© *The Journal of Risk and Insurance*, 1995 Vol. 62, No. 3, 447-482

Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification

Richard A. Derrig
Krzysztof M. Ostaszewski

ABSTRACT

Applications of fuzzy set theory to property-liability and life insurance have emerged in the last few years through the work of Lemaire (1990), Cummins and Derrig (1993, 1994), and Ostaszewski (1993). This article continues that line of research by providing an overview of fuzzy pattern recognition techniques and using them in clustering for risk and claims classification. The classic clustering problem of grouping towns into rating territories (DuMouchel, 1983; Conger, 1987) is revisited using these fuzzy methods and 1987 through 1990 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.

Introduction

In 1961, Ellsberg presented the following paradox. An experiment was designed with two urns, each containing 100 balls, of which the first one was known to contain 50 red balls and 50 black balls, while no further information was given about the contents of the other urn. If asked to bet on the color of a ball drawn from one of the urns, most people were found indifferent as to which color they would choose no matter whether the ball was drawn from the first or the second urn. On the other hand, Ellsberg found that if people were asked which urn they would prefer to use for betting on either color, they

---

Richard A. Derrig is Senior Vice President of the Automobile Insurers Bureau of Massachusetts and Vice President–Research for the Insurance Fraud Bureau of Massachusetts. Krzysztof M. Ostaszewski is Associate Professor of Mathematics and Actuarial Program Director at the University of Louisville.

Krzysztof Ostaszewski has worked on this project at the University of Louisville with financial support from the Actuarial Education and Research Fund, and this support from AERF is gratefully acknowledged. The authors thank Jeff Strong and Robert Roesch of the Automobile Insurers Bureau for invaluable help in programming and performing calculations involved in this project, Herbert I. Weisberg for suggesting the fuzzy clustering of fraud assessment data, Ruy Cardoso for helpful comments on an early draft, Julie Jannuzzi for production of the document, and one anonymous reviewer.

consistently favored the first urn (no matter what color they were asked to bet on).

What seems to be present in this experiment is the participants' perception of uncertainty. When we say "uncertainty," the usual association is with "probability." The Ellsberg paradox illustrates that some other form of uncertainty can indeed exist. Probability theory provides no basis for the outcome of the Ellsberg experiment.

Klir and Folger (1988) analyze the semantic context of the term "uncertain" and arrive at the conclusion that there are two main types of uncertainty, captured by the terms "vagueness" and "ambiguity." Vagueness is associated with the difficulty of making sharp or precise distinctions among objects. "Ambiguity" is caused by situations where the choice between two or more alternatives is unspecified. The basic set of axioms of probability theory originating from Kolmogorov, rests on the assumption that the outcome of a random event can be observed and identified with precision. Any vagueness of observation is considered negligible, or not significant to the construction of the theoretical model. Yet one cannot escape the conclusion that forms of uncertainty represented by vagueness of observations, human perceptions, and interpretations, are missing from probabilistic models, which equate uncertainty with randomness (i.e., a sophisticated version of ambiguity).

Several reasons may exist for wanting to search for models of a form of uncertainty other than randomness. One is that vagueness is unavoidable. Given imprecision of natural language, or human perception of the phenomena observed, vagueness becomes a major factor in any attempt to model or predict the course of events. But there is more. When the phenomena observed become so complex that exact measurement involving all features considered significant would be impossible, or longer than economically feasible for study, mathematical precision is often abandoned in favor of more workable simple, but vague, "common sense" models. Thus, complexity of the problem may be another cause of vagueness.

These reasons were the driving force behind the development of the fuzzy set theory (FST). This field of applied mathematics has become a dynamic research and applications field, with success stories ranging from a fuzzy logic rice cooker to an artificial intelligence in control of Japan's Sendai subway system. The main idea of fuzzy set theory is to propose a model of uncertainty different from that given by probability, precisely because a different form of uncertainty is being modeled.

Fuzzy set theory was created in Zadeh's (1965) historic article. To present this basic idea, recall that a *characteristic function* of a subset E of a universe of discourse U is defined as

$$\chi_E(x) = \begin{cases} 1 \text{ if } x \in E \\ 0 \text{ if } x \notin E. \end{cases}$$

In other words, the characteristic function describes the membership of an element x in a set E. It equals one if x is a member of E, and zero otherwise.

Zadeh challenged the idea that membership in all sets behaves in the manner described above. One example would be the set of "tall people." We consistently talk about the set of "tall people," yet understand that the concept used is not precise. A person who is 5'11" is tall only to a certain degree, and yet such a person is not "not tall." Zadeh writes,

> The notion of fuzzy set provides a convenient point of departure for the construction of a conceptual framework which parallels in many respects the framework used in the case of ordinary sets, but is more general than the latter and, potentially, may prove to have a much wider scope of applicability, particularly in the fields of pattern classification and information processing. Essentially, such a framework provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables.

In the fuzzy set theory, membership of an element in a set is described by the *membership function* of the set. If U is the universe of discourse, and E is a fuzzy subset of U, the membership function μ_E:U→[0,1] assigns to every element x in the set $\tilde{E}$ its degree of membership $\mu_E(x)$. We write either (E,μ_E) or E~ for that fuzzy set, 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 sets theory. They are considered a special case—a fuzzy set is crisp if, and only if, its membership function does not have fractional values.

On the basis of this definition, one then develops such concepts as set theoretic operations on fuzzy sets (union, intersection, etc.), as well as the notions of fuzzy numbers, fuzzy relations, fuzzy arithmetic, and approximate reasoning (known popularly as "fuzzy logic"). 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 Zadeh (1965) introducing the principal concepts of FST. A complete presentation of FST is provided in Zimmerman (1991).

The first recognition of FST applicability to the problem of insurance underwriting is due to DeWit (1982). Lemaire (1990) sets 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 includes such areas as economics of risk, time value of money, individual and collective models or risk, classification, assumptions, conservatism, and adjustment. Cummins and Derrig (1993, 1994) complement that work by exploring applications of fuzzy sets to property-liability insurance forecasting and pricing problems.

Here, we present a method of fuzzy pattern recognition for risk and claims classification. We apply fuzzy pattern recognition to two problems in Massachusetts private passenger automobile insurance: defining rating territories and classifying claims with regard to their suspected fraud content. Dubois and

Prade (1980), Bezdek (1981), and Kandel (1982) provide overviews of fuzzy techniques in pattern recognition. Zimmerman (1991) and Bezdek and Pal (1992) provide other valuable references on the subject.

The concept of a fuzzy set and the mathematical algorithms needed to implement classification using fuzzy techniques is described in the next section. Grouping towns in Massachusetts into rating territories for risk classification purposes is viewed as a fuzzy clustering problem because many towns can be strongly related to two or more territories, thereby creating a border problem: to which of several related territories should a town be assigned. We also explore the influence of geographical proximity on the resulting fuzzy territories and classification of claims by their suspected fraudulent content. A final section summarizes and provides some alternative and future directions for FST in risk and claims classification problems.

Algorithms for Fuzzy Classification

Lemaire (1990) and Ostaszewski (1993) point out that insurance risk classification often resorts either to vague methods—as in the case of using multiple ill-defined personal criteria to identify good risks to underwrite—or methods that are excessively precise—as in the case of a person who fails to classify as a preferred risk for life insurance application because his or her body weight exceeds 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 that we propose to utilize the methodology of fuzzy clustering in territorial classification and to extend that method to classifying claims for suspected fraud.

Kandel (1982) classifies various techniques of fuzzy pattern recognition. *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. According to Bezdek and Pal (1992), the first choice faced by a pattern recognition system designer is that of process description. The designer may choose from among 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 preprocessing, displays, and extraction. Next, semantic clustering algorithms, generating actual structures in data, are identified. Finally, the designer addresses cluster validity and optimality.

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 semantic algorithm. Because the data we analyze are in the form of numerical vectors (i.e., vectors in a euclidean space), with a number of clusters sought predetermined, we consider the

fuzzy c-means technique most appropriate. Bezdek et al. (1987) discuss the convergence properties of the algorithm.

The task is to divide n objects, where n is a natural number, each represented by a vector in a p-dimensional euclidean space

$$x_1, x_2,..., x_n$$

(coordinates of the vectors are known as *features*), into c, $2 \leq c < n$, categorically homogeneous subsets 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 specified in advance. If the membership function of objects in clusters takes on fractional values, 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).

