of standard mathematics to deal with complex and ill-defined systems (Bezdek and Pal, 1992). The formal development began with the paper of Zadeh (1965) introducing the principal concepts of FST. A complete presentation of all aspects of FST is available in Zimmerman (1991).

The first recognition of FST applicability to the ambiguous problem of insurance underwriting is due to DeWit (1982). Lemaire (1990) set out a more extensive agenda for FST in insurance theory, most notably in the financial aspects of the business. Under the auspices of the Society of Actuaries, Ostaszewski (1993) assembled a large number of possible applications of fuzzy set theory in actuarial science. His presentation included such areas as economics of risk, time value of money, individual and collective models, classification, assumptions, conservatism and adjustment. Cummins and Derrig (1991, 1993) complemented that work by exploring applications of fuzzy sets to property-casualty insurance forecasting and pricing problems.

The uniqueness of fuzzy sets lies in modelling forms of uncertainty that cannot be fully understood by the use of probability theory. One such uncertainty is vagueness inherent in human perceptions and descriptions of phenomena. Another area is the decision problem involving increased complexity, especially when it combines a human inability to perform controlled experiments concerning the complexity. Ostaszewski (1993) discusses the meaning of uncertainty from the perspective of fuzzy set theory.

In this work we present a method of fuzzy pattern recognition for risk and claim classification. One specific emphasis will be on Massachusetts private passenger automobile insurance territory definitions. A similar application is made to classifying claims with regard to their suspected fraud content. Dubois and Prade (1980), Bezdek (1981) and Kandel (1982) give overviews of fuzzy techniques in pattern recognition. Zimmerman (1991) and Bezdek and Pal (1992) provide other valuable references on the subject.

Acknowledgement

Krzysztof Ostaszewski has worked on this project at the University of Louisville with financial support from the Actuarial Education and Research Fund. This support from AERF is gratefully acknowledged. We thank Jeff Strong and Robert Roesch of the Automobile Insurers Bureau for invaluable help in programming and performing calculations involved in this project and Julie Jannuzzi for the production of the document.

We describe the concept of a fuzzy set and the mathematical algorithms needed to implement classification using fuzzy techniques in Section 2. Grouping towns in Massachusetts into rating territories for risk classification purposes is viewed as a fuzzy clustering problem because many towns are naturally "between" territories rather than within territories. We also explore the influence of geographical proximity on the resulting fuzzy territories in Section 3. Classification of claims by their suspected fraudulent content is discussed in Section 4. A final section summarizes and provides some alternative and future directions for FST in risk and claim classification problems.

2. Algorithms for Fuzzy Classification

Let X be the universe of discourse; i.e., a certain set whose subsets are of interest to us. A *fuzzy subset* \tilde{E} of X is defined by its *membership function* $\mu_{\tilde{E}}: X \rightarrow [0,1]$. For any $x \in X$, the value $\mu_{\tilde{E}}(x)$ specifies to what degree x belongs to the fuzzy set considered. We write either $(\tilde{E}, \mu_{\tilde{E}})$ for that fuzzy set or \tilde{E} , to distinguish from the standard set notation E . The membership function is a generalization of the characteristic function of an ordinary set. Ordinary sets are termed *crisp sets* in fuzzy set theory. They are considered a special case - a fuzzy set is crisp if, and only if, its membership function does not have fractional values.

Lemaire (1990) and Ostaszewski (1993) have pointed out that insurance risk classification often resorts to rather vague and uncertain methods, or methods that are excessively precise - as in the case of a person who may fail to classify as a preferred risk for life insurance application because his body weight exceeded the stated limit by half a pound. Kandel (1982), writing from a different perspective, says: "In a very fundamental way, the intimate relation between the theory of fuzzy sets and the theory of pattern recognition and classification rests on the fact that most real-world classes are fuzzy in nature." This is exactly the reason why we propose to utilize the methodology of fuzzy clustering in rating territory classification and to extend that method to classifying claims for suspected fraud.

Kandel (1982) classifies various techniques of fuzzy pattern recognition, and distinguishes among techniques evolving around the generation of fuzzy languages. *Syntactic* techniques apply when the pattern sought is related to the formal structure of the language. *Semantic* techniques apply to those producing fuzzy partitions of data sets. Bezdek and Pal (1992) give an excellent presentation of pattern recognition systems and their classification. According to their classification the first choice faced by a pattern recognition system designer is that of process description. They indicate the following choices in this area: syntactic, numerical, contextual, rule-based, hybrid, and fuzzy

process descriptions. Feature analysis is the next design step, in which data (generally given in the form of a data vector containing information about the analyzed objects) may be subjected to pre-processing, displays, and extraction. Next, semantic clustering algorithms, generating actual structures in data, are identified. Finally, the classifier design addresses cluster validity and optimality.

In our work, we use a fuzzy pattern recognition technique given by Bezdek (1981). In the classification of Bezdek and Pal (1992) it can be described as a numerical process description, fuzzy c-means iterative algorithm. As the data we analyze is given in a form of numerical vectors (i.e., vectors in a euclidean space), with a number of clusters sought predetermined, we considered the fuzzy c-means technique most appropriate. Bezdek et al (1987) discuss the convergence of the algorithm.

The task is to divide n *patterns*, where n is a natural number, represented by vectors in a p -dimensional euclidean space

$$x_1, x_2, \dots, x_n,$$

(coordinates of the vectors are known as *features*), into c , $2 \leq c < n$, categorically homogeneous subsets which are called *clusters*. The objects belonging to the same cluster should be similar and the objects in different clusters should be as dissimilar as possible. The number of clusters, c , is known in advance. If the membership of objects in clusters is fractional, then we have fuzzy clusters. The process is called *clustering*.

Any clustering method must answer two fundamental questions: which properties of the data set should be used, and in which way should they be used to identify clusters. Once the algorithm meeting those two conditions is specified, there are, of course, more technical questions, such as whether the algorithm is effective for all possible sets of data, as well as the question of validity of clusters (see Kandel, 1982, and Bezdek and Pal, 1992, for a discussion of this problem).

In risk classification, the purpose is to distinguish among risks that are significantly different for the purposes of rating and underwriting. In claim processing, the purpose is to identify claims suspected of fraud for special processing while routing those non-suspicious claims through normal adjusting channels. Insurance risks and claims are all described here by certain data patterns. The pattern recognition algorithm is given the job of doing the "detective work" of arriving at clusters of similar risks and claims.

Let the data set be

$$X = \{x_1, x_2, \dots, x_n\}.$$

X is assumed to be a finite subset of a p -dimensional euclidean space \mathbb{R}^p . Each

$$x_k = (x_{k1}, x_{k2}, \dots, x_{kp}), \quad k = 1, 2, 3, \dots, n$$

is called a *feature vector*, while each x_{kj} , where $j = 1, 2, \dots, p$ is the j -th feature of the pattern x_k .

A partition of the data set X into clusters is described by the membership functions of the elements of the cluster (note that such a description could also apply to crisp clusters, with the membership function meaning simply the characteristic function). The clusters are denoted by $\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_c$ with the corresponding membership functions $\mu_{S_1}, \mu_{S_2}, \dots, \mu_{S_c}$. In other words, we will construct c clusters that are fuzzy sets.

A $c \times n$ matrix containing the values of the membership functions of the fuzzy clusters

$$\tilde{U} = [\mu_{S_i}(x_k)]_{i=1,2,\dots,c; k=1,2,\dots,n}$$

is a *fuzzy c -partition* if it satisfies the following conditions:

$$\sum_{i=1}^c \mu_{S_i}(x_k) = 1 \text{ for each } k = 1, 2, \dots, n, \quad (1)$$

$$0 \leq \sum_{k=1}^n \mu_{S_i}(x_k) \leq n \text{ for each } i = 1, 2, \dots, c. \quad (2)$$

Condition (1) says that each feature vector x_k has its total membership value of one divided among all clusters, while (2) states that the sum of membership degrees of feature vectors in a given cluster does not exceed the total number of feature vectors.

Given the above definition, let us now present the fuzzy c -means algorithm of Bezdek (1981) also used in Ostaszewski (1993). It consists of four steps; we add a fifth step to make the result operational.

STEP 1:

Choose c , an integer between 2 and n , the number of clusters into which the data is partitioned.

Choose a positive parameter m , and a symmetric, positive-definite $p \times p$ matrix G .

