PERGAMON

Safety Science 41 (2003) 1–14

SAFETY SCIENCE

www.elsevier.com/locate/ssci

Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea

So Young Sohn*, Sung Ho Lee

Department of Computer Science and Industrial Systems Engineering, Yonsei University, 134 Shinchon-dong, Sudaemon-ku, Seoul, South Korea

Abstract

Increasing amount of road traffic in 1990s has drawn much attention in Korea due to its influence on safety problems. Various types of data analyses are done in order to analyze the relationship between the severity of road traffic accident and driving environmental factors based on traffic accident records. Accurate results of such accident data analysis can provide crucial information for road accident prevention policy. In this paper, we use various algorithms to improve the accuracy of individual classifiers for two categories of severity of road traffic accident. Individual classifiers used are neural network and decision tree. Mainly three different approaches are applied: classifier fusion based on the Dempster–Shafer algorithm, the Bayesian procedure and logistic model; data ensemble fusion based on arcing and bagging; and clustering based on the *k*-means algorithm. Our empirical study results indicate that a clustering based classification algorithm works best for road traffic accident classification in Korea.

Keywords: Neural network; Decision tree; Dempster–Shafer; Bayesian fusion; Bagging; Arcing; Clustering algorithm

1. Introduction

A good deal of pedestrians are killed on the roads in many countries. Such road accidents continue to impose unacceptable costs on the community in both human and economic terms. Although there has been a steady reduction in the number of

* Corresponding author. Tel.: +82-2-2123-4014; fax: +82-2-364-7807.
E-mail address: sohns@yonsei.ac.kr (S.Y. Sohn).

0925-7535/02/$ - see front matter
PII: S0925-7535(01)00032-7

accidents, a comparison study with 16 other countries, using the international database on road traffic and accident (IRTAD), shows that the rate of pedestrian fatalities per total traffic fatalities in 1993 varied from more than 30% in Great Britain and Ireland to 12% in France, The Netherlands and New Zealand (Lupton and Bolsdon, 1999). These differences between countries could be due to multiple factors: urban density, road infrastructure, amount of walking done in relation to mobility in general, regulations on alcohol and speed, and an improvement in vehicle safety. In order to help prevent traffic accidents, it is necessary to answer questions about the circumstances and the characteristics of traffic accidents.

A possible avenue to identify factors related accident severity is to utilize retrospective data available such as traffic accident records (TAR). TAR typically contain information related not only to the accident severity but also to potentially influential factors on the accidents such as road facility, traffic mode, weather condition, driver characteristics, and vehicle conditions.

The main feature of TAR is that the large amount of data is recorded in the form of categorical variables with various levels. Therefore, categorical dimension reduction technique is necessary for effective pattern recognition for accident severity. Various kinds of mutivariate analyses have been used to utilize such TAR information (Oh and Ko, 1992; Fontaine and Gourlet, 1997). The most up-to-date-form of analysis would be data mining (Sohn and Shin, 2001; Vorko and Jovic, 2000). Several data mining techniques are available for association, classification and prediction. However, not much research has been done to deal with dimension reduction and pattern extraction for multivariate categorical variables comprising TAR.

Recognizing these features, Sohn and Shin (2001) employed both neural network and decision tree algorithms to find the classification model for road traffic accident severity (bodily injury or property damage) as a function of potentially related categorical factors. However, the classification accuracy of the individual algorithm was relatively low. In this case, use of data fusion or ensemble algorithms may be able to increase the accuracy. Data fusion techniques try to combine classification results obtained from several single classifiers and are known to improve the classification accuracy when some results of relatively uncorrelated classifiers are combined. Data ensemble combines various results obtained from a single classifier fitted repeatedly based on several bootstrap resamples. The resulting performance is known to be more stable than that of a single classifier. On the other hand, a clustering algorithm groups data into several categories and tries to fit a separate classification model for each cluster. This is expected to perform well when there is a significant amount of variation in data.

The main purpose of this paper is to apply data fusion, ensemble and clustering algorithms to improve the accuracy of the single classifier applied to the road traffic accident data analysis. Organization of this paper is as follows. In Section 2, we review data fusion and ensemble algorithms along with application areas. In Section 3, we introduce several fusion, ensemble and clustering algorithms for data classification. In Section 4, a case is presented to compare those algorithms. In Section 5, results are summarized and significant findings of our study are highlighted.