Risk classification seeks to distinguish risks for the purposes of rating and underwriting. In claims processing, the purpose is to identify claims suspected of fraud for special processing and route nonsuspicious claims through normal adjusting channels. Insurance risks and claims are both described here by certain data patterns. The pattern recognition algorithm does the "detective work" of finding clusters of similar risks and claims.

Let the data set be

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

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

$$x_k = (x_{k,1}, x_{k,2},...x_{k,p}), \; k = 1, 2, 3,..., n$$

is called a *feature vector,* while each $x_{k,j}$, where $j = 1, 2,..., p$, is the jth *feature* of the vector x_k.

A partition of the data set X into fuzzy clusters is described by the set of membership functions of the clusters (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 $S_1, S_2,..., S_c$ with the corresponding membership functions $\mu_{S_1}, \mu_{S_2}, ..., \mu_{S_c}$. In other words, we will construct c clusters that are fuzzy sets.

A c x n matrix containing the values of the membership functions of the fuzzy clusters

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

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

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

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

Condition (1) says that each feature vector x_k has its total membership value of one divided among all clusters, and condition (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). The iterative algorithm consists of four steps; we add a fifth step to make the result operational. The first step sets out a working definition of distance between feature vectors (the vector norm) and an initial starting partition. The second step identifies the center of each cluster in the partition. The third step recalculates the membership functions of the partition as normalized distances from the step 2 centers. The fourth step checks the distance between successive partitions to determine if the iteration procedure should be stopped. The fifth step discards small membership values (below some predetermined α, $0 < \alpha < 1$) to make the partition operational. The five formal steps follow.

Step 1

Choose c, an integer between two and n, as the number of clusters into which the data is partitioned. Choose a positive parameter m, and a symmetric, positive-definite p x p matrix G. Define the vector norm $\| \ \|_G$, using the transpose operator T, by

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

Set the iteration counting parameter ℓ equal to zero, and choose the initial fuzzy partition

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

Choose a parameter $\varepsilon > 0$ (this number will indicate when to stop the iteration process).

Note that the columns of the fuzzy partition matrix, numbered one through n, correspond to data vectors, and each column gives degrees of membership of the data point in clusters one through c. The matrix norm $\| \ \|_G$ is suitably chosen in such a way that two data vectors with great similarities are relatively

close to each other, while dissimilar data are set apart. Although no perfect measure of such relationship exists, we can adjust the scale of x_k coordinates by introducing appropriate diagonal entries, and any known correlations of coordinates can be represented in the nondiagonal entries. The size of the matrix G corresponds to the number of coordinates in data vectors.

The main idea of the algorithm is to produce reasonable centers for clusters of data, and then group data vectors around cluster centers which are reasonably close to them. Unlike in standard crisp algorithms, fractional cluster membership is allowed, which gives us flexibility to adjust for any otherwise desirable phenomena.

Step 2

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

$$v_i^{(\ell)} = \frac{\sum_{k=1}^{n} \left(\mu_{S_i}^{(\ell)}(x_k)\right)^m x_k}{\sum_{k=1}^{n} \left(\mu_{S_i}^{(\ell)}(x_k)\right)^m} \tag{4}$$

for i = 1, 2,..., c.

The cluster centers are merely weighted averages of data vectors. Weights are given by the mth powers of the membership degree. Bezdek et al. (1987) discuss the influence of the scaling factor m, as well as convergence of the resulting algorithm.

Step 3

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

$$\tilde{U}^{(\ell+1)} = \left[\mu_{S_i}^{(\ell+1)}(x_k)\right]_{1\le i\le c, 1\le k\le n},$$

where

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

where i = 1, 2,..., c, and k = 1, 2,..., n.

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

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

This step of the algorithm carries us from the previous membership matrix (numbered ℓ) to the next one (numbered $\ell + 1$). One can interpret formula (5) as follows: if the vector norm measures the similarity of two data vectors, the (m-1)st root of its reciprocal is a form of measure of dissimilarity, and formula (5) assigns a new membership degree by relating the dissimilarity with a given cluster center to the "total dissimilarity present." Formula (5) is, however, a result of a longer optimization procedure discussed further by Bezdek et al. (1987).

Step 4

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

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

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

This "stopping procedure" is a standard numerical analysis technique—if yet another iteration does not change much, the result is the best possible. Clearly, the procedure rests on the assumption of the algorithm's convergence, but luckily the proof of that convergence exists, by Bezdek et al. (1987).

Step 5

The final fuzzy matrix $\tilde{U}^{\ell^*}$ 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 is still fuzzy; for large α (or *max-cuts,* where the largest membership value is set equal to one all others are zero), the resulting partitions are likely to be crisp.

Automobile Rating Territories in Massachusetts

As Conger (1987) points out,

> 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 ambiguous nature of "similar expected losses," a decidedly fuzzy concept.

The methodology used for territorial 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] Because frequent switches from one territory to another are undesirable but inevitable, 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.

Such difficulties and imprecisions in groupings warrant an investigation of fuzzy clustering. Resulting fuzzy clusters would be much more flexible, because a town belonging partially to two or more territories could be assigned to 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 responsiveness criterion for risk classification. 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 and B), personal injury protection (A-2), property damage liability (PDL), collision, comprehensive, and a sixth category comprising the five individual coverages combined. We use those values as the coordinates of vectors x_k, $k = 1, 2, 3,..., 350$, representing the towns in the data space. This implies that we treat the data space as six-dimensional, as six parameters are used to describe towns. In our calculations, we use either the five coverage indices (five-dimensional vectors) or the combined index (one-dimensional vectors) but not both. The data for the 1993 indices (based on the 1987 through 1990 data) for towns in Bristol County are given in Appendix A. Data for all 350 towns and Boston are available in Automobile Insurers Bureau (1992) or from the authors. We begin by illustrating the algorithm for a manageable set of towns: the twenty towns of Bristol County, Massachusetts.

---

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

The Bristol County Algorithm

The initial clustering for Bristol County is the indicated 1993 territory assignment groupings relabeled one to five.[2] The initial five-coverage partition matrix is

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

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

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

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

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

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

where

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

where the subscript p refers to one of the five pure premium coordinates of a town, and $i = 1, 2,..., 5$, $k = 1, 2,..., 20$, 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

[2] For illustrative purposes, the town of Fairhaven, which was assigned to 1993 Territory 9, is included with those towns in Territory 8. Fall River is included with New Bedford. Actual 1993 rating territories are subject to judgmental adjustments and capping and are not always those shown here.

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

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

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

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

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.

Table 1
Fuzzy Town Cluster Membership Values
for Bristol County, Massachusetts

		Membership Values					
Town Name	*Initial Cluster*	μ_{s_1}	μ_{s_2}	μ_{s_3}	μ_{s_4}	μ_{s_5}	*Sum*
Mansfield	1	1	0	0	0	0	1
North 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 five-coverage data pattern, ninth 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

Figure 1
Initial Territorial Town Clustering by Combined Index Territory
for Bristol County, Massachusetts

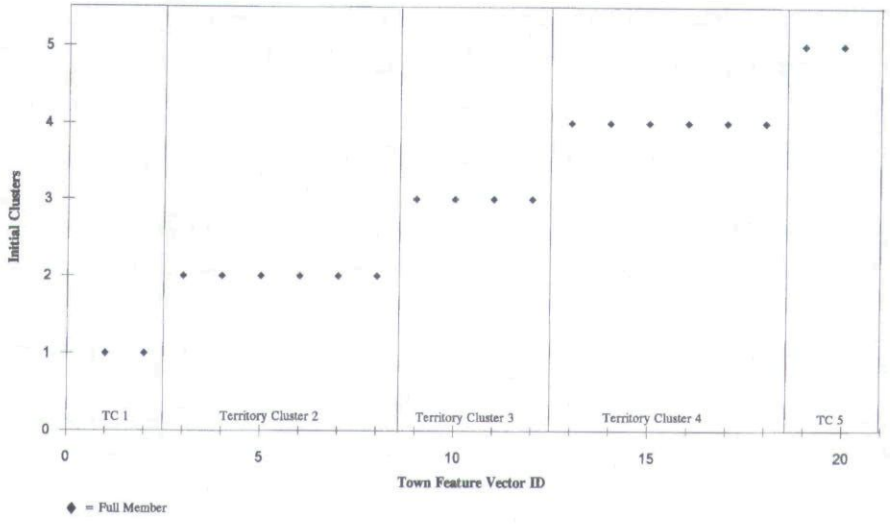
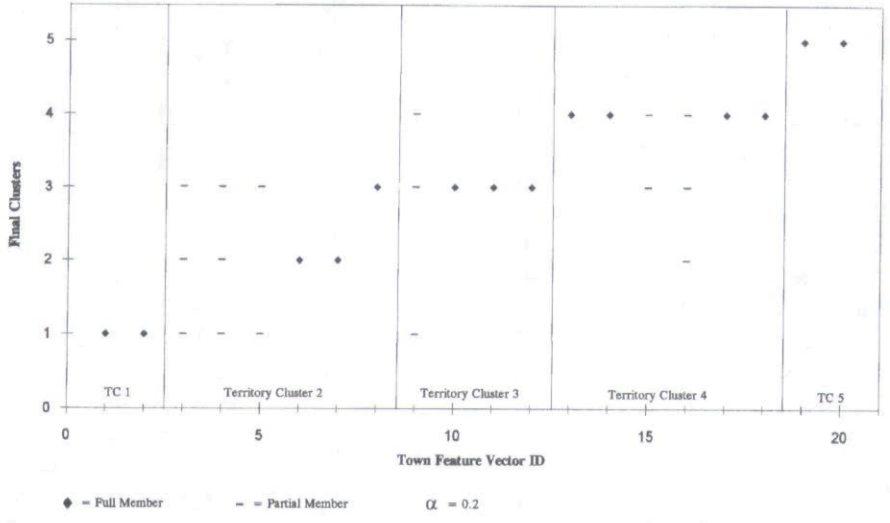