Define the vector norm $\| \cdot \|_G$ given by

$$\begin{aligned} \|x_k - v_i\|_G &= \sqrt{(x_k - v_i)^T G (x_k - v_i)} \\ &= \sqrt{\sum_{j=1}^p \sum_{l=1}^p g_{jl} (x_{kj} - v_{il})^2}. \end{aligned} \quad (3)$$

Choose the initial fuzzy partition

$$\tilde{U}^{(0)} = [\mu_{s_1}^{(0)}(x_k)]_{1 \leq i \leq c, 1 \leq k \leq n}.$$

Choose a parameter $\epsilon > 0$ (this number will tell us when to stop the iteration process). Set the iteration counting parameter l equal to 0.

STEP 2:

Calculate the fuzzy cluster centers $\{v_i^{(l)}\}_{i=1,2,\dots,c}$ given by the following formula:

$$v_i^{(l)} = \frac{\sum_{k=1}^n (\mu_{s_1}^{(l)}(x_k))^m x_k}{\sum_{k=1}^n (\mu_{s_1}^{(l)}(x_k))^m} \quad (4)$$

for $i = 1, 2, \dots, c$.

STEP 3:

Calculate the new partition (i.e., membership matrix)

$$\tilde{U}^{(l+1)} = [\mu_{s_i}^{(l+1)}(x_k)]_{1 \leq i \leq c, 1 \leq k \leq n},$$

where

$$\mu_{S_i}^{(l+1)}(x_k) = \frac{\sqrt[m-1]{\frac{1}{\|x_k - v_i^{(l)}\|_G^2}}}{\sum_{j=1}^c \sqrt[m-1]{\frac{1}{\|x_k - v_j^{(l)}\|_G^2}}} \tag{5}$$

where $i = 1, 2, \dots, c$ and $k = 1, 2, \dots, n$.

If $x_k = v_i^{(l)}$, however the formula (5) cannot be used. In that case we set

$$\mu_{S_i}^{(l+1)}(x_k) = \begin{cases} 1 & \text{if } k = i, \\ 0 & \text{if } k \neq i, \end{cases} \quad i = 1, 2, \dots, c. \tag{6}$$

STEP 4:

By using the natural matrix norm, or the extension of $\| \cdot \|_G$ to the matrix norm, or by choosing a different matrix norm, more suitable to the problem, calculate

$$\Delta = \|\tilde{U}^{(l+1)} - \tilde{U}^{(l)}\|_G$$

If $\Delta > \varepsilon$, repeat steps 2, 3 and 4. Otherwise, stop at some iteration count l^* .

STEP 5:

The final fuzzy matrix \tilde{U}^{l^*} is structured for operational use by means of the normalized α -cut, for some $0 < \alpha < 1$. Quite simply, all membership function values less than α are replaced with zero and the function is renormalized (sums to one) to preserve partition condition (1). For small α , the resulting partition will still be fuzzy; for large α (or *max-cuts* where the largest membership value is set equal to one, all others are zero) the resulting partitions will likely be crisp.

3. Automobile Rating Territories in Massachusetts

As Conger (1987) describes, "In Massachusetts, the past ten years have witnessed the evolution of an increasingly sophisticated system of methodologies for determining

the definitions of rating territories for private passenger automobile insurance. In contrast to territory schemes in other states, which tend to group geographically contiguous towns, these Massachusetts methodologies have had as their goal the grouping of towns with similar expected losses per exposure, regardless of the geographic contiguity or non-contiguity of the grouped towns." Note the deliberate ambiguous use of "similar expected losses," a decidedly fuzzy concept.

The methodology used for territory rating results in a final combined five-coverage pure premium index for each of the 360 towns (or, more precisely, 350 towns and ten areas into which Boston is divided for automobile rating purposes). A complete description of the empirical bayes procedure for determining the biennial individual and combined coverage town indices from four years of data is given in DuMouchel (1983). The indices, which are numbers relatively close to 1 representing expected losses in relation to those of the entire state expected losses, are then ordered and territories are created by partitioning that linear ordering.¹ Since frequent switches from one territory to another are inevitable and undesirable, numerous restrictions on moving towns from one territory to another exist in actual regulatory practice. Once territory clusters are set for a rating year, five individual coverage rates are determined using that single clustering, one which may or may not be appropriate for each coverage, but which is assumed to be equitable overall.

We believe that such difficulties and imprecisions in groupings warrant an investigation of fuzzy clustering. Resulting fuzzy clusters would be much more flexible, as a town belonging partially to two territories could in the final assignment belong to the one of them if regulatory limitations dictate unique assignments of towns to territories. Although stability of territory assignment is desirable and convenient, the system of clustering towns into territories should meet the standard criterion for risk classification - it should be responsive. Towns have an incentive to reduce their relative loss costs by maintaining their roads, safety engineering, and law enforcement, if those actions bring about lower premiums. When the system is not responsive, or slow to respond, the incentives can be diminished or lost.

The pure premium indices are calculated for the following coverages for all 350 towns; Bodily Injury Liability (A-1 & B), Personal Injury Protection (A-2), Property Damage Liability (PDL), Collision, Comprehensive, and Combined. We will use those

¹In general, the partitioning is accomplished by grouping towns within five to six percent intervals on either side of the statewide average index of one.

values as the coordinates of patterns x_k , $k = 1, 2, 3, \dots, 350$, representing the towns in the data space. Note that this implies that we treat the data space as six-dimensional, as the number of parameters used to describe towns is six. In our calculations we will use either the five coverage indices (5-dimensional patterns) or the combined index (1-dimensional patterns) but not both. The data for the 1993 indices for towns in Bristol County are given in Appendix 1. Data for all 350 towns and Boston are available in AIB (1992) or from the authors. We will begin by illustrating the algorithm for a manageable set of towns: the twenty towns of Bristol County, Massachusetts.

The Bristol County Algorithm

The initial clustering for Bristol County is the indicated 1993 territory assignment groupings² relabeled 1 to 5. The initial 5-coverage partition matrix is

$$\tilde{U}^{(0)} = \left[\mu_{S_i}^{(0)}(x_k) \right]_{1 \leq i \leq 5, 1 \leq k \leq 20},$$

where $\mu_{S_i}^{(0)}(x_k)$ represents the membership of town x_k in cluster S_i , and it equals 1 if the town is in the territory, or 0 if it is not.

We also set the stopping parameter $\varepsilon = 0.05$, and $m = 2$. The initial cluster centers are calculated as

$$v_i^{(0)} = \frac{\sum_{k=1}^{20} \left(\mu_{S_i}^{(0)}(x_k) \right)^2 x_k}{\sum_{k=1}^{20} \left(\mu_{S_i}^{(0)}(x_k) \right)^2} \quad (7)$$

for $i = 1, 2, \dots, 5$. We proceed to evaluate the new partition matrix

²For illustrative purposes, the single town of Fairhaven in 1993 Territory 9 is joined with those towns in Territory 8. Fall River is also combined with New Bedford. Actual 1993 Rating Territories are subject to judgmental adjustments and capping and are not always those shown here.

$$\tilde{U}^{(1)} = \left[\mu_{S_i}^{(1)}(x_k) \right]_{1 \leq i \leq 5, 1 \leq k \leq 20}, \quad (8)$$

where

$$\mu_{S_i}^{(1)}(x_k) = \frac{\frac{1}{\sqrt{\sum_{p=1}^5 g_{ip} (x_k)_p - (v_i^{(0)})_p)^2}}}{\left(\sum_{j=1}^5 \frac{1}{\sqrt{\sum_{p=1}^5 g_{jp} (x_k)_p - (v_j^{(0)})_p)^2}} \right)} \quad (9)$$

where the subscript p refers to one of the five pure-premium coordinates of a town, and $i = 1, 2, \dots, 5$, $k = 1, 2, \dots, 20$, and g_{ip} are weights representing the distribution of losses across coverage.³

If $x_k = v_i^{(0)}$, however, formula (6) must be used. In that case we set

$$\mu_{S_i}^{(1)}(x_k) = \begin{cases} 1 & \text{if } k = i, \\ 0 & \text{if } k \neq i, k = 1, 2, \dots, 20, i = 1, 2, \dots, 5. \end{cases}$$

Now we calculate the distance between the old partition matrix $\tilde{U}^{(0)}$ and the new partition matrix $\tilde{U}^{(1)}$, by taking the simple matrix norm

$$\Delta = \sqrt{\sum_{j=1}^5 \sum_{k=1}^{20} (\mu_{S_k}^{(1)}(x_k) - \mu_{S_j}^{(0)}(x_k))^2}. \quad (10)$$

If $\Delta < \varepsilon = 0.05$, the process is stopped. Otherwise, the iterative algorithm continues. The results of the calculation, with an α -cut of 0.2, are presented in Table 1.

