©*The Journal of Risk and Insurance*, 1998, Vol. 65, No. 2, 245-274.

Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud

Patrick L. Brockett
Xiaohua Xia
Richard A. Derrig

Abstract

Claims fraud is an increasingly vexing problem confronting the insurance industry. In this empirical study, we apply Kohonen's Self-Organizing Feature Map to classify automobile bodily injury (BI) claims by the degree of fraud suspicion. Feed forward neural networks and a back propagation algorithm are used to investigate the validity of the Feature Map approach. Comparative experiments illustrate the potential usefulness of the proposed methodology. We show that this technique performs better than both an insurance adjuster's fraud assessment and an insurance investigator's fraud assessment with respect to consistency and reliability.

Introduction and Background

One vexing problem confronting the property-casualty insurance industry is claims fraud. Individuals and conspiratorial rings of claimants and providers unfortunately can and do manipulate the claim processing system for their own undeserved benefit (Derrig and Ostaszewski, 1994; Cummins and Tennyson, 1992). The National Insurance Crime Bureau (NICB) estimates that the annual cost of the insurance fraud problem is $20 billion, which is equivalent to the cost of a Hurricane Andrew each year (NICB, 1994). According to the National Health Care Association, insurance fraud in health insurance represented an estimated 10 percent surcharge on the U.S. $550 billion annual health care bill in 1988 (Garcia, 1989). A recent Insurance Research Council report on automobile insurance fraud stated that "the excess injury payments as a result of fraud and/or buildup are estimated to be between 17 and 20 percent of total paid losses, or $5.2 to $6.3

Patrick Brockett is the Gus S. Wortham Memorial Chairholder in Risk Management and Professor of Finance, Mathematics, and Management Science and Information Systems at the University of Texas at Austin. Xiaohua Xia is Vice President, Research and Risk Management at AutoBond Acceptance Corporation in Austin, Texas. Richard Derrig is Senior Vice President of the Automobile Insurers Bureau of Massachusetts and Vice President, Research for the Insurance Fraud Bureau of Massachusetts.

The authors wish to thank the Automobile Insurers Bureau of Massachusetts for providing the data and computing support for this research. This article was reviewed and accepted by the previous editor.

billion additional for all injury claims in 1995." (IRC, 1996) Outside of the United States, fraud claims are also increasing. For example, arson was thought to be costing the United Kingdom £500 million a year in 1991 (Wilmot, 1991).

Private passenger automobile bodily injury (BI) insurance is the largest line of property-casualty insurance in the U.S. It is estimated that about 40-50 percent of BI claims, for Massachusetts at least, contain some degree of suspicion concerning fraud (Derrig, Weisberg and Chen, 1994). The proportion of fraudulent claims also appears to be increasing as evidenced by ever higher rates of bodily injury claims per accident. The Insurance Research Council documents a countrywide change from 22 BI claims per property damage liability claim (the proxy for claims per accident) in 1987 to 29 BI claims per accident in 1992 (IRC, 1994). While the entire increase may not be due to an increase in fraudulent BI claims, the increase is indicative of fraudulent or abusive insurance lotteries (Cummins and Tennyson, 1992).

With the awareness of the increasing frequency of suspected claims fraud, more and more rigorous techniques and empirical databases are being created for the purpose of fraud detection. One such database is the National Insurance Crime Bureau (NICB) Database System, which contains 200 million records of claims and stolen vehicles and which was recently made available to member insurance companies (Dauer, 1993). Once a company enters a claim in the database, either the NICB or that company's special investigation unit (SIU) will commence an investigation if some suspicious information arises for that particular claim.[1] In Massachusetts, a detail claim database (DCD) of all auto BI claims has been assembled commencing January 1, 1994. This database, available to company special investigative units (SIUs) and the Insurance Fraud Bureau, is expected to provide detailed information on approximately two hundred thousand claims annually.

In addition to databases, people began to use other approaches to analyze the automobile bodily injury (BI) claims fraud problem. Using statistical methods, Weisberg and Derrig (1991) determined the mechanisms behind automobile BI claims fraud, such as relationships between injury type and treatment, for example. Their studies of 1985/1986 and 1989 BI claims found that the overall level of suspected or apparent fraud was about 10 percent of the claims, while the apparent build-up[2] level was 35 percent in 1985/86 and 48 percent in 1989.

Nearly all companies rely on the training of personnel as the primary method of recognizing claims suspected of fraud. Specified objective and subjective claim characteristics, such as "no witnesses to the accident," have become known as potential fraud indicators or red flags. Three-quarters of the companies rely on the presence of these red flags to assist the claim adjuster in recognizing suspicious claims, and one-quarter of those companies use automated methods of tracking red flags (IRC, 1992).

Studies have also shown that the insurance industry does not share a consensus definition regarding what constitutes claims fraud. Weisberg and Derrig (1993) found that different BI claims handling professionals had ambiguous perceptions of what constitutes BI claims fraud. For example, in a coding of the same set of claims by two sets of adjusters, each set of adjusters classified approximately 9 percent of claims as apparently fraudulent, but ironically only 1.8 percent of the identified claims were simultaneously considered to be apparently fraudulent by both sets of adjusters. Derrig and Ostaszewski (1994) further verified the lack of concordance of the fraud perceptions among different BI handling professionals, such as between insurance company claims adjusters and insurance investigators. In order to study the problem Derrig and Ostaszewski (1994) applied a fuzzy set-based clustering technique. The results of the study again showed the lack of concordance among people with respect to which claims were fraudulent. Based upon their findings, the analysts concluded that the use of an adjuster's judgment, as compared to that of an investigator, can serve well in first-pass screening of BI claims regarding suspicion levels.

Weisberg and Derrig (1993) used regression models to discern which objective and subjective fraud indicators are more significant than others in effectively identifying suspicion levels of BI claims fraud. If the goal is to identify individual fraudulent claims then their studies exposed several problems. For example, they used only the adjuster's and investigator's subjective assessments of BI claims fraud as dependent variables. As noted above, however, these dependent variables were not consistent with each other, and there was apparent ambiguity and overlap between them. Another problem was that the reliability of each dependent variable couldn't be verified in the real world, due to data limitations.[3] Additionally, statistical approaches, including regression methods used by Weisberg and Derrig (1993), have difficulty handling the 65 binary fraud indicators as independent variables unless the sample size is sufficiently large. Thus, based upon correlation analysis, some 25 indicators were chosen as the independent variables in the regression models; the other 40 indicators were not utilized for practical reasons.[4] Due to these limitations, Derrig and Ostaszewski (1993) did not use the fraud indicators of claims to extract characteristics of fraudulent claims and construct a screening device directly. Rather, they used fuzzy clustering of multiple suspicion levels pertaining to the accident, the claimant, the insured, the treatment, the injury and the lost wages.

[1] The Insurance Research Council reported that about half the property and casualty premium volume was written by companies with special investigative units (SIUs) (IRC, 1984). Given the recent increased emphasis on fraud detection, the establishment of SIUs has expanded greatly. In Massachusetts, for example, all companies writing private passenger auto as a servicing carrier must establish an SIU.

[2] The term "build-up" is defined as an attempt on the part of the claimant and/or the health provider to inflate the damages for which compensation is being sought (Derrig and Ostaszewski, 1994).

[3] Unlike other fraud detection problems, such as credit card fraud, most automobile bodily injury claims cannot ultimately be verified. It is either too costly or impossible to determine and classify without doubt a fraudulent BI claim unless a reliable court decision is available. However, insurance companies tend not to resolve a claim in this manner because it is both risky and costly. As a result, the data used by Derrig, Weisberg and Ostaszewski, and in this study, consists of only objective and subjective indicators or subjective assessments and are not based upon the legal conclusions of whether or not legal fraud was probably present with respect to the BI suspected claims fraud.

[4] For logistical and expense considerations, real company claim operations may be more inclined to track from 10 to 25 indicators systematically rather than 65 indicators. Hence, parsimonious solutions may have more practical value.

Weisberg and Derrig (1991) claimed that at that time it was premature to address the ultimate goals of quantifying the amount of fraud and developing guidelines for detecting and controlling fraud. Since then analysts studying the automobile BI claims fraud problem have been working to ultimately develop a BI claims fraud detection system or claim classification system. Besides the work done by the Automobile Insurance Bureau of Massachusetts (Weisberg and Derrig, 1993; Weisberg and Derrig, 1996), there have been other attempts in this direction. For instance, Artis, Ayuso and Quillen (1997) model the behavioral characteristics of the claimants and insureds in the Spanish automobile insurance market. An expert system has been developed in Canada "to aid insurance company adjusters in their decision making and to ensure that they are better equipped to fight fraud" (Belhadji and Dionne, 1997). In this empirical study, we intended to apply a different approach to build a BI claim fraud detection or classification system. Specifically, we apply a neural network approach, Self-Organizing Feature Maps (Kohonen, 1982, 1989, and 1990), to construct a claim classification system that uses similar collections of fraud indicators as the classifier.