2. Literature review

Data fusion is a technique that combines data from multiple sensors, and related information from associated databases, in order to achieve improved accuracy and to make better inferences than could be achieved by the use of a single sensor or data set alone. Therefore, the emergence of new sensors, advanced processing techniques and hardware would make fusion more effective. While the coverage of methodological areas of data fusion systems includes artificial intelligence, pattern recognition, and statistical inference, application areas of data fusion are widespread. Military application includes automated target recognition, guidance for autonomous vehicles, remote sensing, battlefield surveillance, and automated threat recognition systems, such as identification-friend-foe-neutral (IFFN) systems. Non-military conventional application includes monitoring of manufacturing process, condition-based maintenance of complex machinery, robotics, and medical pattern recognition (Hall and Llinnas, 1997).

Modeling procedures involved in a data fusion consist of association, estimation, and identity declaration. First of all, association determines which pairs of observations belong together, by representing observations of the same entity. Commonly used association measures include correlation coefficients, distance measures, association coefficients, and probabilistic similarity measures. Next, parameters of the fusion model are estimated using maximum likelihood estimator, least square estimator, or Kalman filter estimator. Finally, for identity declaration, typically one of the following three level fusion is used: data-level fusion, feature-level fusion and decision-level fusion level (Hall and Llinnas, 1997; Kam et al., 1997).

In the data-level fusion, data from commensurate sensors are fused directly. For imagery sensors, data-level fusion is often referred to as pixel level fusion.

Feature-level fusion, referred to state-level fusion, involves the extraction of representative features from sensor data. In feature-level fusion, features are extracted from multiple sensor observations, and are combined into a single concatenated feature vector which is used as an input to pattern recognition approaches based on one of neural network, fuzzy logic, clustering algorithms or template methods.

Decision-level fusion involves fusion of sensor information, after each sensor has made a preliminary determination of an entity's location, attributes, and identity. In this paper, we are concerned with the decision level fusion in order to improve the accuracy of the severity classification of road traffic accident in Korea. Examples of decision level fusion methods include classical approaches such as Bayesian inference, Dempster–Shafer algorithm, and logistic fusion along with weighted decision methods (ensemble or voting techniques). Before we get into the main discussion, we review the application literature on data fusion.

Buede and Girardi (1997) demonstrated how Bayesian and Dempster–Shafer algorithms can address the same target identification problem involving multiple levels of abstraction, such as identification based on type, class, and nature. In the process of demonstrating target identification with these two reasoning methods, the authors compare their convergence time to a long run asymptote for a broad range of aircraft identification scenarios that include missing reports and mis-associated reports. Xufeng et al. (1997) discussed the use of decision-level fusion for improving classification accuracy of fault type of automobile transmission system. Wavelet transform is used for feature extraction from signal data. Based on these feature vectors, neural networks and grey system theory are used to diagnose fault in multiple sensor modes. Dempster–Shafer fusion algorithm is used to improve the precision of decision. Dar and Vachtsevanos (1989) proposed active perception scheme that performs classification of objects by using a subset of available features. A Dempster–Shafer theory based criterion calculated using a fuzzy classifier guides the active perception algorithm to determine the feature subset required for classification task at hand. Shaw and Garvey (1992) developed a novel signal representation based on the theory of Dempster–Shafer. This model, referred to as an evidential signal, represents many competing or overlapping discrete signal hypotheses within a single entity. Kia Information System (1997) used a fusion algorithm by way of a weighted average of driving times obtained from three sensors such as CCTV, detector and probe vehicle in order to forecast the time it takes to derive a certain segment. Blanco et al. (1999) constructed a new score of banking customer that combines both a bureau credit score and an application credit score using linear, logistic and Bayesian combination methods. A bureau credit score is obtained based only on the credit history of individuals as reported on a credit bureau report. On the other hand, an application credit score is based on the data solicited from each credit applicant at the time he or she applies for credit.

Decision level fusion methods include weighted decision methods such as ensemble or voting techniques. Data ensemble combines various results obtained from a single classifier fitted repeatedly based on bootstrap resamples. Such algorithms can be divided into two types: those that adaptively change the distribution of the bootstrap training set based on the performance of previous classifiers, as in Boosting methods or Arcing (Adaptive Resampling and Combining) and those that do not, as in Bagging (Bootstrap AGGregatING).