³The coverage weight distribution, using 1990 exposures times 4-year pure premiums, is $[(g_{ij}) = (0.2229, 0.1109, 0.2048, 0.3210, 0.1404); (g_{ij}) = 0 \text{ if } i \neq j, 1 \leq i, j \leq 5]$.

<i>Town Name</i>	<i>Initial Cluster</i>	<i>Membership Values</i>					<i>Sum</i>
		μ_{S_1}	μ_{S_2}	μ_{S_3}	μ_{S_4}	μ_{S_5}	
MANSFIELD	1	1	0	0	0	0	1
N. ATTLEBOROUGH	1	1	0	0	0	0	1
DIGHTON	2	0.32	0.22	0.46	0	0	1
REHOBOTH	2	0.40	0.23	0.38	0	0	1
NORTON	2	0.58	0	0.42	0	0	1
FREETOWN	2	0	1	0	0	0	1
BERKLEY	2	0	1	0	0	0	1
RAYNHAM	2	0	0	1	0	0	1
SEEKONK	3	0.25	0	0.43	0.32	0	1
EASTON	3	0	0	1	0	0	1
ATTLEBORO	3	0	0	1	0	0	1
DARTMOUTH	3	0	0	1	0	0	1
SOMERSET	4	0	0	0	1	0	1
SWANSEA	4	0	0	0	1	0	1
TAUNTON	4	0	0	0.37	0.63	0	1
WESTPORT	4	0	0.37	0.30	0.33	0	1
ACUSHNET	4	0	0	0	1	0	1
FAIRHAVEN	4	0	0	0	1	0	1
FALL RIVER	5	0	0	0	0	1	1
NEW BEDFORD	5	0	0	0	0	1	1
Sum		3.54	2.82	6.35	5.29	2	20

NOTE: C-means fuzzy clustering algorithm, with 5-coverage data pattern, 9th iteration stopping parameter $0.0499 < 0.05$, α -cut = 0.2, no geographical variables.

Figures 1 and 2 display the results of the transition from initial territory clusters to final fuzzy clusters.

Figure 1 displays the 20 Bristol County towns grouped into their initial clusters in increasing combined index order. For example, Town 1 (Mansfield) has the lowest combined index value (0.8018) and is in the lowest ranked territory (5), while Town 20 (New Bedford) has the highest index (1.2977) and is in the highest ranked territory (13).

Figure 1

1993 Town Clustering for Bristol County by Combined Index Territory

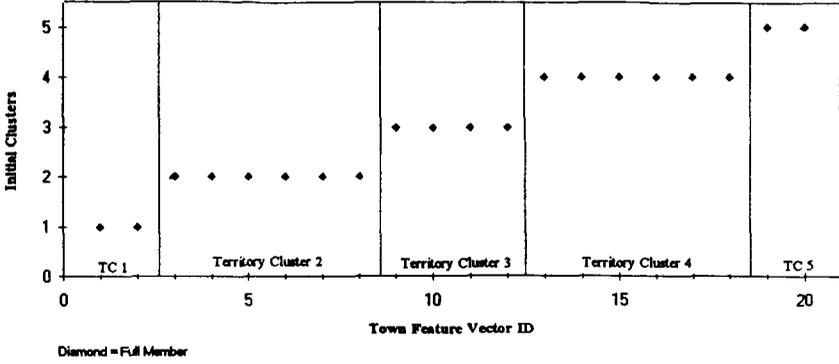
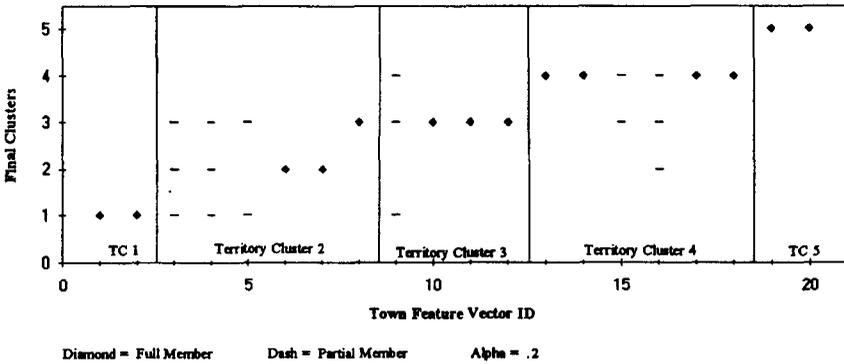


Figure 2 shows the fuzzy clustering results that provide for the incorporation of 5-dimensional data (individual coverages), as well as the fractional assignments (fuzziness) to the clusters. Note that with fuzzy clustering towns tend to become associated with "nearby" clusters as well as with their "home" cluster.

Figure 2

Fuzzy Town Clustering for Bristol County by Five Coverage Indices



One town (Raynham) becomes associated with fuzzy cluster 3 and has little association (less than 0.2) with its original home cluster 2. Another town (Norton) with home cluster 2 splits into fuzzy clusters 1 and 3. These movements are typical of fuzzy clustering results.

Geographical Proximity

We have also performed a calculation adding two more coordinates for each town - its geographical coordinates divided by the coordinates of Nantucket (the division is performed to adjust the scale for the numbers to match the other coordinates, which are all close to 1).⁴ By performing the algorithm on these patterns including geographical coordinates we increase the chance of arriving at clusters which are not only actuarially similar, but relatively close geographically. Recall that other states use geographical proximity as an important factor in determining rating territories.

Table 2
Fuzzy Town Cluster Membership Values with Geographical Variables

Town Name	Initial Cluster	Membership Values					Sum
		μ_{S_1}	μ_{S_2}	μ_{S_3}	μ_{S_4}	μ_{S_5}	
MANSFIELD	1	1	0	0	0	0	1
N. ATTLEBOROUGH	1	1	0	0	0	0	1
DIGHTON	2	0.39	0.20	0.61	0	0	1
REHOBOTH	2	0.38	0.22	0.40	0	0	1
NORTON	2	0.60	0	0.40	0	0	1
FREETOWN	2	0	1	0	0	0	1
BERKLEY	2	0	1	0	0	0	1
RAYNHAM	2	0	0	1	0	0	1
SEEKONK	3	0.25	0	0.44	0.30	0	1
EASTON	3	0	0	1	0	0	1
ATTLEBORO	3	0	0	1	0	0	1
DARTMOUTH	3	0	0	1	0	0	1
SOMERSET	4	0	0	0	1	0	1
SWANSEA	4	0	0	0	1	0	1
TAUNTON	4	0	0	0.38	0.62	0	1
WESTPORT	4	0	0.37	0.28	0.35	0	1
ACUSHNET	4	0	0	0	1	0	1
FAIRHAVEN	4	0	0	0	1	0	1
FALL RIVER	5	0	0	0	0	1	1
NEW BEDFORD	5	0	0	0	0	1	1
Sum		3.62	2.59	6.52	5.27	2	20

NOTE: C-means fuzzy clustering algorithm, with 5-coverage and 2-geographical coordinate data, 11th iteration stopping parameter $0.03 < 0.05$, coverage and geographical data equally weighted.

⁴In this application, the fourth root of the ratio is used to bound the geographical coordinates between 0.49 and 1, making them more comparable in scale to the pure premium indices.

This calculation is performed in the same manner as before, but with 7 variables (pure premium data weighted 50 percent, geographical variables 50 percent). The results are presented in Table 2. Note that we were required to put a full 50 percent weight on the two geographical coordinates in order to produce even the slight differences shown in Table 2. A map of Bristol County is shown in Appendix II.

After an adjustment for geographical location, all towns retained nearly identical membership values within each of the five clusters.⁵ A comparison of the cluster centers shown in Table 3, with and without the geographical variables, reveals how little effect the proximity variables had on the pure premium cluster centers. Either the geographical proximity variable is already accounted for in the 5-coverage pattern or it is relatively weak in relation to the pure premium patterns, at least for this data set.