In the second stage of the study, we do a comparative study between the feature map BI claim classification system and both the adjuster's assessment and the investigator's assessment. Claim adjusters and investigators represent the two primary forces in claim processing and fraud detection. Their expertise will serve as a good benchmark for a novel quantitative approach such as the feature map method. The tool used in the comparative study is another neural network model. Specifically, a feed-forward neural network model combined with a back-propagation learning algorithm. Particularly, we would like to see whether the feature map approach can perform better than the adjuster's and investigator's subjective assessments as measured by the consistency achieved in assessing suspicion levels and clustering BI claims

An overview of the paper is as follows: following this background information, the second section describes the empirical BI claims data used in the study. This data sample is used to construct the feature map models and apply them to BI fraud detection problem in the next two sections. Feed-forward neural network models are constructed to test the feature map approach in the penultimate section. A summary section concludes the paper. The Kohonen's Self-Organizing Feature Map Algorithm is presented in an Appendix.

BI CLAIMS FRAUD DATA

The data set was generated in a study of Massachusetts BI claims and previously was analyzed by Weisberg and Derrig (1991,1992, and 1993). The entire BI claim data set consists of 127 claims, selected from among 387 claims for accidents in 1989. The data production process was completed in two steps. In a first pass through all 387 claims data, each claim was independently examined by two claim adjusters. Of these, 62 claims were deemed to be apparently fraudulent by at least one of the adjusters. The other 65 claims out of the total 127 claims were randomly sampled from the remaining 325 apparently non-fraudulent claims. In a second pass through the data, these 127 claims were again independently coded, this time by an insurance adjuster and by an investigator from the Insurance Fraud Bureau of Massachusetts. Each claim in the data set consists of a claim ID number, a vector of fraud indicators (we will use claim vector, pattern, and pattern vector interchangeably) and two professional assessments of the suspicion level of fraud, i.e., the adjuster's assessment, and the investigator's assessment. In total, there are 65 fraud indicators which have been divided into six categories based upon the practice used in automobile insurance claim processing: characteristics of the accident, the claimant, the insured, the injury, the treatment and the lost wages. Some indicators, such as the accident characteristics, were based on the police report and witness testimony, while others were collected from actual claim files. Every indicator is a dummy variable and assumes a binary value based upon the answer to a yes-no question. The adjuster's and investigator's assessments, which reflect their opinions of the level of suspicion for the claim, fall into a range between 0 and 10, with 10 standing for a virtually certain fraudulent claim, and 0 for a virtually certain valid claim. The distribution of their assessments is shown in Table 1. Adopting the convention used by Weisberg and Derrig (1992), we transform the ten-point suspicion level variable into four discrete categories: not suspicious (0), slightly suspicious (1-3), moderately suspicious (4-6) and strongly suspicious (7-10).

Table 1
Breakdowns by Adjuster Assessment and Investigator Assessment.

Adjuster Assessment						
	Training Set		Holdout Set		Combined	
Suspicion	Count	%	Count	%	Count	%
0	35	45%	19	38%	54	43%
1-3	11	14%	16	32%	27	21%
4-6	12	16%	10	20%	22	17%
7-10	19	25%	5	10%	24	19%
Total	77	100%	50	100%	127	100%
Investigator Assessment						
	Training Set		Holdout Set		Combined	
Suspicion	Count	%	Count	%	Count	%
0	32	42%	19	38%	51	40%
1-3	5	6%	6	12%	11	9%
4-6	11	14%	10	20%	21	17%
7-10	29	38%	15	30%	44	35%
Total	77	100%	50	100%	127	100%

It is necessary to emphasize the fact that, besides the subjective assessments of two professionals, there are no so-called observed (true) fraudulent levels available in the data set[5]. This makes statistical prediction methods, such as logistic

[5] It is normally true since an insurance claim is rarely identified in practice as a fraudulent claim unless some extensive special investigation and/or legal procedures are involved.

regression, impossible to implement in the first place unless we are willing to assume the validity of the professionals' subjective assessments. Not surprisingly, building a reliable dependent variable to measure the level of suspicion is one of the major goals of this study.

We realized that neural network models are normally very computationally intensive. Kohonen Feature Maps, as described in an Appendix, are no exception. Based upon our experience, training a feature map until it reaches a stable state often takes hours on clustered Unix stations. It often becomes too time and resource consuming to conduct extensive experiments based upon advanced sampling techniques. In our first experiment, we thus decided not to do multiple samples or use other advanced sampling techniques. All 127 claim vectors were randomly mixed together before 77, or about 60 percent of the claims, were randomly chosen as the training data set, with the remaining 50 claims used as the holdout data set. We selected a 60:40 proportion for the purpose of having as many claims in the training set as possible but reserved "enough" claims to test the validity of the methodology. The distributions of the professionals' assessments based upon this first sampling in terms of the level of suspicion are summarized in Table 1. We can easily see that the percentage of valid claims perceived by the adjuster and the investigator are comparable. However, among the same 127 BI claims the investigator found more claims to be "strongly suspicious" than did the adjuster. For example, the investigator found 10 more claims (29 versus 19) "strongly suspicious" claims than did the adjuster in the training data set alone.

In our second experiment, we took 9 additional random samples from our data set of 127 claims. The size of the training samples was still 77 claims, or 60 percent of the 127 claims. These 9 random samples plus the original sample were run on smaller feature maps (smaller number of output units). With the multiple samplings, the common results are expected to be more robust.

APPLYING KOHONEN FEATURE MAPS

It must be assumed that if two claims have common or similar fraud indicator patterns, the result would be an approximately equivalent level of suspicion. Consequently, two claims whose indicator vectors have a sufficiently short distance between them should be assigned similar suspicion level values. Hence, if there is a mapping from the claim patterns (vectors) to a suspicion level assessment, such that similar assignments go to the similar patterns and different assignments to different patterns, then we can claim such a mapping system is consistent and reliable. This is the consistency or continuity principle of unsupervised pattern recognition approaches (Schalkoff, 1992).

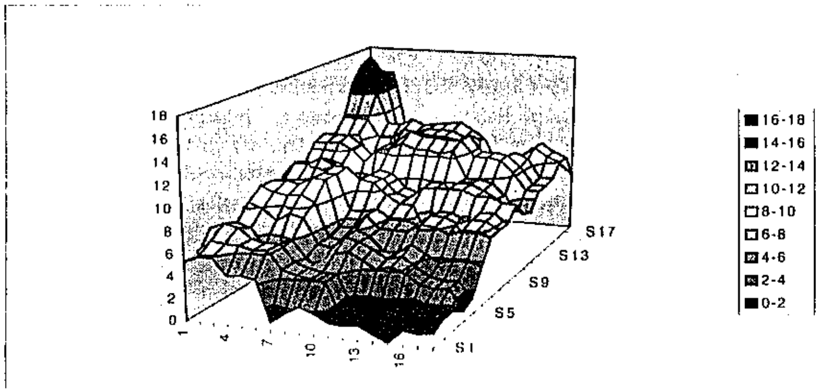
Cluster Analysis has been a popular method dealing with unsupervised learning problems (lack of observed values of output variables). Statistical software packages such as SAS include various cluster analysis models. Certain optimization criteria are applied to cluster analyses to split a set of observations into a number of groups or a hierarchical structure. The distribution of observations among groups is determined by the optimization criteria. Hence, it might happen that one of the groups has a single or very few vectors while another group has 99 percent of the vectors. The application of optimization rules in cluster analysis leads to a lack of control over the numerical process that might have better performance if the process were interactive and certain *a priori* knowledge were to be incorporated. In cases where the data sample is small but its dimensionality is high, or simply the data is quite noisy, conventional cluster analyses might leave little of the necessary freedom for professionals to provide expert inputs into the decision making process. Another disadvantage associated with conventional cluster analyses is that it is not easy to identify the groups in terms of the nature of the observations in each group, such as which group represents the "strongly suspicious" insurance claims and which group represents the apparently "valid" claims in this particular case. Kohonen's Feature Maps (Kohonen, 1982, 1989, and 1990) are capable of dealing with this unsupervised problem while overcoming the weaknesses inherent in conventional cluster analyses. See Appendix for the description of a typical feature map algorithm.

For the demonstration, we used an 18×18 square feature map and trained it for 2,000 epochs. Each cell in the 18×18 map is assigned a random "weight vector" of the same dimension as the claim or pattern vector – namely the number (65 in this case) of fraud indicator variables. The training data set is the one from our first sampling and consists of 77 claims and their accompaning pattern vectors. We know that a distance can be calculated between each cell's weight vector and a pattern vector. For any given pattern vector, we first found the maximum distance among all the weight vectors to the pattern vector and then subtracted each distance from the maximum. The result became a measure of the "closeness" between the weight vectors and the pattern. We then depicted the "closeness" in a three-dimension space as shown in Figures 1 through 4. In the graphs, the height of the landscapes measures the "closeness" of the weight vectors to the pattern vector, i.e., the higher the landscape at location (i, j), the better the matching between the pattern vector and the weight vector (or the output unit) at (i, j).

Figure 1 shows the landscape for a claim pattern p_A after 2,000 epochs of learning, and adjusting the cells' weight vectors in the manner to be specified subsequently. It is clear that the highest peak lies in the upper corner, which implies that this claim pattern finds its best-matching output unit in that corner. Notice the clear pattern of the landscape decreasing towards the lower corner, this means that the matching between pattern p_A and the output units (their weight vectors) gets worse and worse towards the lower corner. It seems from this landscape that the training for the demo feature map was perfectly done. But the next figure shows the presence of noise. Figure 2 is the landscape for another claim pattern, p_B. There are two things clearly different in this landscape. First, the highest peak is located at the lower corner. This implies that claim pattern p_B might be quite different than the claim p_A. Second, besides the highest peak, there are other noticeable peaks, i.e., the landscape does not present a near perfect pattern like the one in Figure 1. This suggests that the training might not have been perfectly executed, or that the training data set contains enough noise to make the identity of this claim ambiguous.