Figure 2
Fuzzy Town Clustering by Five Coverage Indices
for Bristol County, Massachusetts



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, while Town 20 (New Bedford) has the highest index (1.2977) and is in the highest ranked territory.

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

Town 8 (Raynham) becomes associated with fuzzy cluster 3 and has little association (less than 0.2) with its original home cluster 2. Town 5 (Norton) with home cluster 2 splits into fuzzy clusters 1 and 3. These movements are typical of fuzzy clustering results.

Geographical Proximity

We also perform a calculation adding two more features for each town—its geographical coordinates divided by the coordinates of the town with the largest Massachusetts coordinates, Nantucket (the division is performed to adjust the scale and to match the other features, which are all close to one).[4] By performing the algorithm on these vectors, including geographical coordinates, we increase the chance of arriving at clusters that are not only actuarially similar, but relatively close geographically.

This calculation is performed in the same manner as before, but with seven feature variables. We show results for pure premium data weighted 50 percent and geographical variables 50 percent, but any relative weighting scheme can be used to reflect the modeler's preference for geographic dependence of territories. The 50/50 results are presented in Table 2. Note that a full 50 percent weight on the two geographical coordinates produced only slight differences from the five variable centers shown in Table 1. Recall that other states use geographical proximity as an important factor in determining rating territories (DuMouchel, 1983, p. 76). A map of Bristol County is shown in Appendix B.

After the inclusion of 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 geographic variables had on the pure premium cluster centers. Either the geographical variable is already accounted for in the five-coverage pattern or it is relatively weak in relation to the pure premium patterns, at least for this data set.[6]

---

[4] In this application, the fourth root of the ratio is used to bound the geographical coordinates between 0.49 and one, making them more comparable in scale to the pure premium indices. This is equivalent to applying a dilation operator to the simple coordinate comparison to Nantucket in order to produce comparable fuzzy membership values (see Lemaire, 1990, p. 44, and Cummins and Derrig, 1993, p. 452).

[5] Although 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.

[6] The inclusion of location variables to cluster is not necessarily limited to geography. Brockett, Xia, and Derrig (1995) provide two-dimensional location variables that are topographically faithful neural networks for the fraud data discussed below.

Table 2
Fuzzy Town Cluster Membership Values with Geographical Variables for Bristol County, Massachusetts

Town Name	Initial Cluster	Membership Values					Sum
		μ_{s1}	μ_{s2}	μ_{s3}	μ_{s4}	μ_{s5}	
Mansfield	1	1	0	0	0	0	1
North 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 five-coverage and two-geographical coordinate data, eleventh iteration stopping parameter $0.03 < 0.05$, coverage and geographical data equally weighted.

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.

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 five-dimensional patterns and by allowing fractional (fuzzy) membership. In order to isolate the fuzzy effect, we perform one additional clustering exercise using the single combined index variable. The results are displayed graphically in Figure 3.

Table 3
Comparison of Fuzzy Cluster Centers
for Bristol County, Massachusetts

Final Cluster	*A-1&B*	*A-2*	*Property Damage Liability*	*Collision*	*Comprehensive*	*Y-Map*	*X-Map*
Fuzzy Cluster Centers, Loss Data Only							
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	—	—
Fuzzy Cluster Centers, Loss and Geographical Data							
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

Table 4
Fuzzy Town Clusters
Geographical Variables Only

Final Cluster	*Town Name and Initial Cluster Number*
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)

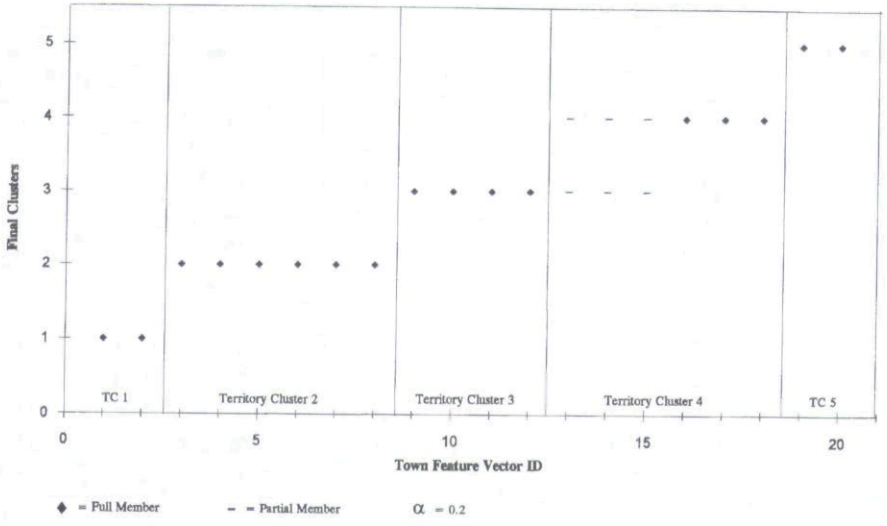
Note: C-means fuzzy clustering algorithm, with two-geographical coordinate data, tenth iteration stopping parameter $0.03 < 0.05$, α-cut = 0.2.

The power of fuzzy clustering to handle towns that are related to two or more territories is demonstrated by the 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.

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. Figure 4 illustrates the results of the fuzzy algorithm applied to the one-dimensional combined index variable currently used to set the 20 indicated territory

Figure 3
Fuzzy Town Clustering by Combined Index
for Bristol County, Massachusetts



boundaries in Massachusetts.[7] The current cluster boundaries are identified in Figure 4 by the vertical territory cluster (TC) delineations. The ability of the fuzzy clustering to model the town association with more than one territory 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 Automobile Insurers Bureau recommended separation of the first three clusters, which are now combined for rating in Massachusetts.

Figure 5 displays the fuzzy clusters using the five-dimensional feature vectors for all 350 towns. Although the fuzzy clusters retain the general pattern of increasing combined index order, the partial membership values reveal much greater dissimilarity among towns that had been grouped together. Fuzzy clusters typically spread across five territory clusters; conversely, each territory cluster typically spreads across four or five 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 indicates the imperfectly correlated nature of the relationship of the five coverage relativities (one town can experience high comprehensive pure premiums be-

[7] Twenty non-Boston territories are indicated in the Automobile Insurers Bureau (1992) filing. Those territories include a split of current territory 1 into three parts (TC1, TC2, TC3 in Figure 4) and a split of three towns (Lowell, Somerville, and Springfield) into two additional indicated territories (TC16, TC18).