Optiz and Maclin (1997) evaluated Bagging and Boosting as methods for creating an ensemble of neural networks and decision tree. The authors tested both algorithms with 14 data sets. They found that Bagging is appropriate for most problems, but when properly applied, Boosting may produce even larger gains in accuracy. Quinlan (1996) reported results of applying both bagging and boosting by decision trees (C4.5) on 18 data sets. While both approaches substantially improve predictive accuracy, boosting shows the greater benefit on average. On the other hand, boosting also produces severe degradation on some data sets. Merz and Pazzani (1999) suggested Principal Component Regression (PCR) for combining the predictions of multiple classification models.

3. Fusion algorithms

This section introduces data fusion, ensemble and clustering algorithms that can improve the classification ability of the neural network and decision tree that were applied to classify the severity level of road traffic accident (Sohn and Shin, 2001). Sohn and Shin (2001) fitted both a neural network and a decision tree in order to classify accident severity into either bodily injury or property damage using categorical input variables such as road width, shape of car body, speed before accident, violent drive, and protective device. However, accuracy of the individual classifier was about 70%.

In this section, we describe various fusion (Dempster–Shafer, Bayesian and Logistic methods) and ensemble algorithms (Bagging and Arcing) used to improve the severity classification accuracy and discrimination power.

First, the Dempster–Shafer algorithm can be applied to our case as follows. We define elemental propositions for outcome of traffic accident (A_1: bodily injury and A_2: property damage) and belief functions m_{N}, m_{D} that can be proposed by two classifiers, neural network and decision-tree, respectively. The number of hypotheses for the Dempster–Shafer algorithm is the $4(=2^2)$ propositions of the power set of the elemental proposition such as (A_1: bodily injury and A_2: property damage). Also, we define the fused proposition u_l where u_1 is for bodily injury (BI), u_2 is for property damage (PD), u_3 is for PD$\bigcup$BI, and u_4 is for Φ. Then $m(u_l)$ which represents the degree of belief related to u_l, can be obtained as Eq. (1) (Hall, 1992; Choi et al., 1998).

$$
\begin{aligned}
m(u_l) &= \frac{\sum\limits_{A_i, A_j = u_l} m_{\mathrm{N}}(A_i) m_{\mathrm{D}}(A_j)}{1 - C} \\
C &= \sum_{A_k A_m = \phi} m_{\mathrm{N}}(A_k) m_{\mathrm{D}}(A_m)
\end{aligned} \tag{1}
$$

For instance, the degree of belief that the accident occurred is property damage, when it is sentenced is PD by both decision tree and neural network, can be calculated as (2), where necessary information can be found in Table 1 (Fig. 1).

$$
\begin{aligned}
m(\mathrm{PD}) &= \frac{m_{\mathrm{N}}(\mathrm{PD}) m_{\mathrm{D}}(\mathrm{PD}) + m_{\mathrm{N}}(\mathrm{PD}) m_{\mathrm{D}}(\mathrm{PD} \bigcup \mathrm{BI}) + m_{\mathrm{N}}(\mathrm{PD} \bigcup \mathrm{BI}) m_{\mathrm{D}}(\mathrm{PD})}{1 - C} \\
C &= m_{\mathrm{N}}(\mathrm{PD}) m_{\mathrm{D}}(\mathrm{BI}) + m_{\mathrm{N}}(\mathrm{BI}) m_{\mathrm{D}}(\mathrm{PD})
\end{aligned} \tag{2}
$$

Table 1
Dempster's combining rule

Decision tree	Neural net		
	m_{N}(BI)	m_{N}(BI)	m_{N}(BI∪PD)
m_{D}(BI)	m_{N}(BI)×m_{D}(BI)	m_{N}(BI)×m_{D}(BI)	m_{N}(BI∪PD)×m_{D}(BI)
m_{D}(BI)	m_{N}(BI)×m_{D}(BI)	m_{N}(BI)×m_{D}(BI)	m_{N}(BI∪PD)×m_{D}(BI)
m_{D}(BI∪PD)	m_{N}(BI)×m_{D}(BI∪PD)	m_{N}(BI)×m_{D}(BI∪PD)	m_{N}(BI∪PD)×m_{D}(BI∪PD)

Other kinds of degrees of belief, m(BI), m(PD$\bigcup$BI), can be calculated in the same way as (2). Consequently, accident is classified as PD when m(PD) is greater than the degrees of belief of other kinds.