FIGURE 1

Matching BI Claim p_A by 18×18 Demo Square Feature Map after 2000 Epochs



Note: In this graph, the height of the landscape implies the "closeness" (rather than the distance) of weight vectors to the claim p_A. That is, the "closer" a weight vector to the claim, the higher the landscape at the corresponding location.

FIGURE 2

Matching BI Claim p_B by 18×18 Demo Square Feature Map after 2000 Epochs

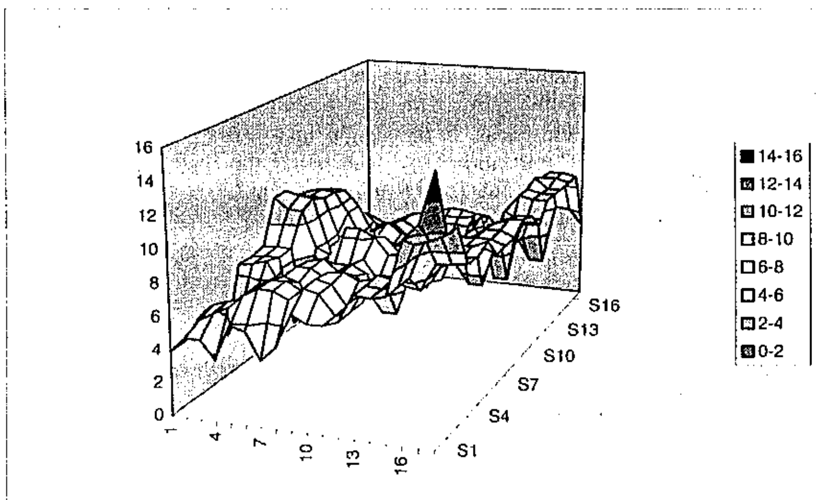


Figure 3

Matching BI Claim p_A by 18×18 Demo Square Feature Map after 0 Epoch

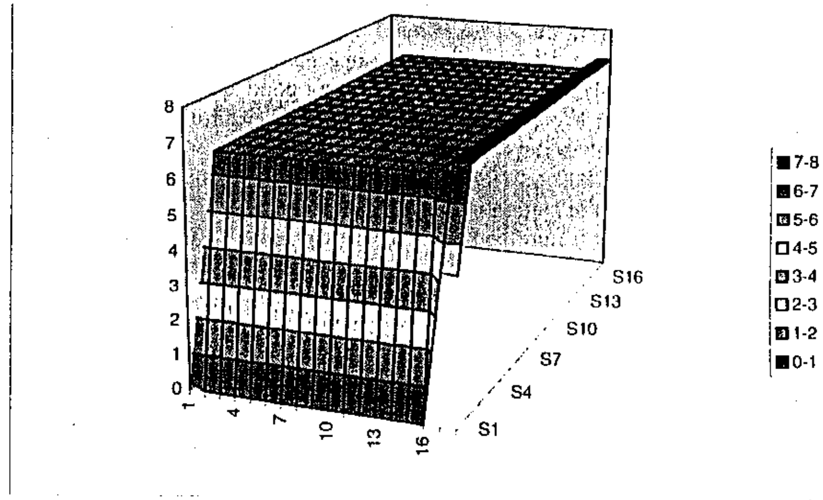
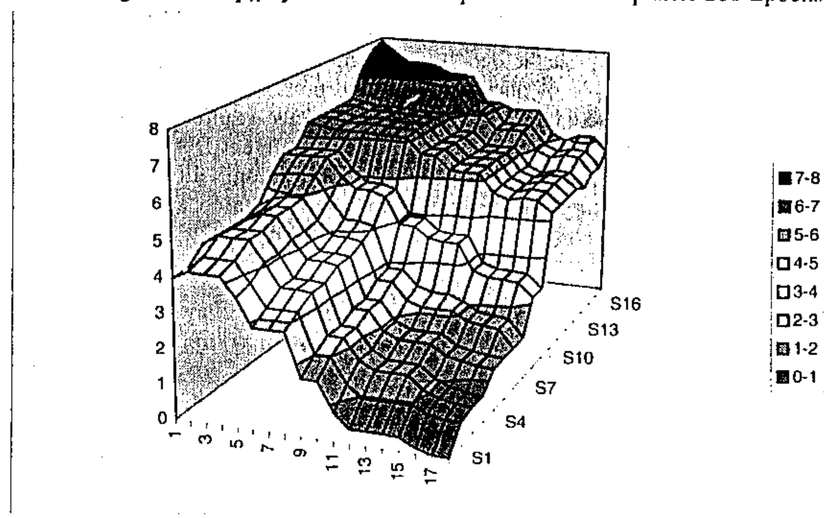


Figure 4

Matching BI Claim p_A by 18×18 Demo Square Feature Map after 200 Epochs



It is interesting to see the evolution of a feature map during the course of training (learning). The landscapes in Figure 3 and 4 were generated for claim p_A

from the feature map right after initialization and after the feature map had been trained for 200 epochs. Recall that we are particularly interested in the highest peak when we read a landscape. However, we fail to locate a single highest peak on the landscape in Figure 3. It can be easily understood since the weight vectors were initialized with small random numbers. The feature map was thus in a chaotic status initially. However, the chaos should gradually be replaced by a more ordered status as training proceeds. For the same pattern p_A, the landscape after 200 epochs (see Figure 4) had already been in pretty good shape. It looks similar to that shown in Figure 1 except that the highest peak in Figure 1 (between 16 and 18) seems much higher than the one in Figure 4 (between 7 and 8). This reflects the fact that the extra 1,800 training epochs enable the feature map to detect or capture this particular claim pattern better.

We now look at this feature map from a different point of view. Assume that there is no external information in terms of the identity of all the BI claims such as the adjuster's assessment or the investigator's assessment but we would like to identify a claim from the feature map. How would such a feature map be perceived or understood, if the feature map has learned from a data sample consisting of a set of independent variables but no observed dependent variable?

We know, from the landscape graphs and the feature map algorithm description in the Appendix, that the weight vectors of a feature map evolve from a set of numerical vectors, composed of small random numbers initially, to their stable status. These weight vectors form a classifying system. With a set of binary BI claim vectors as the training sample, the value of certain components of some weight vectors will increase in order to match certain types of BI claims as training goes on. In other words, "valid" claims are mostly represented by a vector containing many zeros, meaning the non-presence of the corresponding fraud indicators while "strongly suspicious" claims are more likely represented by a vector with a larger number of ones, implying the presence of the corresponding fraud indicators. Hence, to capture a "strongly suspicious" claim, the weight vector of an output unit in a feature map should be quite different from a null vector. However, if an output unit is to match well with "valid" claims, all the components of its weight vector will more likely keep small values. Based upon this notion, define a quantity S_m on the weight vectors $m=[m_{ij}]$, the distance to the null pattern, as:

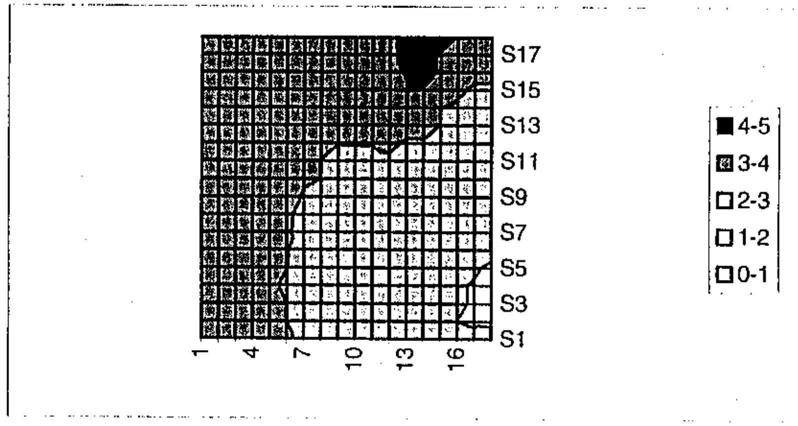
$$S_m = \|m_{ij}\| = \sqrt{\sum_{d=1}^{D} m_{ijd}^2} \qquad (1)$$

where D is the dimension of the weight vector (65 in our application). The quantity S_m provides a measure of whether the output unit corresponds to a highly suspicious claim (high S_m) or a non-suspicious (low S_m).

From the same feature map, we calculated the values for all 324 (18×18) weight vectors using the formula shown above. The calculated values are displayed in Figure 5. In the graph, an output unit is located at each intersection; the scale of

gray at each intersection represents the magnitude of the value for the corresponding weight vector. That is, the darker the color at an intersection, the larger the quantity S_m for the corresponding output unit, the more "suspicious" the claims that match the best with the output unit.