Figure 4
Fuzzy Town Clustering by Combined Index
for Twenty Territories in Massachusetts

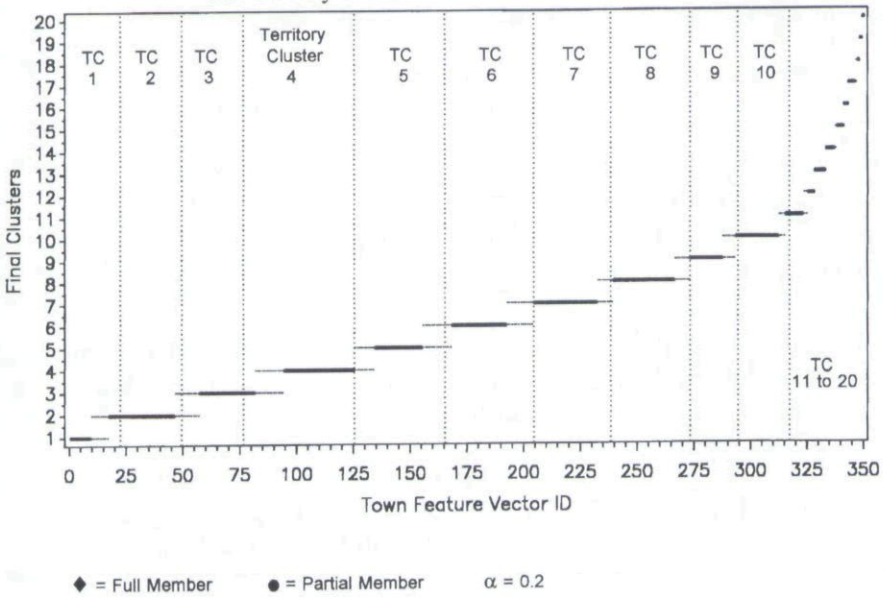
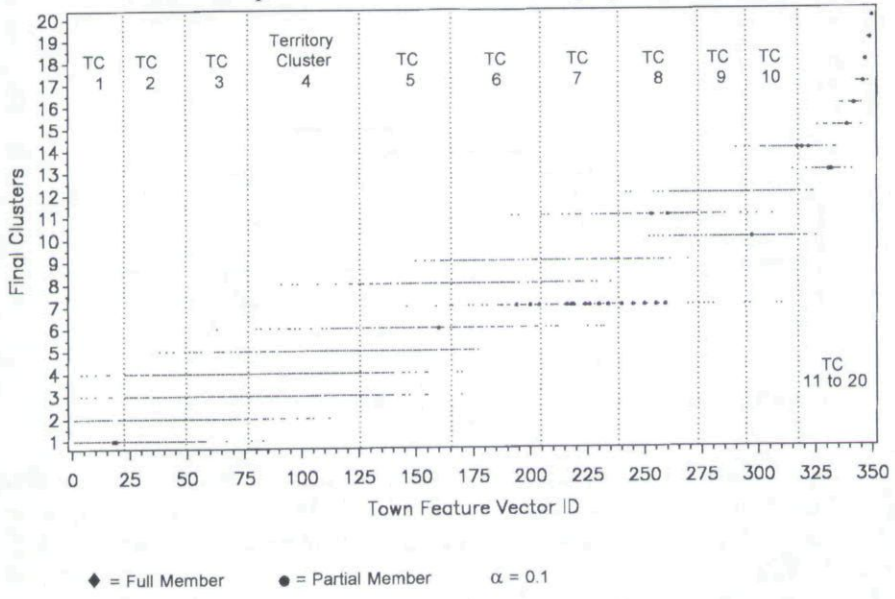


Figure 5
Fuzzy Town Clustering for Massachusetts by Five Coverage Indices
Membership Value Cut at 0.1, No Geographic Variable



cause of a singular theft problem while another town can experience high bodily injury claims experience because of fraudulent claims).

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

		Town Maximum Membership Values			
Max Value	*0-20%*	*21-40%*	*41-60%*	*61-80%*	*81-100%*
Number of Towns	88	189	48	16	9
		With Geographic Variables			
Number of Towns	147	157	31	9	6

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

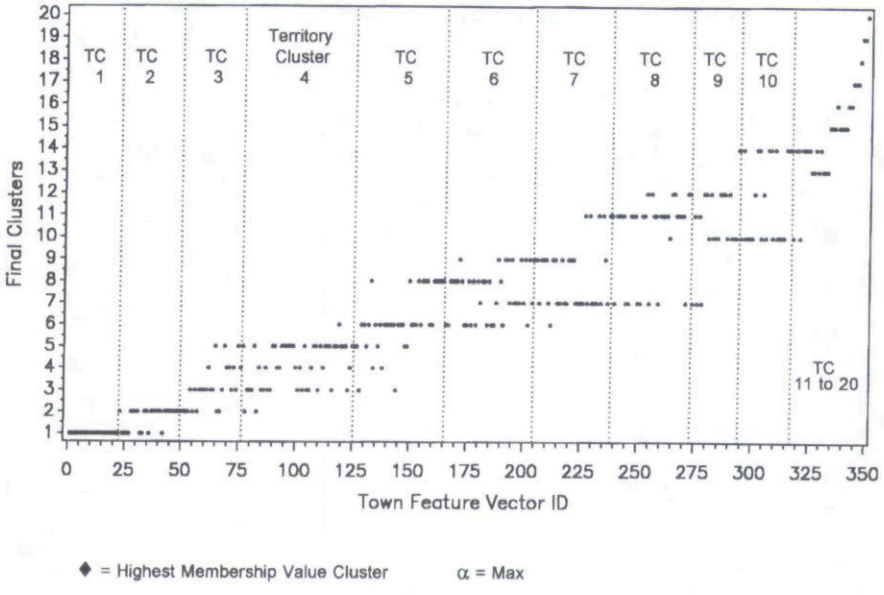
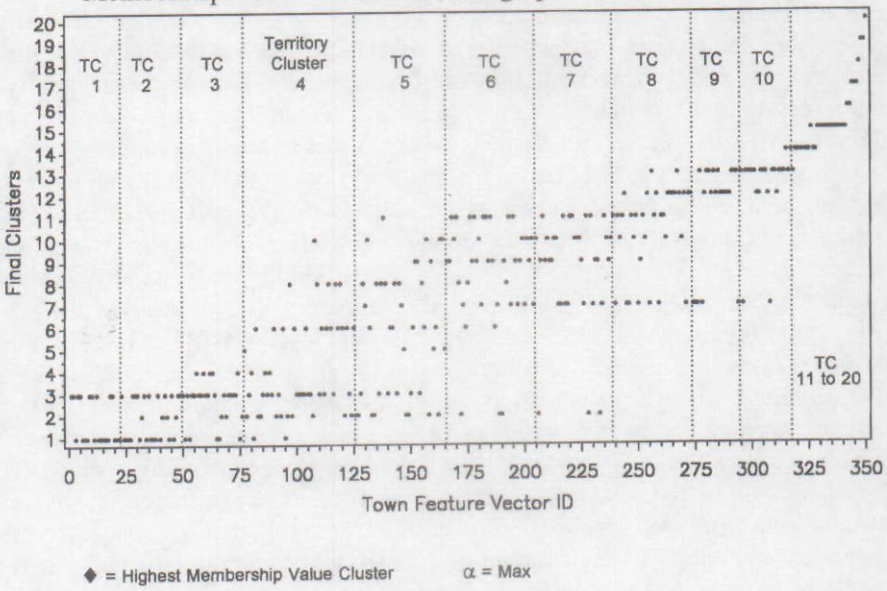


Table 5 also shows the dispersion using 50 percent weighted geographic variables. Unlike the Bristol County examples above, the geographic variables do affect both the cluster centers and the membership values (compare Figures 6 and 7). This result is due to the fact that low territory cluster towns are situated in rural western Massachusetts as well as in eastern Cape Cod. Those town groups have similar five-dimensional pure premiums but have very dis-

Figure 7

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



similar geographic variables that affect the final fuzzy clusters (compare final fuzzy clusters 2 and 3 in Figures 6 and 7).

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 claims processing system for their own 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 data bases, 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 (SIUs). Other than the SIU activities, there is no mention of how the claims processing system deters or detects fraud.

Weisberg and Derrig (1991, 1992) study automobile bodily injury liability and personal injury protection 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 1985/1986 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.[8] The study reveals 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 insurer, 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 claims handlers.

Ideally, one would like to construct a screening device that could be applied by claims 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 bodily injury claims according to the Weisberg-Derrig (1992) study. The difficulty in constructing a fraud/build-up claim screen lies in the ambiguous information (with respect to fraud) provided in support of the claim and subjective categorizations by the claims adjusting observer. Initial attempts at replacing the subjective fraud/no-fraud judgments 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 was assessed by two independent coders for apparent fraud; one reviewed the bodily injury (BI) file and the other reviewed the personal injury protection (PIP) file. Table 7 shows the wide variation in perception of fraud, with only 1.8 percent of the sample claims perceived as fraudulent by both prior coders.

---

[8] The 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 Weisberg and Derrig's (1992) study of post-reform law 1989 claims can be found in "The System Misfired," *Best's Review*, December 1992.