<i>Final Cluster</i>	<i>A-1&B</i>	<i>A-2</i>	<i>PDL</i>	<i>Collision</i>	<i>Comp</i>	<i>Y-Map</i>	<i>X-Map</i>
1	0.80	0.72	0.84	0.87	0.81	-	-
2	0.83	0.88	0.78	0.91	1.10	-	-
3	0.89	0.85	0.92	0.92	0.94	-	-
4	1.09	0.97	0.97	0.94	0.93	-	-
5	1.33	1.33	1.22	1.13	1.48	-	-
<i>Fuzzy Cluster Centers, Loss & Geographical Data</i>							
1	0.80	0.72	0.84	0.87	0.81	0.88	0.90
2	0.83	0.88	0.78	0.91	1.10	0.92	0.92
3	0.90	0.84	0.92	0.92	0.94	0.90	0.91
4	1.09	0.97	0.96	0.93	0.93	0.93	0.92
5	1.33	1.33	1.22	1.13	1.48	0.94	0.91

As a note of passing interest, we calculated the fuzzy clusters based only on the geographical variables. The resulting territory clusters were crisp at $\alpha=0.2$. Table 4 shows those geographic town/territory clusters.

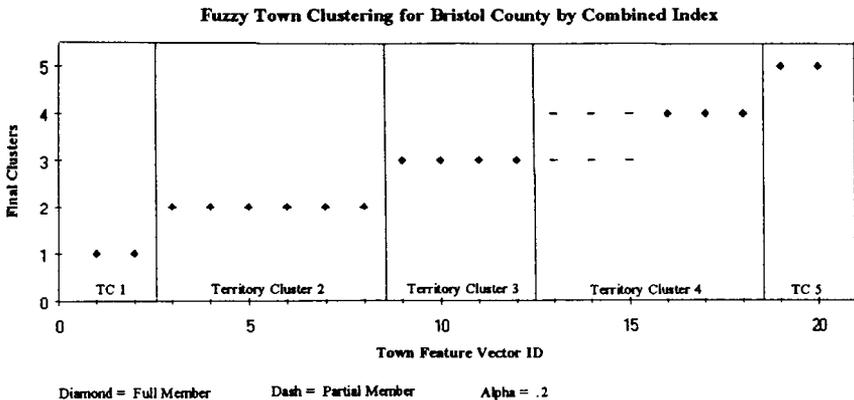
⁵Although the membership values for Dighton appear to be quite different in Tables 1 and 2, the actual pre 0.20-cut values for Dighton are 0.205 and 0.199.

Table 4	
Fuzzy Town Clusters	
Geographical Variables Only	
<i>Final Cluster</i>	<i>Town Name and Initial Cluster Number</i>
1	Mansfield (1), N. Attleborough (1), Norton (2), Easton (3)
2	Berkley (2), Raynham (2) Taunton (4)
3	Dighton (2), Rehoboth (2), Seekonk (3), Attleboro (3)
4	Freetown (2), Somerset (4), Swansea (4), Acushnet (4), Fall River (5)
5	Dartmouth (3), Westport (4), Fairhaven (4), New Bedford (5)
NOTE: C-means fuzzy clustering algorithm, with 5-coverage and 2-geographical coordinate data, 10th iteration stopping parameter $0.03 < 0.05$, α -cut = 0.2.	

Fuzzy Combined Index Clusters

The fuzzy territories shown in Table 1 provide a refinement to the simple territory clusters in use in Massachusetts both by incorporating the 5-dimensional patterns and by allowing fractional (fuzzy) membership. In order to isolate the fuzzy effect, we performed one additional clustering exercise using the single combined index variable. The results are graphically shown in Figure 3. The power of fuzzy clustering to handle those towns that are "between" territories is clear; three towns in the fourth cluster. Somerset (64/36), Swansea (63/37), and Taunton (47/53) have been partially reassigned to Cluster 3 although they were members of territory Cluster 4.

Figure 3



Fuzzy Clustering of 350 Towns

The application of the fuzzy clustering algorithm to the entire set of 350 Massachusetts towns produces results similar to those for Bristol County above. Figure 4 shows the results of the fuzzy algorithm applied to the 1-dimensional combined index variable currently used to set the 20 indicated territory boundaries in Massachusetts.⁶ The current boundary clusters are identified in Figure 4 by the vertical territory cluster (TC) delineations. The ability of the fuzzy clustering to model the "between" territory towns by using partial membership values is clearly evident. The fuzzy clusters also reveal that the current territory boundaries come close, but do not satisfy, the minimum euclidean distance to the centers condition of the algorithm. The fuzzy clustering also reinforces the separation of the first three clusters which are now combined for rating in Massachusetts.

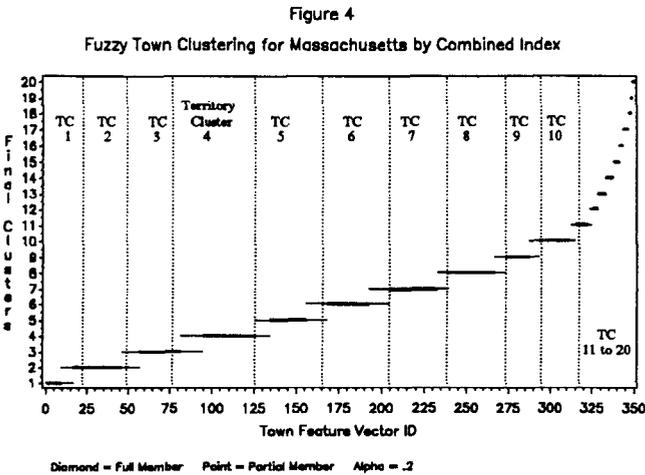
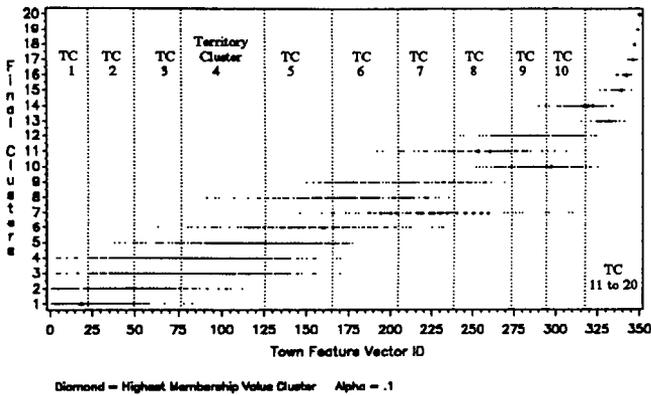


Figure 5 displays the fuzzy clusters using the 5-dimensional feature vectors for all 350 towns. While the fuzzy clusters retain the general pattern of increasing combined index order, the partial membership values reveal much greater similarity among towns that had been grouped together. Fuzzy clusters typically spread across 5 territory

⁶Twenty non-Boston territories are indicated in the AIB filing. Those territories include a split of current territory 1 into three parts (TC1, TC2, TC3) and a split of three towns (Lowell, Somerville, and Springfield) into two additional indicated territories (TC16, TC18).

clusters, while conversely each territory cluster typically spreads across 5 fuzzy clusters for the bulk (325/350) of the towns. The remaining high index towns generally retain their dissimilarity. The spread of the membership values across fuzzy clusters is indicative of the non-monotone nature of the relationship of the 5 coverage relativities (one town can experience high comprehensive pure premiums because of a singular theft problem while another town can experience high bodily injury claims experience because of fraudulent claims).

Figure 5
 Fuzzy Town Clustering for Massachusetts by Five Coverage Indices
 Membership Value Cut at 0.1, No Geographic Variable



One measure of the true extent of the dispersion is the distribution of maximums of town membership values as shown in Table 5.

Table 5					
Fuzzy Town Clustering for Massachusetts					
<i>Town Maximum Membership Values</i>					
Max Value	0-20%	21-40%	41-60%	61-80%	81-100%
# of Towns	88	189	48	16	9

Figure 6 shows the resulting fuzzy clusters using a max cut rather than a 0.1 cut. The graph reveals that the dispersion among fuzzy clusters is fundamental to the algorithm applied to this data.