Figure 5

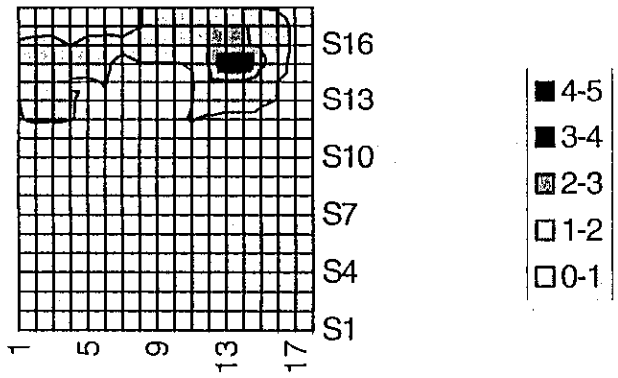


Calculating the quantity on 18x18 demo feature map's weight vectors using formula (1). Each intersection corresponds to one output unit. It is suspected that the darker the color at an intersection, the better the output unit matches with "strongly suspicious" claims.

Based on the interpretation of the S_m measure, it seems that the small area in black near the upper right corner in Figure 5 is the area for "strongly suspicious" claims. The surrounding area in dark gray might be for "moderately suspicious" claims while the neighboring area in slightly lighter color might be for "slightly suspicious" claims. Accordingly, the "valid" claim zone must be the small area in light color near the lower right corner.

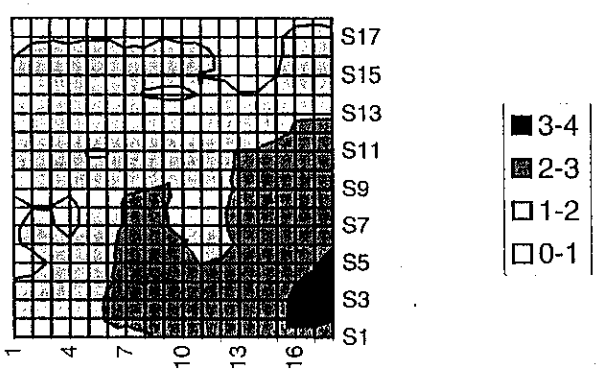
Next, we would like to identify two different claims, p_C and p_D, based upon the interpretation of the feature map illustrated in Figure 5. Similar to the process of generating three-dimension landscapes as described previously, a value measuring "closeness" was calculated between each weight vector and each claim pattern. The "closeness" values were displayed in Figures 6 and 7 for claims p_C and p_D. In each graph, the darker the color at an intersection, the closer the corresponding weight vector to the claim pattern. In Figure 6, we see that the darkest area for claim p_C is close to the upper right corner. This area was designated as "strongly suspicious" area from Figure 5. This suggests that claim p_C (see Figure 7) ought to be considered "strongly suspicious." In contrast, claim p_D is more likely to be a "valid" claim. Incidentally, the adjuster gave claim p_C a "7" as its suspicion level while the investigator's mark was "10." However, both the adjuster and the investigator thought claim p_D was valid, i.e., gave "0" as its suspicion score. Readers are reminded here that what we have done so far is just an illustration. Formal evaluations are made below.

Figure 6



Calculating the "closeness" between claim p_c and each of 324 weight vectors. The smaller the distance between a weight vector and the claim vector, the larger the "closeness." Based upon Figure 5, claim p_c is likely to be "strongly suspicious".

FIGURE 7



Calculating the "closeness" of 324 weight vectors to claim p_D. This claim is most likely to be considered "valid", based upon the interpretation on Figure 5.

AN APPLICATION OF FEATURE MAPS TO BI FRAUD DETECTION

Based on our past experience, in order to achieve a good mapping effect, the number of output units of a feature map is usually four times (or more) the number of training pattern vectors. Thus, we designed a 20×20 two-dimensional output

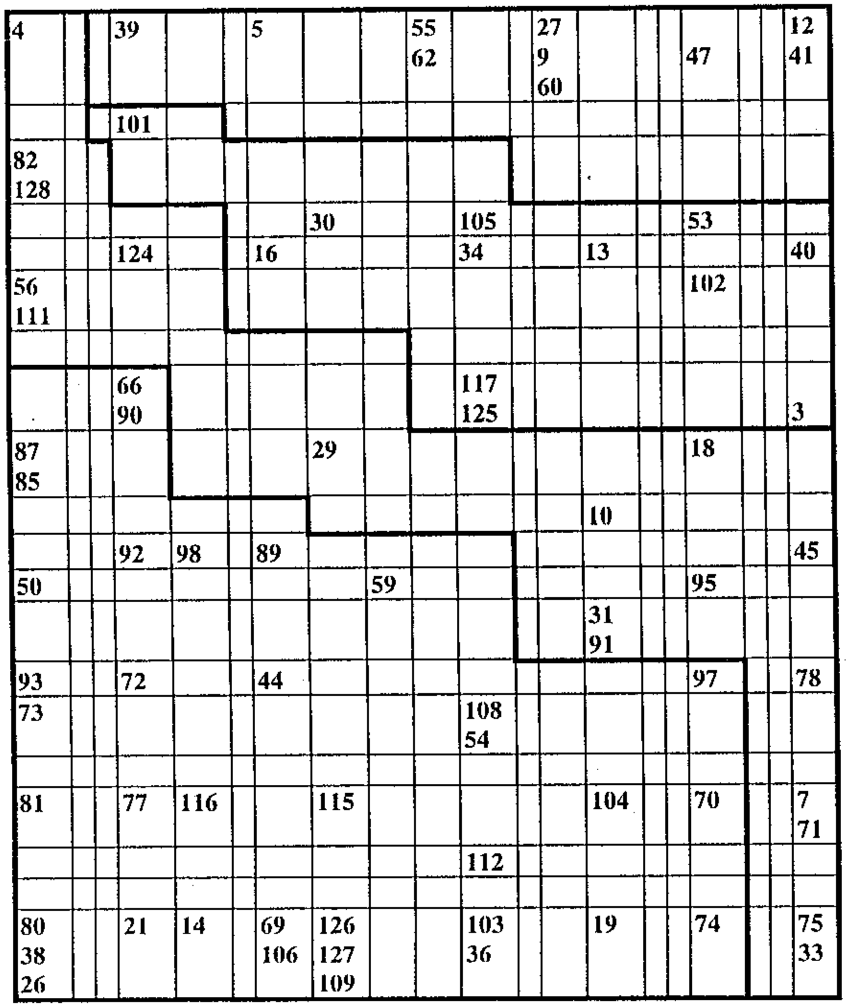
map to match against the training sample composed of 77 pattern vectors. These 400 neural units were arranged in the form of a square. A square feature map form was selected because of its simplicity and tractability. Higher dimensional feature maps are possible but we sacrifice the intuitive graphical realizations of Figures 1-7. We used the algorithm described in the Appendix, to train the feature map for 2000 epochs.

Since it is a square feature map, we chose a 20×20 table to represent the feature map, with each cell representing an output unit. Each claim pattern has its best-matching output unit. We displayed the matching between the claims (their ID numbers) in the training data set in Figure 8. Since each claim in the data set had been assigned a suspicion score by the adjuster and the investigator. Figure 9 and 10 display the adjuster's assessment and the investigator's assessment, respectively. Let's call those figures the ID map, adjuster map and investigator map for the training data set. Each claim in the holdout data set also has its best-matching unit. Figures 11, 12 and 13 are the ID map, adjuster map and investigator map for the holdout data set. Figures 8 through 13 were generated after the training for the feature map had terminated, i.e., after 2,000 epochs.

We could have produced a three-dimension landscape for each individual claim vector. We could also have produced two-dimension graphs as we did in the third section. Instead, we chose to generate the mapping for the whole training set or the whole holdout set. By generating adjuster maps and investigator maps, we could see the overall distribution of their assessments. A well-trained feature map is supposed to provide order rather than disorder for the fraud indicator clustering in mapping. But this is not necessarily true if the adjuster's assessment or the investigator's assessment, unless their assessments are about equal. Accordingly, the map for the adjuster's assessment becomes a measure of how well or how consistently the adjuster and investigator did in evaluating the BI claims. Both the adjuster and the investigator are experienced claim experts. Their evaluations of the possibility of fraud or suspicion level might prove to be less than perfect but their expertise will certainly be of great value. Although we are able to produce an individual landscape for each claim in the data set and find the location of its highest peak, we still have to identify a claim whose highest peak is located in the upper-right corner as well as one whose peak is located in the bottom-left corner. We expect that the maps of the adjuster's assessment and the investigator's assessment provide such an explanation.