Table 6
Massachusetts Bodily Injury Liability Claims

Claim Type	*Approximate Claim Count Percentage*
Apparent Fraud Referable for Criminal Investigation	1.0
Apparent Fraud Only	9.1
Apparent Fraud or Build-up	48.3
Valid	51.7

Table 7
Suspicion of Fraud by Two Sets of Claims Coders

Fraud Suspected by	*Number Coded*	*Percent of 387 Sample*
Personal Injury Protection Coder Only	27	7.0
Bodily Injury Coder Only	28	7.2
Both Coders	7	1.8
Neither Coder	325	84.0

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

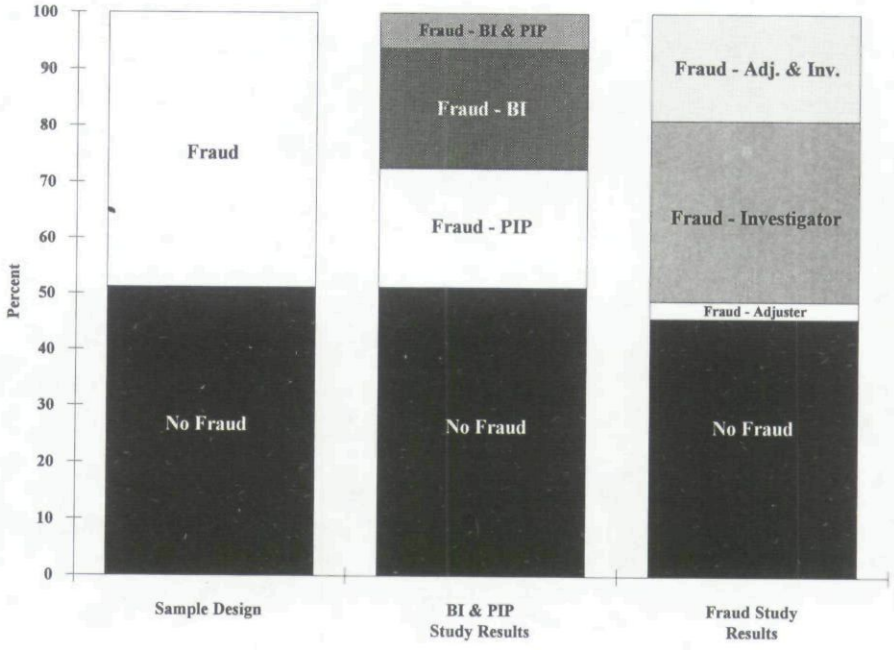
The 1993 Fraud Study Data

The 62 claims deemed fraudulent by at least one BI or PIP coder (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 and Derrig, 1993, 1994). From discussion with claims experts, it was determined that claims investigators would have a different perspective on fraud than insurer 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). Each claim was evaluated by one claims manager and one investigator, giving a total of four independent evaluations of each claim (including the two original BI/PIP study coders).

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.[9] 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 by all four of the coders. Figure 9

[9] Strictly speaking, the 1993 study coders reviewed the entire combined bodily injury and personal injury protection files, while the prior coders, with some exceptions, reviewed one file or the other but not both.

Figure 8
Subjective Assessments of Fraud:
Sample Design vs. Coder Results



shows the decomposition of adjuster and investigator fraud codings by type of bodily injury and personal injury protection coding. Although continued agreement on the bodily injury and personal injury protection no fraud is generally evident, there is a wide disparity of views on the previously coded fraud claims.

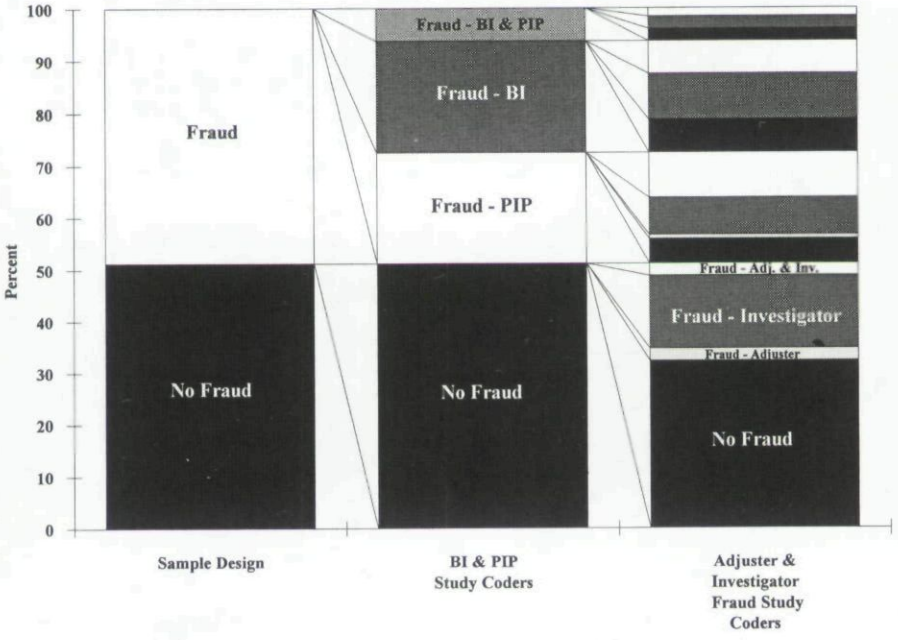
Fuzzy Clustering of Suspicion Measures

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 investigator coders were asked to record their suspicion of fraud on a zero to ten scale. These responses were grouped into five initial clusters based upon adjuster 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, for comparison purposes.[10] The claim data vectors in order are shown in Appendix C.

Similar to our town/territory exercise above, we begin with the initial one-dimensional clustering of claims by adjuster suspicion levels. We apply the

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

Figure 9
Subjective Assessments of Fraud:
Sample Design vs. Coder Results
Plus Decomposition of Adjuster and Investigator Codings



same fuzzy clustering algorithm to the three-dimensional feature vector: the adjuster suspicion value, the investigator suspicion value, and the fraud vote. As with the town/territory exercise, the final fuzzy clustering illustrates both the added dimensions (more information about each claim) and the fractional cluster membership. The fuzzy clusters are shown in Appendix D.

Figure 10 shows graphically how the 127 claims initially clustered by the adjuster suspicion levels are clustered by the fuzzy algorithm using the investigator and fraud vote data and an α-cut of 0.2.[11] 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 five 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 claims handling decisions.

The fuzzy algorithm arranges the claims into clusters of no perceived fraud (center (0,0,0)), little perceived fraud by the adjusters (centers (1,4,1) and (1,8,2)), and strongly perceived fraud by adjusters (center (7,5,2) and center

[11] For illustrative purposes in Figures 10 and 11, the claim pattern data shown in Appendices C and D have been reindexed into three-dimensional lexicographical order. The initial clusters are determined by the adjuster suspicion levels (first feature) zero, 1-3, 4-6, 7-9, and 10.

Figure 10
Fuzzy Claim Clustering by All Suspicion Scores:
Membership Value Cut at 0.2