Figure 6
Fuzzy Town Clustering for Massachusetts by Five Coverage Indices
Membership Value Cut at Max, No Geographic Variable

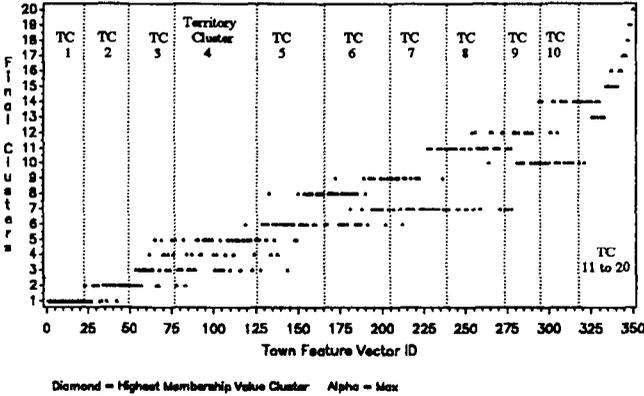
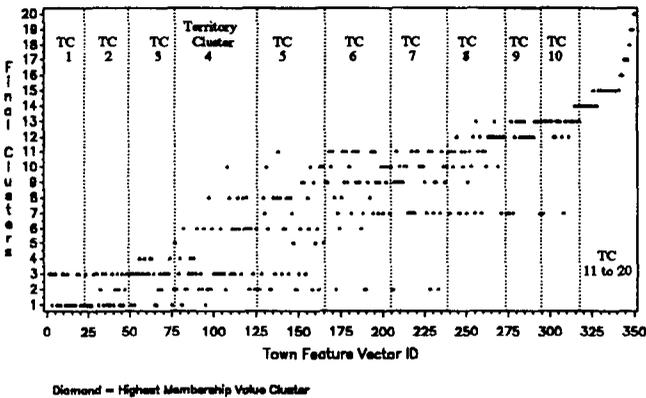


Figure 7 shows the max-cut dispersion using 50 percent weighted geographic variables. Unlike the Bristol County examples above, the geographic variable does affect both the cluster centers and the membership values (compare Figure 6 and 7). This result is due to the fact that low territory cluster towns are situated in rural western Massachusetts as well as in eastern Cape Cod. Those town groups have similar 5-dimensional pure premiums but have very dissimilar geographic variables that affect the final fuzzy clusters (compare final fuzzy clusters 2 and 3 in Figures 6 and 7).

Figure 7
Fuzzy Town Clustering for Massachusetts by Five Coverage Indices
Membership Value Cut at Max, Geographic Variable Included



4. Claim Clustering for Suspicion of Fraud

One vexing problem of property-liability insurance is claim fraud. Individuals and conspiratorial rings of claimants and providers unfortunately can and do manipulate the claim processing system for their own undeserved benefit. When that manipulation violates prevailing law, those actions are called "hard" or criminal fraud. In the criminal sense, we can define *fraud* as a clear and willful attempt proscribed by law to obtain money or value under false pretenses. More often, the information presented in support of an insurance claim bears a vague or ambiguous relationship when measured against the details of a statute or regulation. Unnecessarily prolonged medical treatment and inflated car repair bills to "bury the deductible" are two instances of this so-called "soft" fraud or *build-up*.

Clarke (1990) catalogues the general responses to fraud and build-up in eight Western industrialized nations including the United States. According to Clarke, all insurance claim frauds share the common characteristic of not being self-disclosing. "Their essence is to appear as normal and to be processed and paid in a routine manner. It follows that insurers will only have an idea of the extent of such frauds if they take specific measures to detect them." Clarke's review of anti-fraud activities in the United States is limited to listing various national organizations and databases, such as the National Auto Theft Bureau (now National Insurance Crime Bureau) and the Index System, and to discussing the formation of Special Investigative Units (SIU). Other than the SIU activities, there is no mention of how the claim processing system deters or detects fraud.

Weisberg and Derrig (1991, 1992) have studied automobile bodily injury liability and personal injury protection (PIP) claims in Massachusetts both for their underlying structure and for their fraud and build-up content. They derive a practical definition of *fraud* in the auto context as an attempt to obtain compensation for the alleged consequences of an injury that never happened or was unrelated to the accident. *Build-up* is defined as an attempt on the part of the claimant and/or health provider to inflate the damages for which compensation is being sought. Those studies of 1985/86 and 1989 bodily injury liability claims found that the overall level of suspected or apparent fraud was about 10 percent of the claims, while the apparent build-up level was 35 percent in

the earlier data and 48 percent of the claims in the later data.⁷ The study revealed that although about 10 percent of the claims were apparent frauds, only 1 percent were judged to be candidates for criminal prosecution, under the beyond-a-reasonable-doubt standard, while 2 percent were potentially deniable by the insurance company, under the weaker preponderance-of-the-evidence standard. The use of the terms *suspected* and *apparent* is deliberate and meant to reflect that the numerical estimates of fraud and build-up levels come from judgmental assessments by experienced claim handler coders.

Ideally, one would like to construct a screening device that could be applied by claim adjusters in real time and that would sort incoming claims into various types. Table 6 displays an illustrative categorization of claims and their approximate distribution in Massachusetts BI claims according to the Weisberg-Derrig (1992) study.

<i>Claim Type</i>	<i>Approximate Claim Count Percentage</i>
1 Apparent Fraud Referable for Criminal Investigation	1.0%
2 Apparent Fraud Only	9.1%
3 Apparent Fraud or Build-up	48.3%
4 Valid	51.7%

The difficulty in building a fraud/build-up claim screen lies in the ambiguous (with respect to fraud) information provided in support of the claim and subjective categorizations by the claim adjusting observer. Initial attempts at replacing the subjective fraud/no-fraud judgements by objective facts or patterns of information in the claim file proved unsuccessful. (Weisberg and Derrig, 1991).

A follow-up study clearly demonstrates the unreliability of a single observer's black-and-white, zero-one, fraud/no-fraud judgment and proposes to use numerical scales of suspicion of fraud as a better quantification of the true underlying reality - the coder's perception (Weisberg and Derrig, 1993). A total of 387 Massachusetts claims had been assessed by two independent coders for apparent fraud, one reviewed the BI file and the other reviewed the PIP file. Table 7 shows the wide variation in perception of fraud with only 1.8% of the sample claims perceived as fraudulent by both prior coders.

⁷The Auto Insurance Reform Law of 1988 replaced the previous \$500 medical damage tort threshold with a \$2,000 one, providing greater economic incentive to build-up claims. A short summary of the Weisberg-Derrig 1992 study of post-reform law 1989 claims can be found in "The System Misfired," Best's Review, December, 1992.

<i>Subjective Assessment</i>	<i>Number Coded</i>	<i>Percent of 387 Sample</i>
Apparent Fraud PIP Coder Only	27	7.0%
Apparent Fraud BI Coder Only	28	7.2%
Apparent Fraud Both Coders	7	1.8%
Apparent Fraud Neither Coder	325	84.0%

The fraudulent claim handling problem reflects the fact that suspicion of fraud is a (literally) fuzzy concept and, therefore, a natural candidate for (mathematical) fuzzy set theoretical analysis. We begin by using the fuzzy classification algorithm of Section 2 in an experiment to quantify the level of ambiguity of judgment prevailing in the 1989 study claims.

The 1993 Fraud Study Data

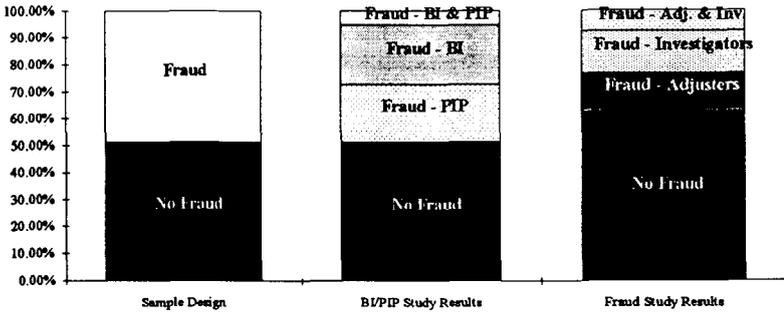
The 62 claims deemed fraudulent by at least one of the coders (Table 7) were supplemented by 65 claims that were representative of the remaining 325 claims in order to study the characteristics of each subsample (Weisberg-Derrig, 1993). From discussion with claim experts, it was determined that claim investigators would have a different perspective on fraud than insurance company adjusters. Thus, parallel coding of the 127 claims was arranged using experienced claims managers to reflect adjuster perceptions and senior personnel from the Insurance Fraud Bureau of Massachusetts to reflect investigator perceptions.