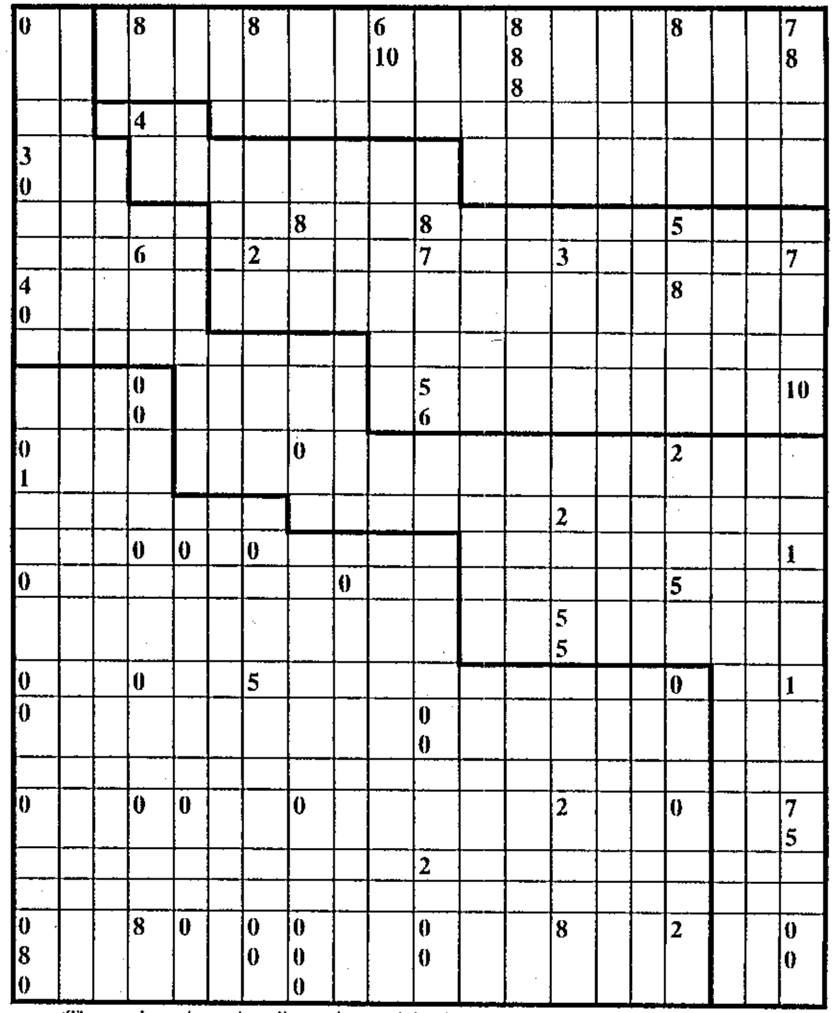
From Figures 9 and 10, it is clear that the larger numbers, indicating higher suspicion level assigned by the adjuster to the claims in the training data set, tend to concentrate in the upper-right area while zeros are mostly distributed in the lower-left area. Hence, these two feature maps on the training data set suggest an overall pattern that "strongly suspicious" claims are most likely to find their "peak" or "best-matching" output unit in the area of upper-right corner and its neighborhood. The suspicion level of the claims to be mapped diminishes gradually towards the opposite corner where the output units act like the magnets abstracting "valid" claims.[6]

Figure 8
1D Map for the Training Data – Locating the Best-Matching Unit for Each Claim in the Data Set



[6] Note that the interpretation of the 18×18 feature map shown in Figure 1-7 is slightly different from that of this 20×20 feature map. This is mainly because (1) training feature map is a stochastic process and (2) the quantity S_m, which is analogous to counting fraud indicators, may not be equivalent to the approaches the adjuster and investigator used in the evaluation of BI claim suspicion.

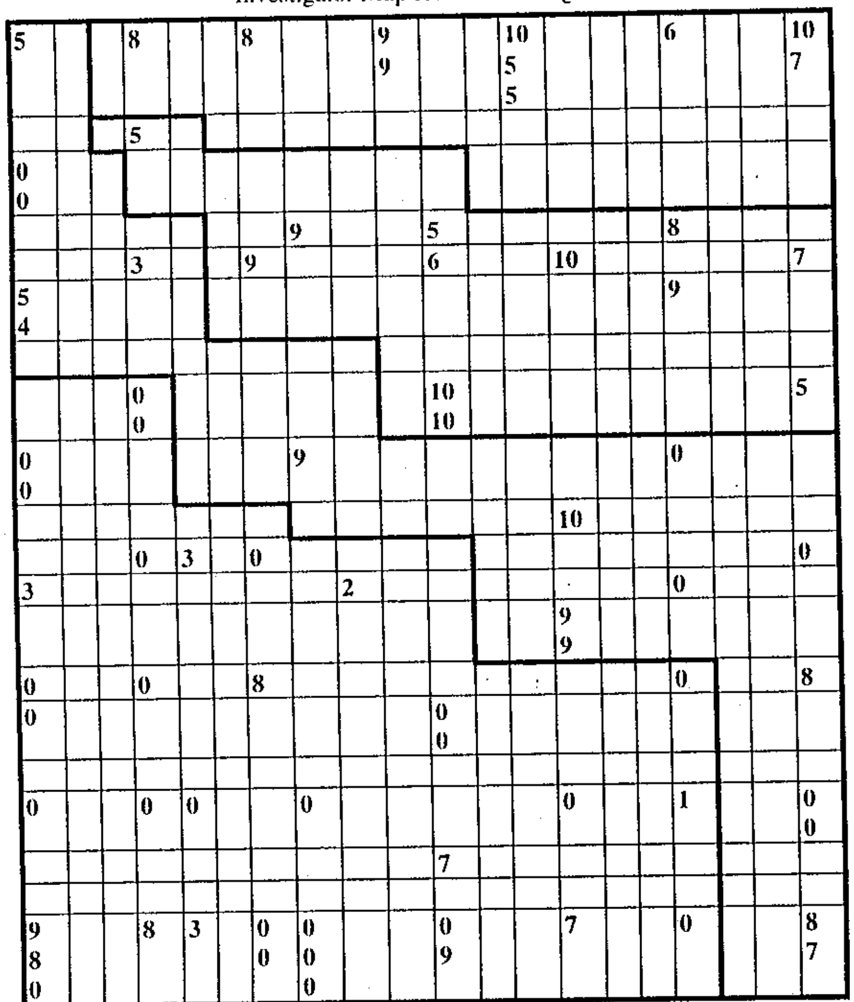
Figure 9
Adjuster Map for the Training Data



(The numbers shown in cells are the suspicion levels assigned to claims by the adjuster)

In Figure 9, however, three 8's were mapped onto the "valid" area (see the row at the bottom). It seems difficult to untangle the 8's from 0's since the claims they represent are mapped to the same output unit or nearby units. Similarly, a few 7's, 8's and 9's are spotted in the lower-left area while a few 0's sneak towards the upper-right corner in Figure 10. These may be the outliers and do not disguise the overall image. Or, these might be evidence that the adjuster and/or the investigator did not assign indicators and suspicion scores consistently within the training data set.

Figure 10
Investigator Map for the Training Data



(The numbers shown in cells are the suspicion levels assigned to claims by the investigator)

Recall that claims p_C and p_D were picked works in Section 3 to show how quantity S_m. We speculated that claim p_C might be "strongly suspicious" and that claim p_D "valid" from the perspective of that 18×18 feature map. Here, we find again from Figure 8 that claim p_C, or #12, is in the "strongly suspicious" area (see the cell in top-right corner) while claim p_D, or #69, (see the 7th cell at the bottom) is considered to be a "valid" claim from this 20×20 feature map.

Finding an overall pattern from Figures 12 and 13 (on the holdout data set) becomes difficult. Although we might still consider the upper-right corner in Figure 12 as the area for "strongly suspicious" claims and the opposite corner is for "valid" claims, the picture is much more vague, as compared to Figure 9. One of the reasons might be the imbalance of the sampling. The claims that were deemed by the adjuster as having high suspicion level (from 7 to 10) account for only 10 percent in the holdout data set while it is 25 percent in the training data set (see Table 1). Hence, the vague image in Figure 12 might be because the number of "strongly suspicious" claims in the holdout data set is not enough for the group to distinguish itself in a single random sampling. An apparent reason is that the performance in extrapolation seems logically inferior to that in interpolation. Presumably, the overall pattern is more ambiguous when the holdout data set is applied. The extrapolation image seems to be more satisfactory when the investigator's assessment is used as the label (see Figure 13). Not surprisingly, Table 1 shows that the distributions of suspicion level by the investigator are more comparable between the training data set and the holdout data set. It also seems clear that, compared to the one from the training data set, the investigator map from the holdout data set seems less ordered, as evidenced by that a few more large suspicion scores can be found in the lower-left area in Figure 13.

In Section 3, we showed that without the help of any external knowledge, a feature map model could be constructed to understand BI claims in a unique way, as shown in Figure 5 through Figure 7. In this section, we showed that with the help of the adjuster and the investigator's assessments, or the operation of an outcome feedback loop, a claim could be roughly identified or categorized on a feature map. Furthermore, a feature map provides a mechanism for evaluating those subjective assessments. For instance, we found that the adjuster and the investigator did not assign the suspicion level to those BI claims in a random fashion. The overall distributions of the claims' level of suspicion suggest that these two experts evaluated the claims' level of suspicion in a somewhat consistent, but apparently less than perfectly consistent, fashion.