In Bayesian fusion algorithm, Bayesian inference updates the likelihood of a hypothesis given a previous likelihood estimate and additional evidence (observations). The technique may be based on either classical probabilities or subjective probabilities.

Generally, this technique assumes that propositions are mutually exclusive. When applied to our case, posterior probability for fusing the result of each classifier is defined as follows (Hall, 1992).

$$P(u_l|I_{\mathrm{N}}, I_{\mathrm{D}}) = \frac{P(u_R)P(I_{\mathrm{N}}|u_R)P(I_{\mathrm{D}}|u_R)}{\sum_{j=1}^{n} P(u_j)P(I_{\mathrm{N}}|u_j)P(I_{\mathrm{D}}|u_j)} \tag{3}$$

Using Eq. (3), probability that the accident occurred is related to property damage, when it is sentenced as bodily injury by decision tree and property damage by neural network, P(PD|$\mathrm{PD_N}$, $\mathrm{BI_D}$) can be obtained as follows:
where P(PD) and P(BI) can be replaced with m(PD) and m(BI) obtained from Dempster–Shafer algorithm. $P(\mathrm{BI}_N|\mathrm{BI})$ represents the probability that neural network classifies it as bodily injury when the severity of accident was really bodily injury.

$$\frac{P(\mathrm{PD})P(\mathrm{PD_N}|\mathrm{PD})P(\mathrm{BI_D}|\mathrm{PD})}{P(\mathrm{PD})P(\mathrm{PD_N}|\mathrm{PD})P(\mathrm{BI_D}|\mathrm{PD}) + P(\mathrm{BI})P(\mathrm{PD_N}|\mathrm{BI})P(\mathrm{BI_D}|\mathrm{BI})} \tag{4}$$

where n = number of class R = selected proposition (R = 1,2) I = outcome of classifier (Neural Network, Decision Tree) $P(u_R)$ = degree of belief for $u_R(m(u_R))$ $P(I_{\mathrm{N}}|u_R)$ = degree of belief that is related to I when the accident occurred is u_R.

This kind of probability is obtained from the retrospective data. After data fusion, we compare $P(\mathrm{PD}|\mathrm{PD_N},\mathrm{BI_D})$ to $P(\mathrm{BI}|\mathrm{PD_N},\mathrm{BI_D})$ and sentence the severity of accident as the one associated with higher probability.

The third approach is to use a logistic fusion method (Blanco et al., 1999). A logistic regression model for fusion can be formed using individual results of classifiers

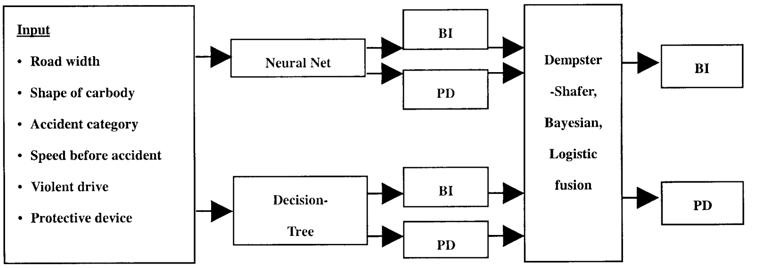


Fig. 1. Data fusion for the severity of a road traffic accident.

as input variables. That is the probability that the accident severity is sentenced as A_1 (bodily injury) is a logistic function of both the classification result due to neural network (I_N) and that of decision-tree (I_D):

$$P(A_1|I_N, I_D) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 I_N - \beta_2 I_D)} \tag{5}$$

where I_N and $I_D = 1$ if the accident occurred is classified as bodily injury, otherwise 0.

For instance, the probability that the accident occurred is bodily injury, when it is sentenced as BI by decision tree and PD by neural network becomes

$$P(PD|PD_N, BI_D) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \times 0 - \beta_2 \times 1)} \tag{6}$$

In addition, there are many other fusion approaches. When there are two classifiers, the following rules can be applied (Madanl et al. 1998): $\max(0,P(I_N) + P(I_D)-1)$, $P(I_N)P(I_D)$, $\min(P(I_N),P(I_D))$, $2P(I_N)P(I_D)/(P(I_N)+P(I_D))$, $\mathrm{root}(P(I_N)P(I_D))$, $(P(I_N)+P(I_D))/2$, $\max(P(I_N),P(I_D))$, $P(I_N)+P(I_D)-P(I_N)P(I_D)$.