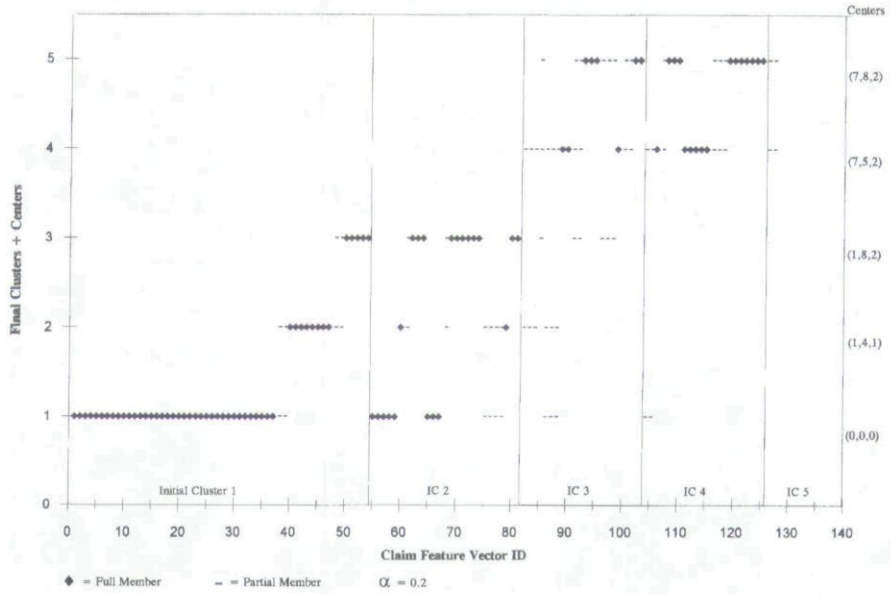
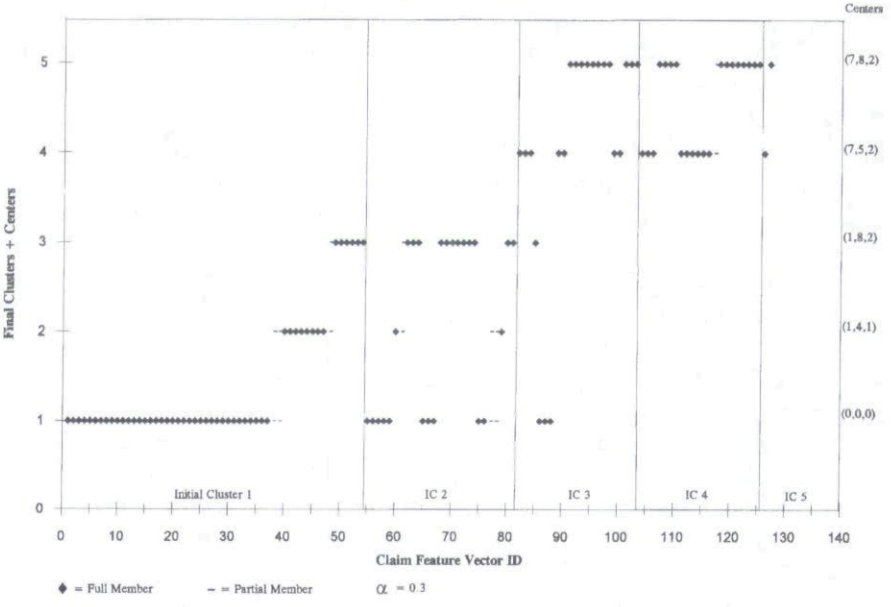


Figure 11
Fuzzy Claim Clustering by All Suspicion Scores:
Membership Value Cut at 0.3



(7,8,2)). Because 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 supports the hypothesis that adjuster suspicion levels can serve well to screen suspicious claims despite the inherent ambiguity in observer perceptions.

Fuzzy Clustering of Fraud Assessment

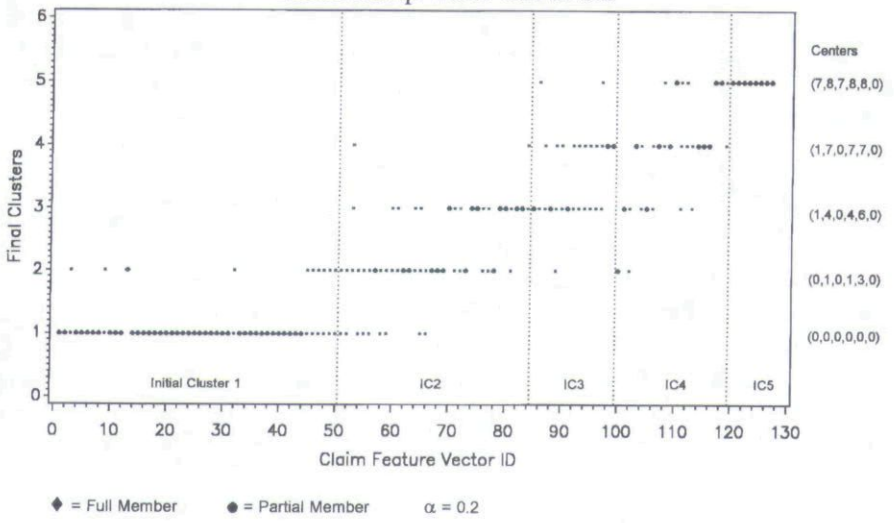
Each adjuster coder was asked to give an overall judgmental assessment in terms of fraud content for each of the 127 claims. The assessment categories were planned fraud, opportunistic fraud, build-up only, and no fraud/build-up. Each claim was also coded for the level of suspicion (zero to ten) for each of six components of the claim: the accident, the injury, the insured, the claimant, the medical treatment, and the wage loss. An application of the fuzzy clustering algorithm allows us to relate the suspicion scores to the fraud assessment in a manner analogous to a regression model. Appendix F contains the feature vector data.

Our five initial clusters are the four fraud assessment categories but with the build-up category divided into two parts: build-up claims with injury suspicion level less than five (IC2) and greater than or equal to five (IC3). Appendix G shows the numerical fuzzy clustering results. Figure 12 shows the results of the application of the fuzzy clustering algorithm and the final cluster centers. Fuzzy cluster 1, with center (0, 0, 0, 0, 0, 0) is clearly the valid nonsuspicious claims. Fuzzy cluster 2 represents treatment build-up only, probably without the involvement of the claimant. Fuzzy cluster 3 has moderate suspicion levels for the injury, the claimant, and the treatment consistent with a deliberate build-up of a minor injury for the purpose of filing a tort claim. Fuzzy cluster 4 shows high suspicion levels for the injury, claimant, and treatment, but not the accident or the insured, consistent perhaps with nonexistent injury and treatment. Fuzzy cluster 5 has high suspicion levels for the first five features consistent with planned fraud. The sixth feature value, wage loss, is zero in all cluster centers reflecting a very low incidence of suspicious wage loss claims. The results of the clustering algorithm provide for alternative groupings of claims by suspicion levels rather than by overall, perhaps vague, assessments of build-up and fraud.

Conclusion

This article explores the usefulness of fuzzy pattern recognition techniques when applied to insurance territorial classification and claims classification problems. We consider 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

Figure 12
Fuzzy Clustering of Fraud Study Claims by Assessment Data: Membership Value Cut at 0.2



sets representative of the two types of problems to illustrate the universality and flexibility of the fuzzy approach.

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 strongly related to two or more 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, 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 fuzzy clustering to be an excellent tool in evaluation of the data provided by claims 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 claims labeling processes is well modeled through fuzzy set techniques.

We do not conclude that fuzzy set theory is either the only way or the best way to analyze fraud-related data. Indeed, Weisberg and Derrig (1993) used regression methods to parse or cluster claims according to discrete model outcome values based upon the presence or absence of so-called fraud indicators. Brockett, Xia, and Derrig (1995) apply a neural network model to the arrays of fraud indicators without using any of the subjective assessment variables.

With the relatively new, and ambiguous, fraud-related data, fuzzy set theory can be one tool in the analytic arsenal to detect claim fraud.

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

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

References

Automobile Insurers Bureau of Massachusetts, 1992, Filing on Automobile Rating Territories, DOI Docket G92-19, Boston.

Bezdek, J. C., 1981, *Pattern Recognition with Fuzzy Objective Function Algorithms* (New York: Plenum Press).

Bezdek, J. C., R. J. Hathaway, M. J. Sabin, and W. T. Tucker, 1987, Convergence Theory for Fuzzy c-Means: Counterexamples and Repairs, *IEEE Transactions on Systems, Man & Cybernetics,* SMC-17: 873-877.

Bezdek, J. C. and S. K. Pal, Eds., 1992, *Fuzzy Models for Pattern Recognition: Methods That Search for Structure in Data* (New York: IEEE Press).

Brockett, P. L., X. Xia, and R. A. Derrig, 1995, Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud, *Special Actuarial Seminar,* Automobile Insurers Bureau, Boston, January 25.

Clarke, M., 1990, The Control of Insurance Fraud: A Comparative View, *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): 1-74.

Cummins, J. D. and R. A. Derrig, 1993, Fuzzy Trends in Property-Liability Insurance Claim Costs, *Journal of Risk and Insurance,* 60: 429-465.

Cummins, J. D. and R. A. Derrig, 1994, Fuzzy Financial Pricing of Property-Liability Insurance, Working Paper, S. S. Huebner Foundation, University of Pennsylvania, Philadelphia.

DeWit, G. W., 1982, Underwriting and Uncertainty, *Insurance: Mathematics and Economics,* 1: 277-285.

Dubois, D. and H. Prade, 1980, *Fuzzy Sets and Systems* (San Diego: Academic Press).

DuMouchel, W. H., 1983, The 1982 Massachusetts Automobile Insurance Classification Scheme, *The Statistician,* 32: 69-81.

Ellsberg, D., 1961, Risk, Ambiguity and the Savage Axioms, *Quarterly Journal of Economics,* 75: 643-669.

Kandel, A., 1982, *Fuzzy Techniques in Pattern Recognition* (New York: John Wiley).