Figures 8 and 9, from the 1993 Weisberg and Derrig fraud study, illustrate graphically the contrasting perceptions among the four independent views of the same set of claim files.⁸

⁸Strictly speaking, the 1993 study coders reviewed the entire combined BI and PIP files while the prior coders, with some exceptions, reviewed one file or the other but not both.

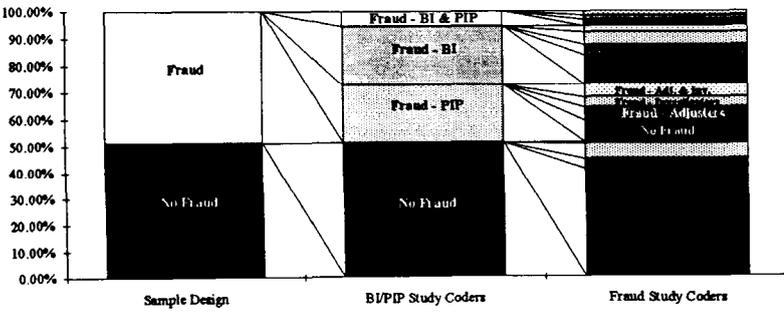
The general equivalence of overall results shown in Figure 8 belies the actual disparity among coders. For example, not one of the 127 claims was coded as fraudulent

FIGURE 8
 1993 MASSACHUSETTS FRAUD STUDY
 Subjective Assessments
 Sample Design vs Coder Results



by all four of the coders. Figure 9 shows the decomposition of adjuster and investigator fraud codings by type of BI and PIP coding. While continued agreement on the BI and PIP no fraud is generally evident, there is a wide disparity of views on the previously coded fraud claims.

FIGURE 9
 1993 MASSACHUSETTS FRAUD STUDY
 Subjective Assessments
 Sample Design vs Coder Results

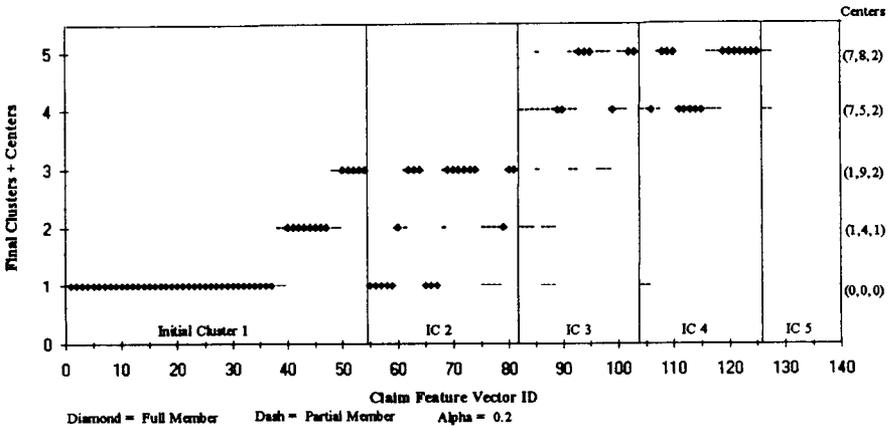


Fuzzy Clustering of Suspicion Measures

Our experiments now focus on the subjective measures of suspicion of fraud that allow for fuzzier identifications than simply zero-one, fraud/no-fraud. Each of the adjusters and investigative coders were asked to record their suspicion of fraud on a zero to ten scale. These responses were grouped into five clusters based upon suspicion levels: None (0), Slight (1-3), Moderate (4-6) Strong (7-9), and Certain (10). Given that four coders had reviewed each claim, a "fraud vote" equal to the number of reviewers who designated the claim fraudulent provided a third suspicion measure, scaled zero to three,⁹ for comparison purposes. The claim data patterns in Study ID order are shown in Appendix III.

Similar to our town/territory exercise of section 3, we begin with the initial one-dimensional clustering of claims by adjuster suspicion levels. We apply the same fuzzy clustering algorithm to the 3-dimensional feature vector of the adjuster suspicion value, the investigator suspicion value and the fraud vote. As with the town/territory exercise, the final fuzzy clustering will be illustrative of both the added dimensions (more information about each claim) and the fractional cluster membership. The fuzzy clusters are shown in Appendix IV.

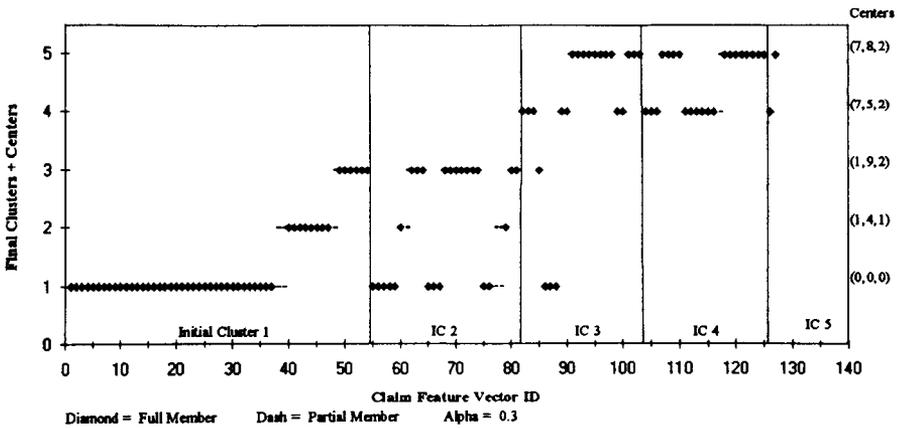
Figure 10
Fuzzy Claim Clustering by All Suspicion Scores



⁹Although the obvious maximum is four, no claim was coded fraudulent by all four coders.

Figure 10 shows graphically how the 127 claims initially clustered by the adjuster suspicion levels,¹⁰ are clustered by the fuzzy algorithm using the investigator and fraud vote data and an α -cut of 0.2. Figure 11 shows the fuzzy clusters at the somewhat higher level α -cut of 0.3, demonstrating the major effect of the multidimensional data pattern rather than the fractional membership values. Only 5 claims retain partial membership at $\alpha=0.3$. In using $\alpha=0.3$, we do, however, mask the uncertainty present in 27 other claims. That uncertainty may be operationally important for claim handling decisions.

Figure 11
Fuzzy Claim Clustering by All Suspicion Scores



It is important to note that the fuzzy algorithm has arranged the claims into clusters of no perceived fraud (center (0,0,0)), little perceived fraud by the adjusters (centers (1,4,1) and (1,9,2)) and strongly perceived fraud by adjusters (center (7,5,2) and center (7,8,2)). Since it is unlikely that an insurer practically can have all claims assessed by four coders, or even assessed in some other multidimensional way, it is especially encouraging that the agreed upon strongly suspicious claims (final clusters 4 and 5) are fully contained within the adjuster suspicion levels 4 and above (initial clusters 3, 4, and 5). This result gives support to the hypothesis that adjuster suspicion levels can serve well to screen suspicious claims despite the inherent ambiguity in observer perceptions.

¹⁰For illustrative purposes in Figures 6 and 7, the claim pattern data shown in Appendices III and IV have been reindexed into 3-dimensional lexicographical order.

5. Summary and Conclusion

This paper has explored the usefulness of fuzzy pattern recognition techniques when applied to insurance classification problems. We considered the ratemaking problem of grouping towns into territories and the operational problem of grouping claims according to their suspected fraud content. Each problem has some degree of ambiguity and judgment in current practices that can be illuminated by fuzzy pattern recognition techniques. The same fuzzy clustering algorithm was applied to Massachusetts data sets representative of the two types of problems to illustrate the universality and flexibility of the fuzzy approach.

As we had pointed out in Section 3, one of the problems of the current automobile insurance rating territories system in Massachusetts is the possibility of frequent switching of certain towns from one territory to another. The fuzzy algorithm is capable of discovering fractional degrees of membership, which may indicate towns "between" territories, towns with transitional behavior, or a need to accommodate the data by increasing the number of clusters. A fuzzy clustering algorithm, by indicating newly developed or increasing fractional membership in another cluster, would provide an early warning of a change occurring. By including a geographical proximity variable, we showed that fuzzy clustering should not change locally (county level) but can change with statewide data.

In the case of claim classification by suspected fraudulent content we found the fuzzy clustering to be an excellent tool in evaluation of the data provided by claim adjusters. As such data is necessarily of subjective nature, modification of it given by the clustering algorithm allows for softening of the sharp distinction (suspicious/not suspicious) which adjusters are forced to make. The inherent ambiguity of suspicious claim labeling process is well modeled and grouped through fuzzy set techniques.