So far we have shown how classical fusion algorithms can be applied to the severity classification for road traffic accidents. We now turn to the ensemble algorithm where a single classifier is repeatedly fitted based on bootstrap resamples of training data. The advantage of ensemble lies in the possibility that the difference of result caused by the variance of input data may be reduced by combining each classifier's output. Bagging and Arcing are popular methods for producing classifier ensembles (Breiman, 1996, 1998; Opitz and Maclin, 1997).

In this paper we evaluate Bagging and Arcing as methods for creating an ensemble of classifiers. Bagging perturbs the training set repeatedly to generate multiple predictors and combines by simple voting (classification) or averaging (regression). Bagging is a "bootstrap" ensemble method that creates individuals for its ensemble by training each classifier on a random redistribution of the training set. For instance, the accident occurred is presented by value 1 or 0 when the outcome of an individual classifier based on each bootstrap resample is sentenced as BI or PD. After many bootstrap resamples are applied, the final result for one case is classified as the most frequent class of the severity of accident voted by each individual classifier. The steps involved in Bagging are as follows (Breiman, 1996; Choi et al., 1999).

1. Suppose $C(x,t)$ is a classifier, producing an output with a dummy variable distinguishing Bodily Injury (one) from Property Damage (zero), at input point t.
2. To bag C, we draw bootsrap samples $x^{*1},\ldots,x^{*B}$ each of size N with replacement from the training data.
3. Classify input point t to the class with largest "vote" in C_{bag} (Fig. 2).

$$C_{bag}(t) = \frac{1}{B}\sum C(x^{*b}, t) \tag{7}$$

Arcing is a more complex procedure. Again, multiple classifiers are constructed and vote for classes, but the construction is sequential, with the construction of the $(b+1)$st classifier depending on the performance of the b previously constructed classifiers.

The basic idea in Arcing is like Bagging, but it takes unequal probability bootstrap samples; that is, puts more weight on observations that are misclassified in order to make the classifier work harder on those points. The steps of Arcing are as follows:

1. Sample with replacement from training set T with probabilities $P_b(i)$ and construct the classifier C_b using the resampled set T_b of size N.
2. Classify T_b using C_b and let $d(i)=1$ if example i is incorrectly classified, else $d(i)=0$.
3. Let

$$\varepsilon_b = \sum_{i=1}^{N} P(i)d(i), \quad \beta_b = \frac{(1-\varepsilon_b)}{\varepsilon_b} \tag{8}$$

4. Update probabilities by

$$P_{b+1}(i) = \frac{P_b(i)\beta_b^{d(i)}}{\sum P_b(i)\beta_b^{d(i)}} \tag{9}$$

5. Let $b=b+1$ and go to 1 if $b<B$.
6. Take a weighted vote of the classification, with weight $\log(\beta_b)$.

After B steps, the $C_1,\ldots,C_B$ are combined using weighted voting with C_B having weight $\log(\beta_b)$. There are two minor revisions to this algorithm. If ε_b becomes equal

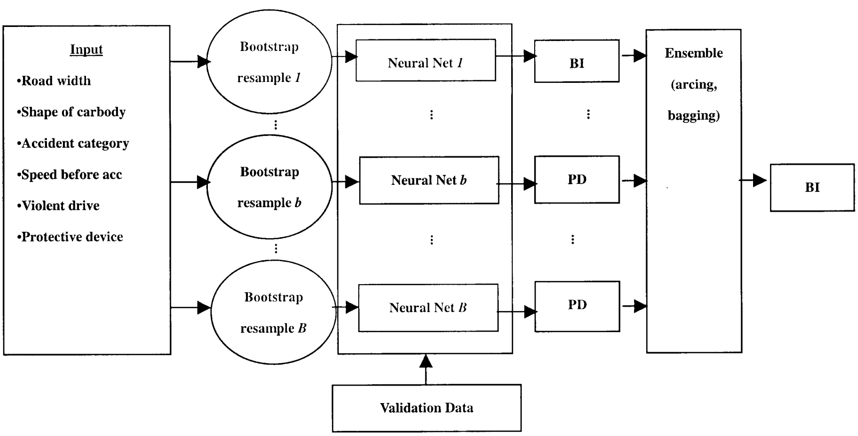


Fig. 2. Emsemble for the severity of a road traffic accident.

to or greater than 1/2, then the voting weights $\log(\beta_b)$ go negative. It is found that good results were then obtained by setting all p(i) to be equal and restarting. If ε_b equals zero, making the subsequent step undefined, we again set the probability equal and restart (Breiman, 1998).