Klir, G. J. and T. A. Folger, 1988, *Fuzzy Sets, Uncertainty, and Information* (Englewood Cliffs, N.J.: Prentice-Hall).

Lemaire, 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 R. A. Derrig, 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 R. A. Derrig, 1992, Massachusetts Automobile Bodily Injury Tort Reform, *Journal of Insurance Regulation,* 10: 384-440.

Weisberg, H. I. and R. A. Derrig, 1993, Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims, *AIB 1994 Fraudulent Claims Payment Filing,* Massachusetts DOI Docket G93-24, Boston.

Weisberg, H. I., R. A. Derrig, and X. Chen, 1994, Behavioral Factors and Lotteries Under No-Fault with a Monetary Threshold: A Study of Massachusetts Automobile Claims, *Journal of Risk and Insurance,* 61: 245-275.

Zadeh, L. A., 1965, Fuzzy Sets, *Information and Control,* 8: 338-353.

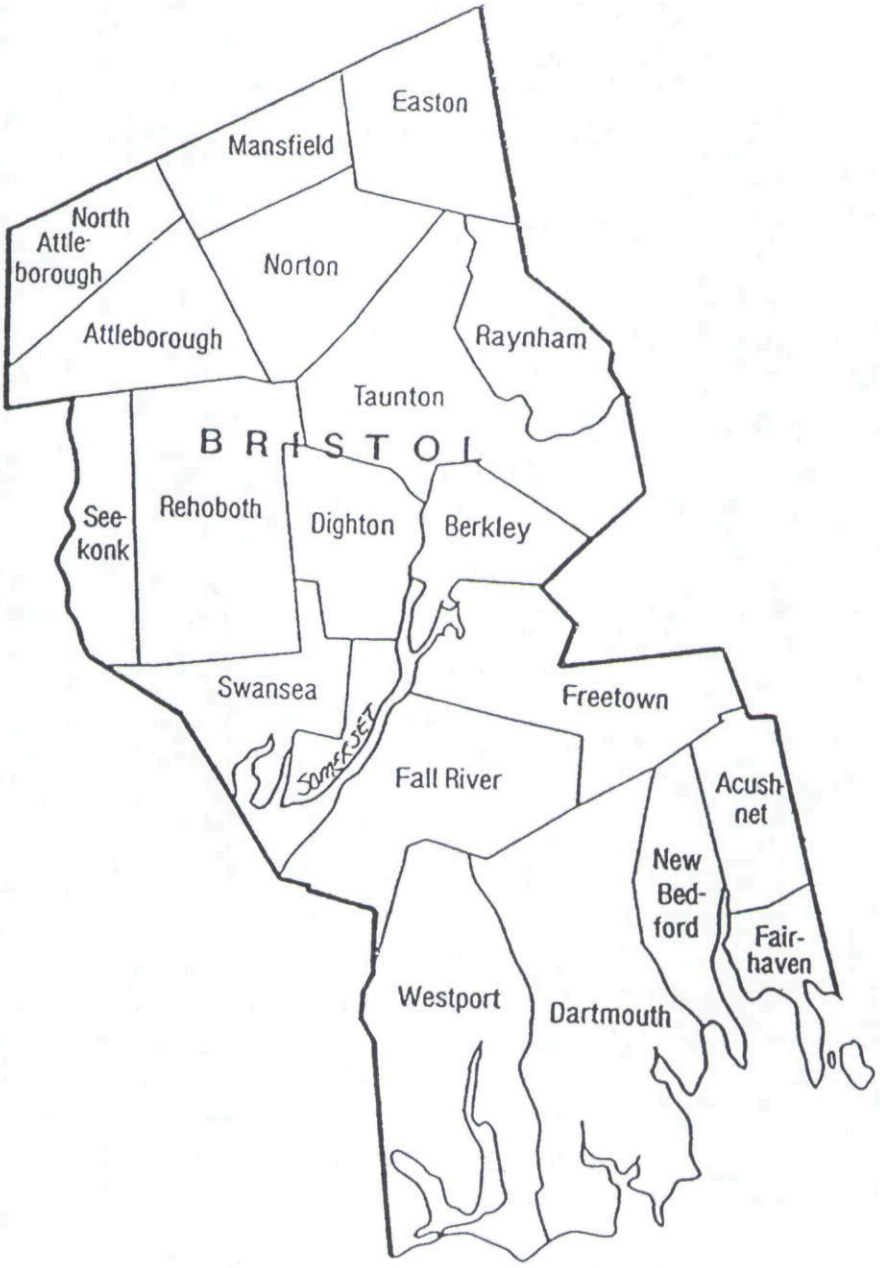
Zimmermann, H. J., 1991, *Fuzzy Set Theory and Its Applications*, Second Edition (Boston: Kluwer Academic Publishers).

Appendix A
Pure Premium and Geographic Indices for Bristol County, Massachusetts

Town Name	1993 Territory	Cluster	Pure Premium Indices by Coverage: Combined	A-1 & B	A-2	Property Damage Liability	Collision	Comprehensive	Y-Map	X-Map
Mansfield	5	1	0.8018	0.7755	0.6849	0.8115	0.8666	0.7699	0.8739	0.9036
North Attleborough	5	1	0.8029	0.7630	0.7170	0.8611	0.8166	0.8167	0.8739	0.8842
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	0.9036	0.9219
Seekonk	7	3	0.9092	0.9010	0.6828	0.9609	0.9149	0.8236	0.9036	0.8842
Easton	7	3	0.9108	0.9172	0.8286	0.9349	0.9443	0.9529	0.8739	0.9036
Attleboro	7	3	0.9171	0.8493	0.8384	0.9577	0.9219	0.9366	0.9036	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	0.9219
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	0.9219

Note: The pure premium indices are ratios of empirical Bayes estimates of town loss pure premium to statewide averages using 1987 through 1990 accident year data (Automobile Insurers Bureau, 1992). 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 fourth root for comparability with the pure premium indices.

Appendix B


Easton
Mansfield
North Attle-borough
Norton
Attleborough
Raynham
Taunton
BRISTOL
See-konk
Rehoboth
Dighton
Berkley
Swansea
Freetown
Somerset
Fall River
Acush-net
New Bed-ford
Fair-haven
Westport
Dartmouth

Appendix C
Suspicion Measures for 1993 Massachusetts Fraud Study Claims

Observation	Adjuster Value	Investigator Value	Fraud Vote	Observation	Adjuster Value	Investigator Value	Fraud Vote	Observation	Adjuster Value	Investigator Value	Fraud Vote
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 values 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 on a zero to four scale. Sample of 1989 Massachusetts bodily injury claims overweighted for fraud judgment (Weisberg and Derrig, 1993).

Appendix D
Fuzzy Claim Clustering by Suspicion Scores
α-cut = 0.2

Cluster	*Adjuster Score*	*Initial Cluster U (0)*	*Final Cluster U (24) [0.2 Cut]*	*Final Center*
1	0	1, 2, 4, 6, 8, 11, 14, 17, 25, 26, 28, 29, 33, 36, 48, 49, 53, 58, 63, 65, 66, 68, 69, 71, 72, 74, 76, 79, 80, 82, 85, 86, 87, 88, 89, 91, 92, 93, 96, 97, 102, 105, 106, 107, 108, 110, 112, 114, 115, 119, 122, 125, 126, 127	1, (0.35)7, 8, (0.48)11, 17, 18, (0.71)22, 25, 26, 44, 48, 53, 56, (0.48)58, (0.35)62, 63, 65, 66, 68, 69, (0.40)70, 71, 72, 73, 76, (0.45)78, 80, (0.74)81, 82, (0.42)83, 84, 86, 87, 88, 89, 91, 92, 93, (0.42)94, 96, 99, 102, 103, 105, 106, 108, 109, 112, (0.42)113, 114, 115, 119, 122, 125, 126, 127	(0, 0, 0)
2	1-3	10, 13, 16, 18, 20, 22, 23, 24, 35, 44, 50, 56, 57, 60, 64, 73, 75, 77, 78, 81, 84, 99, 103, 109, 111, 113, 121	2, 4, (0.52)11, 14, (0.29)22, 23, (0.57)28, (0.31)33, 49, (0.58)50, (0.45)55, (0.52)58, (0.48)67, (0.28)70, 75, (0.55)78, (0.26)81, (0.28)83, 85, (0.28)94, 97, (0.48)100, 107, 110, (0.28)111, (0.58)113	(1, 4, 1)
3	4-6	15, 31, 42, 43, 45, 47, 52, 54, 55, 67, 70, 83, 90, 94, 95, 98, 100, 116, 117, 120, 123, 124	6, 10, 13, 16, 20, 24, (0.43)28, 29, (0.26)31, (0.69)33, 35, 36, (0.24)45, (0.42)50, 57, 60, 64, 74, 77, 79, (0.32)90, (0.72)111, (0.36)116, (0.43)117, (0.29)120, 121	(1, 8, 2)
4	7-9	5, 7, 9, 12, 19, 21, 27, 30, 32, 34, 37, 38, 39, 40, 41, 46, 51, 59, 62, 101, 104, 118	(0.66)3, (0.65)7, 9, 15, (0.35)19, 32, 34, (0.30)40, (0.33)41, (0.75)42, (0.23)45, (0.75)46, (0.55)55, 59, (0.26)61, (0.65)62, (0.52)67, (0.32)70, (0.30)83, (0.30)94, (0.32)95, 98, (0.52)100, 104, (0.29)117, 118, (0.26)120, 123	(7, 5, 2)
5	10	3, 61	(0.34)3, 5, 12, (0.65)19, 21, 27, 30, (0.74)31, 37, 38, 39, (0.70)40, (0.67)41, (0.25)42, 43, (0.53)45, (0.25)46, 47, 51, 52, 54, (0.74)61, (0.68)90, (0.68)95, 101, (0.64)116, (0.28)117, (0.45)120, 124	(7, 8, 2)