Future research should separate the relative advantages of 5-dimensional versus 1-dimensional town clustering by comparing classical (crisp) cluster techniques to their fuzzy counterparts. Accuracy of resulting town rates should be the determining criteria. Suspicious claim clustering should be tested on multi-dimensional data generated, not by several observers as we do here, but by several components of the claim process, fraud indicators being prime examples.

We conclude that fuzzy clustering is a valuable addition to the methods of risk and claim classification, in the areas studied here, as well in further possible applications.

References

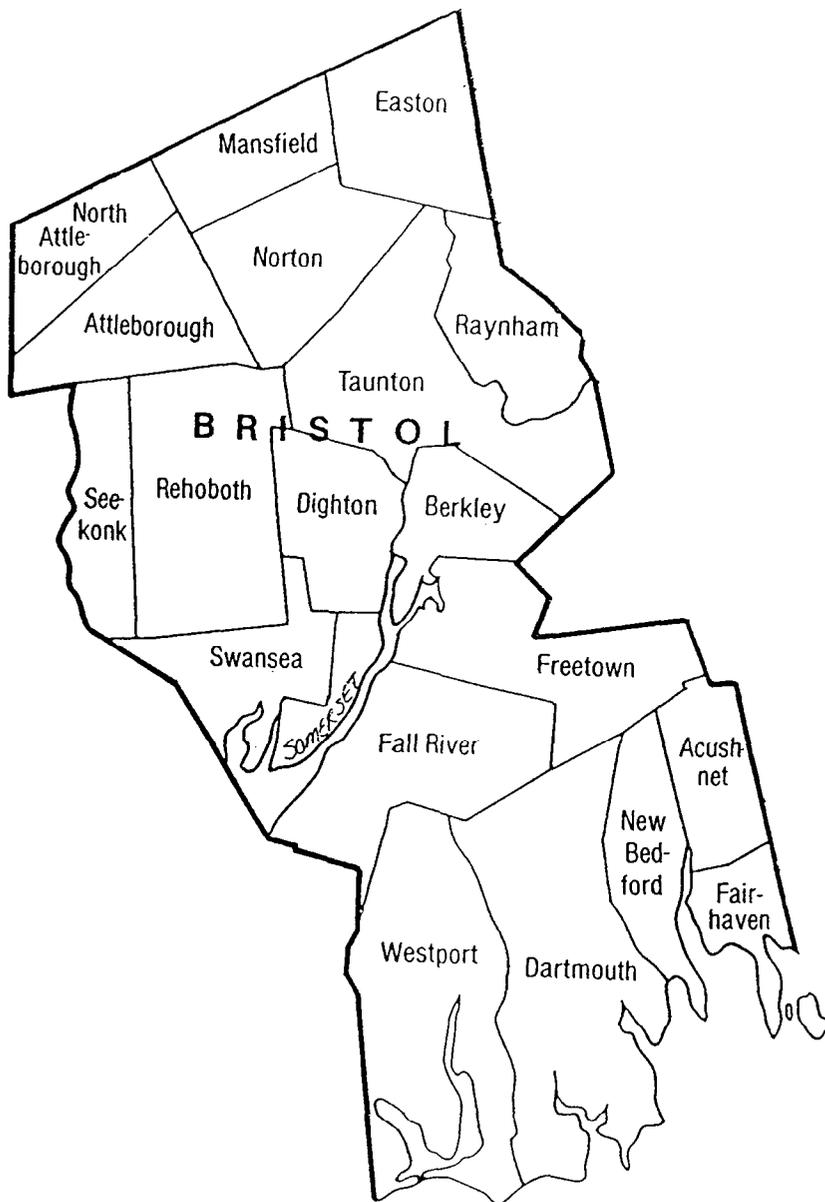
- Automobile Insurers Bureau of Massachusetts (1992), Filing on Automobile Rating Territories, DOI Docket G92-19.
- Bezdek, J.C., (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York, Plenum Press.
- Bezdek, J.C., Hathaway, R.J., Sabin, M.J. and Tucker, W.T., (1987), Convergence Theory for Fuzzy c-Means: Counterexamples and Repairs, *IEEE Trans. Syst. Man., Cybern.*, SMC-17:873-877.
- Bezdek, J.C., and Pal, S.K. (editors), (1992), *Fuzzy Models for Pattern Recognition: Methods That Search for Structure in Data*, IEEE Press, New York.
- Clarke, M., (1990), The Control of Insurance Fraud: A Comparative View, *The British Journal of Criminology*, 30:1:1-23.
- Conger, R.F., (1987), The Construction of Automobile Rating Territories in Massachusetts, *Proceedings of the Casualty Actuarial Society*, 74(141, May):1-74.
- Cummins, J.D., and Derrig, R.A., (1991), Fuzzy Financial Pricing of Property-Liability Insurance, working paper.
- Cummins, J.D., and Derrig, R.A., (1993), Fuzzy Trends in Property-Liability Insurance Claim Costs, *Journal of Risk and Insurance*, forthcoming.
- DeWit, G.W., (1982), Underwriting and Uncertainty, *Insurance: Mathematics and Economics*, 1:277-285.
- Dubois, D. and Prade, H., (1980), *Fuzzy Sets and Systems*, Academic Press, San Diego.
- DuMouchel, W.H. (1983), The 1982 Massachusetts Automobile Insurance Classification Scheme, *The Statistician*, 32:69-81.
- Kandel, A., (1982), *Fuzzy Techniques in Pattern Recognition*, John Wiley and Sons, New York.
- Lemiare, J., (1990), Fuzzy Insurance, *Astin Bulletin*, 20:33-55.
- Ostaszewski, K., (1993), *An Investigation into Possible Applications of Fuzzy Sets Methods in Actuarial Science*, Society of Actuaries, Schaumburg, Illinois.
- Weisberg, H.I. and Derrig, R.A., (1991), Fraud and Automobile Insurance: A Report on the Baseline Study of Bodily Injury Claims in Massachusetts, *Journal of Insurance Regulation*, 9:427-541.
- Weisberg, H.I. and Derrig, R.A., (1992), Massachusetts Automobile Bodily Injury Tort Reform, *Journal of Insurance Regulation*, 10:384-440.
- Weisberg, H.I. and Derrig, R.A., (1993), Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims, *AIB 1994 Fraudulent Claims Payment Filing*, D.O.I. Docket G93-24.
- Zadeh, L.A. (1965), Fuzzy Sets, *Information and Control*, 8:338-353.
- Zimmermann, H.J. (1991), *Fuzzy Set Theory and its Applications*, Second Edition, Kluwer Academic Publishers, Boston, Massachusetts.

Appendix I

Pure Premium and Geographic Indices for Bristol County											
			Final Pure Premium Indices by Coverage								
Town Name	1993 Terr.	Terr. Cluster	Combined	A-1 & B	A-2	PDL	Coll.	Comp.	Y-Map	X-Map	
MANSFIELD	5	1	0.8018	0.7755	0.6849	0.8115	0.8666	0.7699	0.8739	0.9036	
N. ATTLEBOROUGH	5	1	0.8029	0.7630	0.7170	0.8611	0.8166	0.8167	0.8739	0.9036	
DIGHTON	6	2	0.8613	0.8557	0.8207	0.8217	0.8691	0.9171	0.9036	0.9036	
REHOBOTH	6	2	0.8650	0.8756	0.7256	0.8215	0.9450	0.9251	0.9036	0.9036	
NORTON	6	2	0.8705	0.7335	0.7839	0.8852	0.9505	0.7962	0.8739	0.9036	
FREETOWN	6	2	0.8754	0.8330	0.9214	0.7524	0.9235	1.1653	0.9306	0.9219	
BERKLEY	6	2	0.8775	0.8662	0.8640	0.7647	0.8900	1.0825	0.9036	0.9219	
RAYNHAM	6	2	0.8806	0.8408	0.8407	0.8799	0.8786	0.9408	0.8739	0.9219	
SEEKONK	7	3	0.9092	0.9010	0.6828	0.9609	0.9149	0.8236	0.9036	0.9036	
EASTON	7	3	0.9108	0.9172	0.8286	0.9349	0.9443	0.9529	0.8739	0.9219	
ATTLEBORO	7	3	0.9171	0.8493	0.8384	0.9577	0.9219	0.9366	0.8739	0.9036	
DARTMOUTH	7	3	0.9215	1.0212	0.9357	0.9162	0.9089	0.9520	0.9554	0.9219	
SOMERSET	8	4	0.9546	1.1093	0.9566	0.9588	0.8826	0.8316	0.9306	0.9036	
SWANSEA	8	4	0.9551	1.1388	0.9171	0.9552	0.8620	0.9440	0.9306	0.9036	
TAUNTON	8	4	0.9612	1.1071	0.9660	0.9975	0.9278	0.9331	0.9036	0.9219	
WESTPORT	8	4	0.9896	0.9876	0.9602	0.8895	0.9444	1.2262	0.9554	0.9219	
ACUSHNET	8	4	0.9978	1.0150	0.9956	0.9447	0.9770	0.9412	0.9306	0.9391	
FAIRHAVEN	9	4	1.0366	1.0991	1.0444	1.0250	1.0182	0.9875	0.9554	0.9391	
FALL RIVER	12	5	1.2382	1.3310	1.2587	1.2034	1.0752	1.4572	0.9306	0.9036	
NEW BEDFORD	13	5	1.2977	1.3281	1.3958	1.2340	1.1817	1.5076	0.9554	0.9391	