These kinds of ensemble approaches are known to stabilize the classification results by reflecting the variation within a data set. On the other hand, when the variation in observations is relatively large, classification based on clustering would work better than ensemble, by clustering the data set first and fitting a classification model for each cluster accordingly. We propose to use a k-means clustering algorithm to group a training data set. Next, classifier C_{k} is applied to each cluster k where $k=1,\ldots,K$. When a new case needs to be classified, all we have to do is to assign this case into one of K existing clusters. Subsequently, classification result is given following the corresponding classifier for a given cluster (Fig. 3).

4. Case study

Sohn and Shin (2000) used individual algorithms such as a neural network and a decision tree to classify the severity of road accidents occurred in Seoul, Korea in 1996. Input variables used for classification of two levels of severity (bodily injury and property damage) are road width, shape of car body, accident category, speed before the accident, violent drive, and protective device. These variables were selected using the decision tree and all turned out to have better explanatory power than variables representing weather conditions. A sample of 11 564 accidents was taken and 60% of them were used for training while the rest of them were used for validation, respectively. Correct classification rates obtained by both classification models were not significantly different. In order to increase not only the classification accuracy but also discrimination power, we use several fusion algorithms introduced in the previous section.

The discrimination power of the neural network to Bayesian procedure is defined as the difference between degree of belief of the neural network and that of Bayesian procedure. That is, discrimination power is the gain of reliability of a classifier to that of another classifier. For instance, we took a sample of size nine accidents, and calculated the probability of each accident being property damage as displayed in

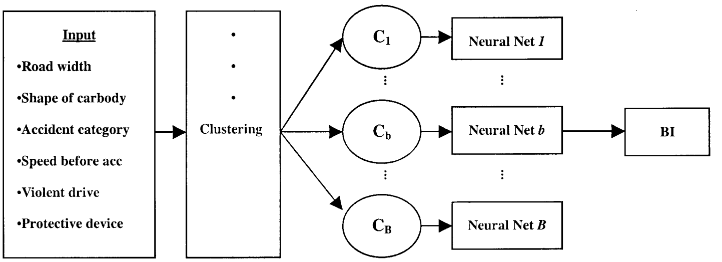


Fig. 3. Clustering for the severity of a road traffic accident.

Fig. 4. In case 1, such probability was obtained as 0.62, 0.60, 0.52, 0.96 by the Dempster–Shafer algorithm, neural network, decision tree, and Bayesian procedure, respectively. One can say that in this case, the Dempster–Shafer algorithm and the Bayesian procedure have discrimination gain by 0.1 and 0.44 compared to the decision tree, respectively.

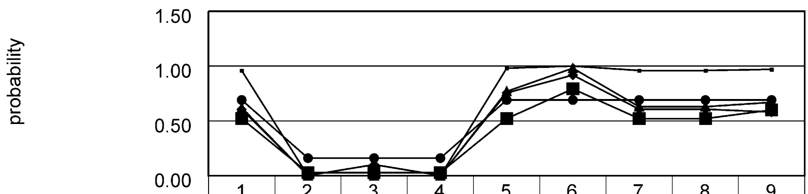
As a result of test data analysis for discrimination power, Bayesian turns out to be the best by showing average gain by 13.81, 17.67, 22.56, and 22.52% compared to the Dempster–Shafer, neural network, decision tree, and logistic fusion, respectively as displayed in Table 2.

In Fig. 4, it is shown that fusion algorithms such as the Dempster–Shafer, Bayesian procedure and logistic model provide greater (smaller) value than 0.5 when the classification probability for property damage is greater (smaller) than 0.5 by the neural network and decision tree. However, fusion algorithms provide less discriminative values when the classification results of both the neural network and the decision tree are contradicting each other. We do not include ensemble and clustering algorithms for discrimination power comparison since they simply provide classification results instead of probabilities.

Next, we compare each algorithm in terms of classification accuracy using test data. We first compare single classifiers to fusion algorithms such as the Dempster–Shafer algorithm, the Bayesian procedure and the logistic model. Information in Table 3 is used as input to the Baysian procedure while the following logistic regression model is fitted for the logistic fusion based on the individual classification results (I_N, I_D) along with actual realization (A_1):

$$P(A_1|I_N, I_D) = \frac{1}{1 + \exp(4.1343 - 0.2796 I_N - 2.1767 I_D)} \tag{10}$$

According to detailed results given in Table 4, the Dempster–Shafer algorithm and the logistic model appear to have highest classification accuracy over individual classifiers.