Figure 11
ID Map for the Holdout Data Set Locating the Best-Matching Unit for Each Claim in the Data Set

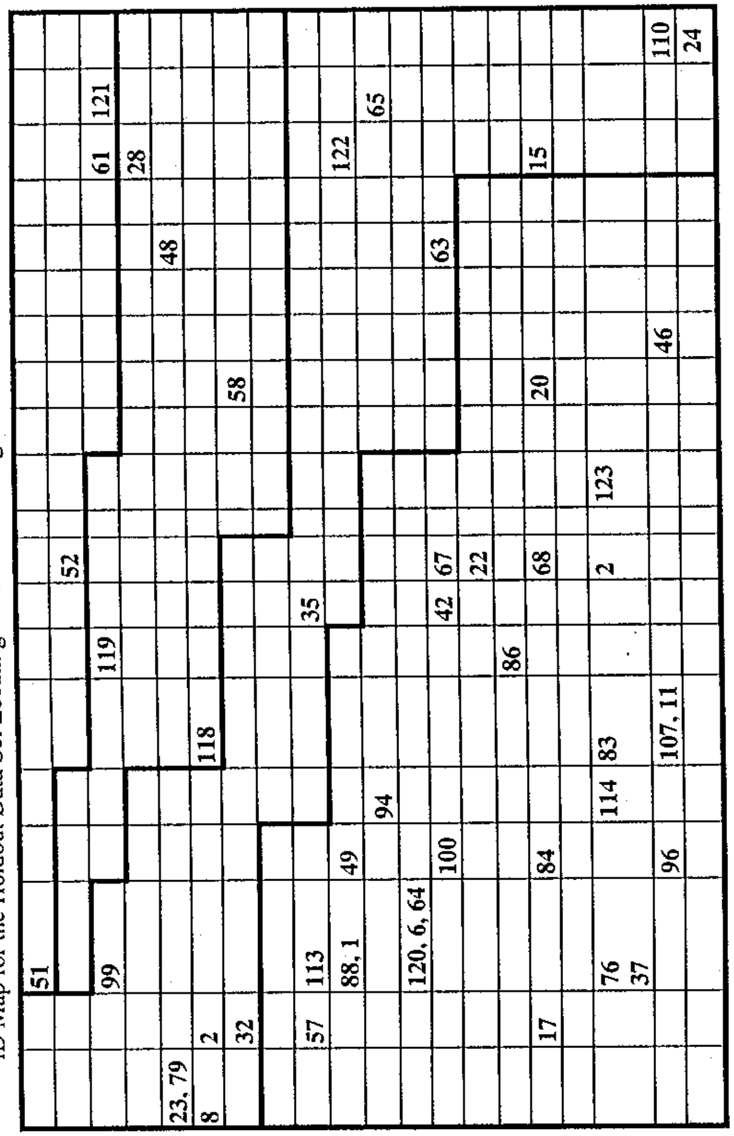
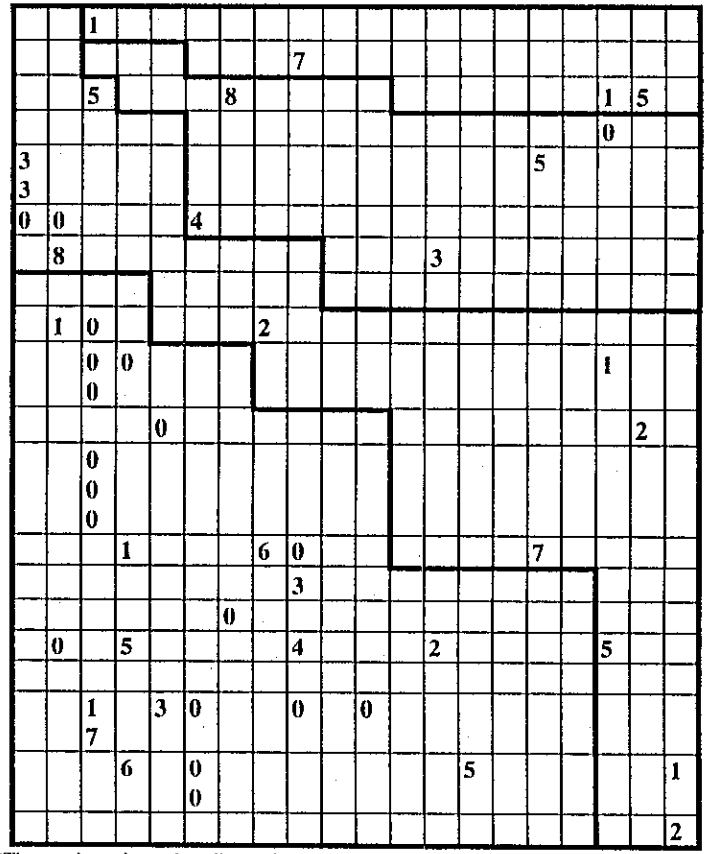
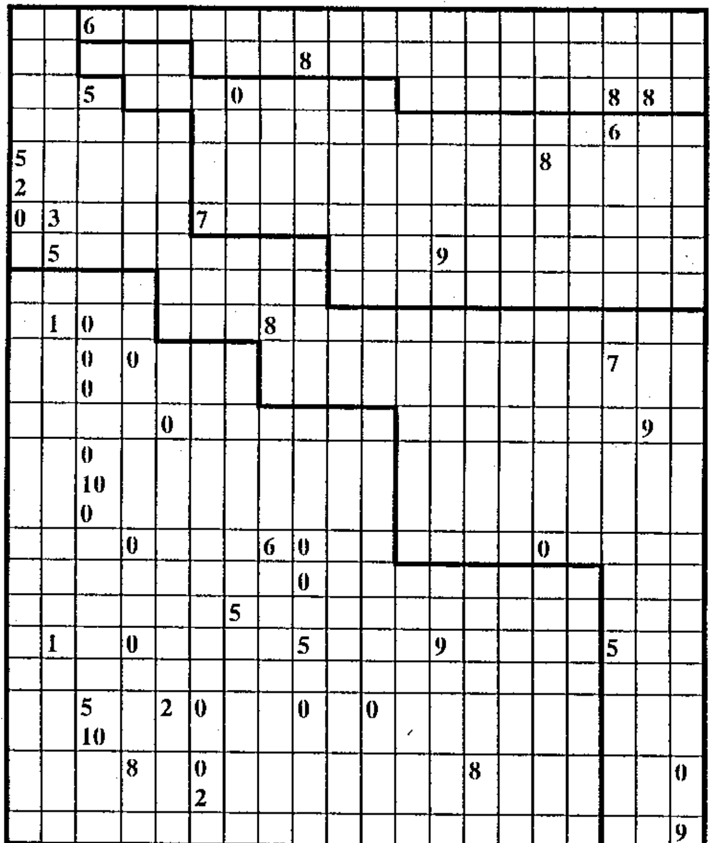


Figure 12
Adjuster Map for the Holdout Data Set



(The numbers shown in cells are the suspicion levels assigned to claims by the adjuster)

Figure 13
Investigator Map for the Holdout Data Set



(The numbers shown in cells are the suspicion levels assigned to claims by the investigator)

A Validation by Feed Forward Neural Networks

It is instructive to construct a test and compare the feature map approach with the experts' claim evaluations. As we know, the adjuster's assessment and the investigator's assessment represent the subjective suspicion level assigned to the claims. Hence, each of their two assessments becomes a dependent variable, a variable depending upon the 65 fraud indicators. To do a fair comparison, we have to construct a dependent variable from the feature map. With this dependent "suspicion" variable, the next task is to find a way of doing the quantitative comparison between the three dependent variables. Constructing a variable measuring the suspicion level of BI claims from the feature map point of view and finding a model to do a comparative study are the two things to be accomplished in this section.

Table 2
Breakdown by Feature Map Categorization 1 (FMC1)

	Training Data Set		Holdout Data Set		Total	
Level of Suspicion	Count	%	Count	%	Count	%
1	37	48%	28	56%	65	51%
2	18	23%	13	26%	31	24%
3	12	16%	5	10%	17	13%
4	10	13%	4	8%	14	11%
Total	77		50		127	

Table 3
Breakdown by Feature Map Categorization 2 (FMC2)

	Training Data Set		Holdout Data Set		Total	
Level of Suspicion	Count	%	Count	%	Count	%
1	32	42%	26	52%	58	46%
2	14	18%	11	22%	25	20%
3	13	17%	6	12%	19	15%
4	18	23%	7	14%	25	20%
Total	77		50		127	

Table 4
Breakdown by Feature Map Categorization 3 (FMC3)

	Training Data Set		Holdout Data Set		Total	
Level of Suspicion	Count	%	Count	%	Count	%
1	28	36%	25	50%	53	42%
2	9	12%	6	12%	15	12%
3	9	12%	6	12%	15	12%
4	31	40%	13	26%	44	35%
Total	77		50		127	

Table 5
Correct Classification Rates by Feed-Forward Neural Networks (1)

Sample Type	Sample Size	Adjuster's	Investigator's	FMC1	FMC2	FMC3
Train	53	100.0%	100.0%	100.0%	100.0%	100.0%
Test	24	44.8%	45.8%	79.2%	66.7%	82.3%
Train+Test	77	82.8%	83.1%	93.5%	89.6%	94.5%
Holdout	50	52.5%	43.0%	67.5%	59.5%	63.5%

Note: Four feed-forward neural network models were constructed, using different initial seeds, to emulate each of five approaches. A claim is correctly classified for an approach if the output by a neural network model matches the value given by the approach on the claim. For instance, if the adjuster considers a claim "strongly suspicious" and the output of a neural network model is 4 ("strongly suspicious") as its output for the claim, the correct classification is obtained by the neural network model on the claim. A correct classification rate for a data set equals the number of correctly classified claims divided by the total number of claims in the data set. The correct classification rates shown above are the average over four neural network models.

Feature Map Categorization

The professional assessment maps shown in Figures 9 and 10 present an overall tendency of claim suspicion level, i.e., high at the upper-right corner and low at the opposite corner. Hence, we can label the claims mapped onto the upper-right area as "strongly suspicious", those onto the lower-left area as "valid." "Moderately suspicious" and "slightly suspicious" zones can be identified from the feature map accordingly. However, only if the feature map is split into four regions, with each output unit (a cell in Figures 8 through 13) belonging to exactly one of the regions, can such a numerical variable be defined. The variable for a claim assumes a value of "4" if the claim finds its best-matching unit in the "strongly suspicious" region, "3" if in the "moderately suspicious" region, "2" if in the "slightly suspicious" region, and "1" if in the "valid" region.