NOTE: The pure premium indices are ratios of empirical bayes estimates of town loss pure premium to statewide averages using 1988-1991 accident year data. AIB Filing, Docket G92-19, May 11, 1992, Exhibit 11. 1993 Territories are those prior to capping. Geographical coordinates are map identifications, $1 \leq y \leq 12$, $1 \leq x \leq 18$, divided by the Nantucket coordinates (12, 18) then exponentiated to 4th root for comparability with the pure premium indices.

Appendix II



Appendix III

Suspicion Measures for 1993 Mass. Fraud Study Claims

<i>Obs</i>	<i>Adj Value</i>	<i>Inv Value</i>	<i>Fraud Vote</i>	<i>Obs</i>	<i>Adj Value</i>	<i>Inv Value</i>	<i>Fraud Vote</i>	<i>Obs</i>	<i>Adj Value</i>	<i>Inv Value</i>	<i>Fraud Vote</i>
1	0	0	1	44	1	0	1	87	0	0	0
2	0	3	1	45	5	8	1	88	0	0	0
3	10	5	3	46	8	6	2	89	0	0	0
4	0	5	2	47	5	8	2	90	5	9	1
5	8	8	3	48	0	0	1	91	0	0	0
6	0	10	2	49	0	3	1	92	0	0	0
7	7	0	1	50	1	6	1	93	0	0	0
8	0	0	1	51	7	8	3	94	5	0	0
9	8	5	2	52	5	8	2	95	6	8	0
10	2	10	2	53	0	0	1	96	0	0	0
11	0	2	1	54	6	9	1	97	0	3	0
12	7	10	3	55	4	5	1	98	5	5	0
13	3	10	2	56	1	0	1	99	1	0	0
14	0	3	2	57	3	9	2	100	4	5	0
15	5	5	2	58	0	2	1	101	8	9	2
16	2	9	1	59	8	5	2	102	0	0	0
17	0	1	1	60	1	8	2	103	2	0	0
18	2	0	1	61	10	9	2	104	8	5	1
19	8	7	2	62	7	0	1	105	0	0	0
20	2	9	3	63	0	0	0	106	0	0	0
21	8	8	2	64	2	9	0	107	0	4	1
22	3	0	1	65	0	0	0	108	0	0	0
23	3	5	1	66	0	0	0	109	1	0	0
24	2	9	3	67	4	5	0	110	0	4	0
25	0	0	2	68	0	0	0	111	2	7	0
26	0	0	1	69	0	1	0	112	0	0	0
27	8	10	3	70	5	0	1	113	3	2	1
28	0	6	2	71	0	0	0	114	0	0	0
29	0	9	1	72	0	0	0	115	0	0	0
30	8	9	3	73	2	0	0	116	5	10	1
31	5	9	3	74	0	8	1	117	4	7	0
32	8	5	2	75	1	5	1	118	8	0	1
33	0	7	1	76	0	0	0	119	0	0	0
34	7	6	2	77	1	8	1	120	5	8	0
35	2	8	3	78	3	2	0	121	1	7	1
36	0	9	2	79	0	9	0	122	0	0	0
37	7	10	2	80	0	0	0	123	6	3	0
38	8	8	2	81	3	0	0	124	6	10	1
39	8	8	3	82	0	0	0	125	0	0	0
40	7	7	3	83	5	0	0	126	0	0	0
41	8	7	3	84	1	0	0	127	0	0	0
42	6	6	2	85	0	5	0				
43	5	8	2	86	0	0	0				

NOTE: *Adjuster Suspicion* and *Investigator Suspicion* value are judgmental assessments of the presence of fraud on a zero (no fraud) to ten scale. *Fraud Vote* value is the sum of coders with judgmental assessment of fraud, zero to 4 scale. Sample of 1989 Massachusetts BI claims overweighted for fraud judgement. Weisberg and Derrig, 1993, Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims, AIB Filing, DOI G93-24, July 2, 1993.

Fuzzy Claim Clustering by Suspicion Scores

α -cut = 0.2

Cluster	Adjuster Score	Initial Cluster <i>U</i> (0)	Final Cluster <i>U</i> (24) [<i>0.2</i> Cut]	Final Center
1	0	1, 2, 4, 6, 8, 11, 14, 17, 25, 26, 28, 29, 33, 36, 49, 50, 54, 59, 64, 66, 67, 69, 70, 72, 73, 75, 77, 80, 81, 83, 86, 87, 88, 89, 90, 92, 93, 94, 97, 98, 103, 106, 107, 108, 109, 111, 113, 115, 116, 120, 123, 126, 127, 128	1, (.35)7, 8, (.48)11, 17, 18, (.71)22, 25, 26, 45, 49, 54, 57, (.48)59, (.35)63, 64, 66, 67, 69, 70, (.40)71, 72, 73, 74, 77, (.45)79, 81, (.74)82, 83, (.42)84, 85, 87, 88, 89, 90, 92, 93, 94, (.42)95, 97, 100, 103, 104, 106, 107, 109, 110, 113, (.42)114, 115, 116, 120, 123, 126, 127, 128	(0, 0, 0)
2	1-3	10, 13, 16, 18, 20, 22, 23, 24, 35, 45, 51, 57, 58, 61, 65, 74, 76, 78, 79, 82, 85, 100, 104, 110, 112, 114, 122	2, 4, (.53)11, 14, (.29)22, 23, (.57)28, (.33)29, (.31)33, 50, (.58)51, (.45)56, (.53)59, (.48)68, (.28)71, 76, (.55)79, (.26)82, (.28)84, 86, (.28)95, 98, (.48)101, 108, 111, (.28)112, (.58)114	(1, 4, 1)
3	4-6	15, 31, 42, 44, 46, 48, 53, 55, 56, 68, 71, 84, 91, 95, 96, 99, 101, 117, 118, 121, 124, 125	6, 10, 13, 16, 20, 24, (.43)28, 29, (.26)31, (.69)33, 35, 36, (.24)46, (.42)51, 58, 61, 65, 75, 78, 80, (.32)91, (.24)96, (.72)112, (.36)117, (.43)118, (.28)121, 122	(1, 9, 2)
4	7-9	5, 7, 9, 12, 19, 21, 27, 30, 32, 34, 37, 38, 39, 40, 41, 47, 52, 60, 63, 102, 105, 119	(.66)3, (.65)7, 9, 15, (.35)19, 32, 34, (.30)40, (.33)41, (.75)42, (.75)47, (.55)56, 60, (.26)62, (.65)63, (.52)68, (.32)71, (.30)84, (.30)95, (.32)96, 99, (.52)101, 105, (.29)118, 119, (.20)121, 124	(7, 5, 2)
5	10	3, 62	(.34)3, 5, 12, (.65)19, 21, 27, 30, (.74)31, 37, 38, 39, (.70)40, (.67)41, (.25)42, 44, (.53)46, (.25)47, 48, 52, 53, 55, 62, (.68)91, (.68)96, 102, (.64)117, (.28)118, (.45)121, 125	(7, 8, 2)

NOTE: C - means fuzzy clustering algorithm, with 3 suspicion measure data patterns for 127 study claims (Appendix III), 24th iteration stopping parameter $0.049 < 0.05$, α - cut = 0.2.