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

Appendix E
Fuzzy Claim Clustering by Suspicion Scores
α-cut = 0.3

Cluster	*Adjuster Score*	*Initial Cluster U (0)*	*Final Cluster U (24) [0.3 Cut]*	*Final Center*
1	0	1, 2, 4, 6, 8, 11, 14, 17, 25, 26, 28, 29, 33, 36, 48, 49, 53, 58, 63, 65, 66, 68, 69, 71, 72, 74, 76, 79, 80, 82, 85, 86, 87, 88, 89, 91, 92, 93, 96, 97, 102, 105, 106, 107, 108, 110, 112, 114, 115, 119, 122, 125, 126, 127	1, 8, (0.48)11, 17, 18, 22, 25, 26, 44, 48, 53, 56, (0.48)58, 63, 65, 66, 68, 69, 70, 71, 72, 73, 76, (0.45)78, 80, 81, 82, 83, 84, 86, 87, 88, 89, 91, 92, 93, 94, 96, 99, 102, 103, 105, 106, 108, 109, 112, (0.42)113, 114, 115, 119, 122, 125, 126, 127	(0, 0, 0)
2	1-3	10, 13, 16, 18, 20, 22, 23, 24, 35, 44, 50, 56, 57, 60, 64, 73, 75, 77, 78, 81, 84, 99, 103, 109, 111, 113, 121	2, 4, (0.52)11, 14, 23, (0.57)28, 49, (0.58)50, (0.52)58, 75, (0.55)78, 85, 97, 107, 110, (0.58)113	(1, 4, 1)
3	4-6	15, 31, 42, 43, 45, 47, 52, 54, 55, 67, 70, 83, 90, 94, 95, 98, 100, 116, 117, 120, 123, 124	6, 10, 13, 16, 20, 24, (0.43)28, 29, 33, 35, 36, (0.42)50, 57, 60, 64, 74, 77, 79, 111, 117, 121	(1, 8, 2)
4	7-9	5, 7, 9, 12, 19, 21, 27, 30, 32, 34, 37, 38, 39, 40, 41, 46, 51, 59, 62, 101, 104, 118	3, 7, 9, 15, (0.35)19, 32, 34, 42, 46, 55, 59, 62, 67, 98, 100, 104, 118, 123	(7, 5, 2)
5	10	3, 61	5, 12, (0.65)19, 21, 27, 30, 31, 37, 38, 39, 40, 41, 43, 45, 47, 51, 52, 54, 61, 90, 95, 101, 116, 120, 124	(7, 8, 2)

Note: C-means fuzzy clustering algorithm, with three suspicion measure data patterns for 127 study claims (Appendix C), 24th iteration stopping parameter 0.049 < 0.05, α-cut = 0.3.

Appendix F

Fraud Assessment and Adjuster Suspicion Levels for Fuzzy Claim Clustering

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

Note: Overall assessment codes are valid claim (1), build-up (2), opportunistic fraud (3), and planned fraud (4). Suspicion levels are on a zero to ten scale.

Appendix G

Fuzzy Claim Clustering of Assessment Data by Suspicion Scores
α-cut = 0.2

Cluster	Assessment	*Adjuster Overall Initial Cluster U (0)*	*Final Cluster U (21) [0.2 Cut]*	*Final Center*
1	Valid	1, 2, 6, 8, 11, 14, 17, 25, 26, 29, 35, 44, 48, 49, 56, 63, 65, 66, 68, 71, 72, 73, 77, 79, 80, 82, 84, 86, 87, 88, 89, 91, 92, 93, 96, 97, 99, 102, 105, 106, 107, 108, 110, 114, 115, 119, 122, 125, 126, 127	1, 2, (0.25)4, (0.65)6, 8, 11, 14, 17, (0.31)22, 25, (0.56)26, (0.25)28, 29, (0.26)35, (0.30)36, (0.30)44, 48, 49, (0.25)50, (0.25)53, 63, 65, 66, 68, (0.25)69, 71, 72, (0.26)73, (0.25)76, (0.30)77, 79, 80, 82, (0.25)84, 86, 87, 88, 89, 91, 92, 93, 96, 97, (0.47)99, 102, 105, 106, 107, 108, 110, (0.25)112, 114, 115, 119, 122, 125, 126, (0.57)127	(0, 0, 0, 0, 0, 0)
2	Build-up (Injury Suspicion Level Less Than Five)	4, 10, 13, 16, 18, 22, 23, 28, 33, 36, 43, 47, 50, 53, 55, 58, 60, 64, 67, 69, 74, 76, 78, 85, 100, 103, 109, 111, 112, 113, 117, 120, 121, 123	(0.75)4, (0.35)6, (0.67)10, (0.51)18, (0.37)22, (0.63)23, (0.31)24, (0.44)26, (0.75)28, 33, (0.74)35, (0.70)36, (0.26)43, (0.70)44, (0.33)47, (0.75)50, (0.75)53, 56, 58, 60, 64, (0.75)69, (0.74)73, 74, 75, (0.75)76, (0.70)77, (0.26)83, (0.75)84, 85, (0.53)99, (0.62)100, (0.51)103, 109, 111, (0.75)112, (0.68)120, (0.38)121, (0.43)127	(0, 1, 0, 1, 3, 0)
3	Build-up (Injury Suspicion Level Greater Than or Equal to Five)	7, 31, 37, 42, 45, 54, 57, 62, 81, 83, 94, 95, 98, 116, 124	(0.33)10, 13, (0.59)15, 16, (0.49)18, 20, (0.32)21, (0.32)22, (0.37)23, (0.69)24, 31, (0.38)32, (0.27)37, (0.33)42, (0.47)43, (0.40)45, (0.67)47, (0.33)54, 55, (0.68)57, 67, 70, 78, 81, (0.49)83, (0.33)90, (0.59)94, (0.33)95, 98, (0.38)100, (0.49)103, 113, (0.52)116, 117, (0.32)120, (0.62)121, (0.54)123, (0.52)124	(1, 4, 0, 4, 6, 0)
4	Opportunistic Fraud	5, 9, 12, 15, 19, 20, 21, 24, 30, 32, 34, 40, 46, 52, 59, 61, 70, 75, 90, 101	5, 7, 9, (0.41)15, (0.33)21, (0.70)30, (0.62)32, 34, (0.29)37, 40, (0.67)42, (0.27)43, 46, (0.42)52, (0.67)54, (0.32)57, 59, (0.44)61, 62, (0.25)83, (0.67)90, (0.41)94, (0.67)95, (0.48)116, (0.46)123, (0.48)124	(1, 7, 0, 7, 7, 0)
5	Planned Fraud	3, 27, 38, 39, 41, 51, 104, 118	3, 12, 19, (0.35)21, (0.30)30, 27, (0.45)37, 38, 39, 41, (0.60)45, 51, (0.58)52, (0.56)61, 101, 104, 118	(7, 8, 7, 8, 8, 0)

Note: C-means fuzzy clustering algorithm, with six suspicion measure data patterns for 127 study claims (Appendix F), 21st iteration stopping parameter 0.044 < 0.05, α-cut = 0.2.