We drew "borders" in Figure 8 to split the feature map into four regions. The feature map separation was made with the following considerations: First of all, the adjuster map and investigator map (Figure 9 and 10) suggest the general trend of suspicion, i.e., "strongly" at the upper right corner and "valid" at the lower left corner. Secondly, it is thought that the "strongly suspicious" claim group, in reality, accounts for roughly ten to fifteen percent of the BI claims while about half of the BI claims appear to be valid. No scientific method was used as guideline for this exercise.[7] The final categorization was thus obtained as shown in Figures 8 through 13. A feature map claim categorization variable was thus defined. We define this as the feature map categorization (FMC1). The summary on FMC1 is shown in Table 2.

[7]This assignment would be accomplished in a practical application by further testing or by a feedback loop from actual outcomes.

Table 6
Correct Classification Rates by Feed-Forward Neural Networks (2)

Approach	Feature Mapping Categorization			Adjuster's Assessment			Investigator's Assessment		
Sample Type	Train	Test	Hold-out	Train	Test	Hold-out	Train	Test	Hold-out
Sample Size	53	24	50	53	24	50	53	24	50
Sample 1 (%)	100	83	**62**	100	46	**48**	100	46	**42%**
Sample 2 (%)	100	71	**70**	100	54	**54**	100	67	**44%**
Sample 3 (%)	100	63	**52**	100	33	**60**	100	63	**42%**
Sample 4 (%)	100	67	**66**	100	46	**38**	100	46	**38%**
Sample 5 (%)	100	71	**72**	100	63	**48**	100	38	**48%**
Sample 6 (%)	100	58	**60**	100	63	**38**	100	46	**40%**
Sample 7 (%)	100	79	**60**	100	50	**40**	100	50	**42%**
Sample 8 (%)	100	71	**58**	100	58	**42**	100	58	**46%**
Sample 9 (%)	100	79	**74**	100	50	**48**	100	46	**40%**
Sample 10 (%)	100	75	**62**	100	46	**48**	100	46	**34%**
Standard Deviation	0.00	0.08	0.07	0.00	0.09	0.07	0.00	0.09	0.04
Average (%)	**100**	**72**	**64**	**100**	**51**	**46**	**100**	**50**	**42%**

Notes: Correct classification rate for this table is defined in the same way as for Table 5, except that one feed-forward neural network model was constructed for each approach and each sampling.

From Table 1, we know that the adjuster and the investigator gave very different views about the suspicion levels over the data sample. For instance, the investigator found more claims to be "strongly suspicious" than did the adjuster. One reason for the lack of concordance between the adjuster and the investigator can be explained by the manner in which different types of insurance professionals perform their job. Usually, investigators tend to either provide hard evidence in order to convict fraudulent claimants or assume validity if the claim under investigation is only slightly suspicious. Hence, the distribution of assessments given by investigators tends to concentrate on the two ends, i.e., apparently fraudulent and valid. However, insurance claim adjusters have a greater degree of freedom; for example, if they consider a claim to be even slightly suspicious, they still hold some bargaining power and can take an aggressive strategy in negotiation.

To obtain a fair comparison between the feature map categorization method and a professional's assessment, we adjusted the borders twice in this study. Based upon FMC1, we adjusted the borders in Figure 8 and Figure 11 to obtain FM Categorization No. 2, or FMC2. Specifically, we moved the three zigzag lines towards the lower-left corner to have higher densities for higher levels of suspicion categories. This adjustment was intended to produce a distribution of claim suspicion level by FMC2 (see Table 3) that is similar to one by the adjuster's assessment (see Table 1). Similarly, the zigzag lines were further moved down towards the lower-left corner in order to obtain a distribution of claim suspicion level by FMC3 (see Table 4) which approximates that of the investigators' assessment (see Table 1). Consequently, the percentage of strongly suspicious claims in FMC3 is even higher than that in FMC2. Likewise, the percentage of strongly suspicious claims in FMC2 is higher than that in FMC1.

Validation by Feed-Forward Neural Networks

Although the main purpose of this study was to search for a quantitative model for us to better identify the suspicion level of BI insurance claims, how well the model performs remains to be tested. Verification from practice is not available. Otherwise we would have used this knowledge in building (possibly different and better) models. One way to validate this new method is to see whether it performs consistently or more consistently than did the adjuster and the investigator approach.

Suppose that the learning algorithm is back-propagation (Rumelhart et al., 1986) that the transformation function is sigmoid shaped and that the number of hidden units is sufficiently large. Then, it has been proven in numerous publications (Hornik and Stinchcombe 1989; Hornik et al., 1990) that a standard feed-forward neural network with one hidden layer can be a universal approximator for virtually any function of interest. It is also widely believed that human decision making (such as providing a claim suspicion level), is both a nonlinear and continuous process. Accordingly, a three-layer feed-forward neural network, using the back-propagation algorithm, was chosen as the validation model.[8] Such a model will produce the comparative results between the feature map categorization and the adjuster's/investigator's assessments. In fact, a side benefit is that the feed-forward network can be used as an alternative approach to the feature map approach. Thus, the numerical values of the feed-forward neural network model might be more acceptable to claim processing practitioners, if interpreted as suspicion level or possibility of fraud.

We used NeuralWorks Predict (a commercial neural network software from NeuralWare, Inc.) to do the validation. All three feature map categorizations (FMC1, FMC2 and FMC3 from the 20×20 feature map) were compared to the adjuster's and investigator's assessment. NeuralWorks Predict uses sophisticated learning algorithms. One feature of the software package is that the software has a built-in genetic algorithm that helps find the optimal three-layered feed-forward neural network. Another nice feature is that the learning process is closely monitored in the sense that there are two data sets involved in the process, i.e., a training set and a testing set. The training data set is used for the purpose of training the model and the other one, the testing set, is used to test whether the training should stop. In this experiment, the 77 claims in the first random sample were further split into the training set and the testing set while the remaining 50 claims were withheld from the training process and used for the purpose of extrapolation. Remembering this point will help when reading Table 5 (and later

[8] A similar model was used in Brockett, et. al. (1994) in a study of predicting insurance insolvency.

Table 6). Table 5 presents the results of a comparison between the two professional's assessments and three feature map categorizations (FMC1, FMC2 and FMC3).

There are a few things worth mentioning in Table 5. First, it is obvious that the program was able to construct a neural network identifying all 53 claims for each of the five approaches. Second, adjusting the decision boundaries does make some difference since the correct classification rates on the testing set and the holdout set from FMC2 are lower than those from FMC1 and FMC3. On this subject, statistically sound evidence is not available although it might prove to be another interesting issue for further study. Third, the adjuster seemed to evaluate the claims more consistently than the investigator did since the correct classification rate of the adjuster's approach is 52.5 percent versus 43 percent for the investigator's approach on the extrapolation. And finally, moving boundaries in an appropriate way does not change the fact that the feature mapping approaches do have an advantage over the two professionals' subjective approaches.

To achieve more statistically significant results, we used nine other random samplings to produce training and holdout data sets in each sampling. In each sampling, fifty claims were reserved in the holdout set while the remaining seventy-seven were used as the training set for the feature map training and were split into the training set and the testing used in NeuralWorks Predict for training. We designed ten 15×15 squared feature maps, with each feature map corresponding to a sampling, and trained them for two thousand epochs.[9] Rather than experiment with multiple region separations on each feature map, we drew separation lines on each feature map only once. For each random sampling, we have three approaches to compare, i.e., the feature map, the adjuster's assessment and the investigator's assessment. Table 6 summarizes the validation results produced by NeuralWorks Predict.

From Table 6, we can basically make similar points to the ones we did on the results from Table 5, i.e., the feature map approach out performs the adjuster's approach, which out performs the investigator's approach in terms of consistency in evaluating BI insurance claims. Specifically, the feature map approach overwhelms the other two approaches in terms of correct classification rates within the combined training and testing sets and over the holdout sets (in spite of a couple exceptions). An average rate of 64 percent was achieved on the holdout data sets for the feature map while the average rates for the adjuster and investigators approaches are only 46 percent and 42 percent, respectively. Furthermore, a 74 percent correct classification rate was obtained for the feature map approach on the holdout set from sampling # 9. However, the highest rate on the holdout sets is at 60 percent for the adjuster's approach and only 48 percent for the investigator's approach. This suggests the great potential of the feature map approach in classifying BI claims, especially when one considers that an even higher rate could

[9] Here, the feature maps are smaller than what we had used in the first experiment, which is 20×20. The only reason we did this was to save resources while we hoped our conclusion would be conservative if these smaller feature maps produce consistently better results from the feed-forward neural network validation, compared to the adjuster's and the investigator's map.

be achieved if the borders to partition the square feature maps were drawn using some type of optimization rather than subjectively.

CONCLUSION AND DISCUSSION

With BI claim characteristic vectors consisting of 65 potential fraud indicators and no true dependent variable regarding fraud (or suspicion), the fraud detection problem becomes a challenging issue. The Kohonen Self-Organizing Feature Map can be used as a fraud detection or claim classification approach. Moreover, we have found that this new method out performs the insurance adjuster's and the insurance investigator's assessments of fraud for the same claim files, as verified by the multi-layered feed-forward neural network. The study presents two alternative neural network approaches in classifying BI insurance claims in terms of their suspicion level, feature maps and feed-forward networks derived from the feature maps. The first approach is capable of producing visual images for individual claims, such as a landscape as shown in Figures 1 and 2 or a planar graph shown in Figure 13 through Figure 15. A derived feed-forward neural network is able to provide numerical outputs or suspicion scores. Fuzzy clustering of the weight vectors may help here (Derrig and Ostaszewski, 1994).

There are a few observations. First, our conclusion concerning the superiority of our new methodology partially depends upon the neural network verification. Although NeuralWorks Predict is robust enough for the conclusion drawn in this comparative study, it is possible to find a better network structure and training parameter design than those used in this empirical study. For instance, we have seen in Tables 5 and 6 that the correct classification rates on all the training sets are 100 percent.

Another concern is the determination of the decision regions. Splitting the feature maps, we used not only the knowledge from the adjuster's and investigator's assessments but also the rough proportion of each suspicion level in the sample. We did an exercise in drawing three separation lines on one of the feature maps and all three feature map categorizations did produce better results than the adjuster and the investigator. However achieving the optimal feature map categorization deserves further investigation.

Concerns may be raised regarding the assumption that categorization is determined completely by fraud indicators, with all 65 indicators being equally weighted in the feature map models without data preprocessing. In the real world, some indicators might be more important than others. For example, Weisberg and Derrig (1993) show that unequal weights, ranging from one to three, arise in the regression modeling of adjuster suspicion levels by fraud indicators.[10] This concern implies a further research direction where statistical or case studies and the use of hybrid models (Goonatilake and Khebbal, 1995) may provide insights.

The correct classification rates on the holdout data sets are in the neighborhood of 64 percent (Table 6). Improvement seems to be quite possible and necessary because of the following reasons. First of all, obtaining a one percent increase in the correct classification rates means a saving of millions of dollars to the insurance industry as insurance fraud has become a multi-billion dollar business (IRC, 1997). Secondly, a preprocessing screening of fraud indicators might lead to a comparable performance by the feature map approach with a smaller number of indicators, collecting variables itself costs money. Building a comparable model using a smaller number of variables is another way of saving claim-handling costs. Thirdly, a data sample containing 127 claims is small. A model built upon a larger data sample would certainly improve the correct classification rates on the holdout set since the holdout set would be more similar to the training set as they both get larger.

[10]Weisberg and Derrig (1993). Appendix F, Table III.1, p. 156.

APPENDIX
A KOHONEN'S FEATURE MAP ALGORITHM

Structurally, a typical Kohonen's Self-Organizing Feature Map (Kohonen, 1982, 1989 and 1990) is a two-layered network. It consists of a set of input units in the input layer and a set of output units within the output layer. Assume that there is a link between each output unit and each input unit and that a weight is associated with each link, measuring the strength of the link. Then, a weight vector composed of the weights on the links to an output unit can equivalently represent the unit itself. The dimension of this weight vector equals the number of input units. Normally, output units are distributed over some topological form, such as a one-dimensional line, a two-dimensional square or a rectangle, or a higher dimensional object such as sphere or a cube, etc. As perceived, neurons in the brain are spatially distributed and a group of neurons in a certain area of the brain adapts to the environment interdependently; i.e., both the activation strengths and the topographical location of a neuron of the brain are meaningful. The learning algorithm, therefore, is designed to develop the meaningful spatial formation, i.e., a feature mapping from input patterns to output units that captures the pattern interrelationships.

Assume that there are N^2 output units for a feature map and that these units are arranged in a square form. Denote by m_{ij} the weight vector, or equivalently, the output unit at location (i, j). Suppose that the feature map learns from a training sample, which contains P patterns. A D-dimensional numerical vector represents each pattern. Denote x_p the p^{th} vector (pattern) in the sample. The learning process for a typical squared feature map is normally iterative. Each iteration is called an epoch. Let t denote the t^{th} epoch in the learning process.

Step 1. Initialization

Set epoch index, t, to 0. Set the components of all weight vectors, $m_{ijd}(t)$, to 0^+, small random numbers, for i, j = 1, ... N and d = 1, ... D. Set $\alpha(t)$ to a number smaller than but close to 1. Set w(t), the neighborhood width, to be between N/2 and N-1.

Step 2. A Three-Step Epoch

Step 2.1. If all patterns have been used in this epoch, then go to Step 3. Otherwise, select a pattern x_p from the unused patterns.

Step 2.2. Find (i_0, j_0) such that

$$\left\|x_p - m_{i_0 j_0}\right\| = \min_{i,j=1,\ldots,N}\left\|x_p - m_{ij}\right\| = \min_{i,j=1,\ldots,N}\sqrt{\sum_{d=1}^{D}\left(x_{pd} - m_{ijd}\right)^2} \qquad \text{(A1)}$$

That is, the weight vector at (i_0, j_0), has the smallest distance to the current pattern x_p. Or, unit (i_0, j_0) is the best-matching unit to pattern x_p.

Step 2.3. Go to Step 2.1 after updating the weight vectors in the neighborhood.

$$m_{ij}(t+1) = \begin{cases} m_{ij}(t) + \alpha(t)\left[x_p - m_{ij}(t)\right] & \text{if } |i - i_0|, |j - j_0| \le w(t) \\ m_{ij}(t) & \text{otherwise} \end{cases} \qquad \text{(A2)}$$

Step 3. Terminate or Continue

If $t = T$, the pre-specified stopping rule, then stop. Otherwise, decrease $\alpha(t)$ and $w(t)$ such that both are close to zero when t approaches T and return to Step 2.

REFERENCES

Artis, M, M. Ayuso and M Quillen. 1997. "Modeling Different Types of Automobile Insurance Fraud Behavior in the Spanish Market." Paper presented at the *First International Conference on Insurance: Mathematics and Economics*, Amsterdam, August 1997.

Belhadji, E. B. and G. Dionne. 1997. "Development of an Expert System for the Automatic Detection of Automobile Insurance Fraud." Working Paper 97-06, Ecole des HEC, August 1997.

Brockett, P. L., W. W. Cooper, L. L. Golden and V. Pitaktong. 1994. "A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency." *Journal of Risk and Insurance* 61:402-424.

Cummins, J. D. and S. Tennyson. 1992. "Controlling Automobile Insurance Claim Costs." *Journal of Economic Perspectives* 6: 95-115.

Dauer, C. 1993. "Automation helps to combat fraud. *National Underwriter, [Property/Casualty/Employee Benefits]* 97(3): 3.

Derrig, R. A. and K. M. Ostaszewski. 1994. "Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification." *The Journal of Risk and Insurance* 62(3): 447-482.

Derrig, R. A. and H. I. Weisberg. 1996. "Coping with the Influx of Suspicious Strain and Sprain Claims." *AIB 1997 Cost Containment and Fraudulent Claims Payments Filing*. D.O.I., Docket R96-37.

Derrig, R. A., H. I. Weisberg and X. Chen. 1994. "Behavioral Factors and Lotteries under No-Fault with a Monetary Threshold: A Study of Massachusetts Automobile Claims." *The Journal of Risk and Insurance* 61(2): 245-275.

Garcia, J. L. 1989. "Fraud: The Hidden Element in Health Care Costs." *Compensation & Benefits Management* 6(1): 49-51.

Goonatilake, Suran and Sukhdev Khebbal. 1995. *Intelligent Hybrid Systems*. Edited by S. Goonatilake and S. Khebbal. John Wiley and Sons.

Hornik, K. and M. Stinchcombe. 1989. "Multilayer Feedforward Networks are Universal Approximators." *Neural Networks* 2: 359-366.

Hornik, K., M. Stinchcombe and H. White. 1990. "Universal Approximation of Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks." *Neural Networks* 3: 551-560.

IRC - Insurance Research Council, 1984, Special Investigative Units: Surveys on Insurance Company Use of SIUs for Fraud Investigations, Oak Brook, IL.

IRC - Insurance Research Council, 1992, Fighting Insurance Fraud in the Insurance Industry, Oak Brook, IL.

IRC - Insurance Research Council, 1994, Auto Injuries: Claiming Behavior and Its Impact on Insurance Costs, Oak Brook, IL.

IRC - Insurance Research Council, 1996, Fraud and Buildup in Auto Injury Claims: Pushing the Limits of the Auto Insurance System, Oak Brook, IL.

Kohonen, T. 1982. "Self-organizing formation of topologically correct feature maps." *Biological Cybernetics* 43: 59-69.

Kohonen, T. 1989. "Self-Organizing Feature Maps." *Self-Organizing and Associative Memory*, 3rd ed. Berlin, Heidelberg, Germany: Spring-Verlag.

Kohonen, T. 1990. "The Self-Organizing Map." *Proceedings of the IEEE* 78(9): 1464-1480.

NICB - National Insurance Crime Bureau, 1994, Factpage: Battling Insurance Fraud, Palos Hill, IL.

Rumelhart, D. E., G. E. Hinton and R. J. Williams. 1986. "Learning Internal Representations by Error Back Propagation." In D. E. Rumelhart and J. L. McClelland (Eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Chap. 8)*. Cambridge, MA: MIT Press.

Schalkoff, R. 1992. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley & Sons.

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*. D.O.I. Docket G93-24.

Wilmot, R.T.D. 1991. "World Insurance Forum: London." *Best's Review, [Property/Casualty]* 92(7): 97-98.