	1	2	3	4	5	6	7	8	9
Neural Net	0.60	0.00	0.10	0.00	0.75	0.92	0.61	0.61	0.58
Decision-Tree	0.52	0.03	0.03	0.03	0.52	0.79	0.52	0.52	0.60
Dempster-Shafer	0.62	0	0	0	0.77	0.98	0.63	0.63	0.67
Bayesian	0.96	0	0	0	0.98	1	0.96	0.96	0.97
Logistic	0.69	0.16	0.16	0.16	0.69	0.69	0.69	0.69	0.69

Neural Net, Decision-Tree, Dempster-Shafer, Bayesian, Logistic

Fig. 4. Probability of being classified as property damage based on each technique.

We analyze the mis-classifcation pattern of each classifier in Tables 5–7. The total number of mis-classifications due to the Bayesian procedure was 998 out of a total of 3469 test cases. When both the neural network and the decision tree classify the outcome correctly, there was no mis-classification due to the Bayesian procedure. However, there were about 77% of mis-classification cases when both the neural network and the decision tree had wrong classification results. The Dempster–Shafer algorithm also showed similar pattern to that of Bayesian procedure as displayed in Table 6.

Table 2
Comparison of discrimination power of other classifiers to Bayesian procedure

Single classifier	Decision tree	−22.56%
	Neural network	−13.81%
Fusion technique	Dempster–Shafer	−17.67%
	Bayesian	0
	Logistic	−22.52%

Table 3
Conditional probability used for the Bayesian method

Neural network		Decision tree	
$P(PD_N \vert PD)$	$P(BI_N \vert BI)$	$P(PD_D \vert PD)$	$P(BI_D \vert BI)$
0.67	0.85	0.66	0.82
$P(BI_N \vert PD)$	$P(PD_N \vert BI)$	$P(BI_D \vert PD)$	$P(PD_D \vert BI)$
0.33	0.15	0.34	0.18

Table 4
Classification accuracy of each algorithm

	Accuracy (%)	The number of classifier
Decision tree	72.30	1
Neural network	70.86	1
Dempster–Shafer	72.79	2
Bayesian	71.23	2
Logistic fusion	72.30	2
Bagging (neural net)	72.70	5
	72.41	10
Bagging (decision tree)	74.78	5
	73.80	10
Clustering method (neural net)	73.94	3
Clustering method (decision tree)	76.10	3

Logistic fusion provided 961 out of 3469 mis-classification cases as in Table 7. Unlike both the Bayesian and the Dempster–Shafer algorithm, the percent of mis-classification of PD\BI\PD\BI, BI\PD\BI\PD appeared to be 0%. However, the overall misclassification pattern was similar to those two fusion algorithms.

In the previous section, we suggested using data ensemble and clustering algorithms to avoid the tendency of the mis-classification patterns potentially generated by general fusion algorithms when individual classifiers do not perform well. Next, we apply those algorithms in an effort to improve the accuracy rates of fusion algorithms.

We first use ensemble approach by generating both five and 10 bootstrap resamples and applying them in order to fit both neural network and decision trees for bagging purpose. Classification accuracy of bagging with five bootstrap resamples applied to decision tree turned out to be 74.78% while that with 10 resamples

Table 5
Misclassified cases by the Bayesian method

Real\Neural network\Decision tree\Bayesian			
PD\BI\BI\BI	23.35%	PD\BI\PD\BI	8.12%
BI\PD\PD\PD	53.91%	BI\BI\PD\PD	2.20%
BI\BI\BI\PD	0%	PD\PD\BI\BI	10.52%
PD\PD\PD\BI	0%	BI\PD\BI\PD	1.90%

Table 6
Misclassified cases by the Dempster–Shafer method

Real\Neural network\Decision tree\Dempster–Shafer			
PD\BI\BI\BI	24.68%	PD\BI\PD\BI	6.57%
BI\PD\PD\PD	56.99%	BI\BI\PD\PD	2.33%
BI\BI\BI\PD	0%	PD\PD\BI\BI	7.42%
PD\PD\PD\BI	0%	BI\PD\BI\PD	2.01%

Table 7
Misclassified cases by the Logistic fusion method

Real\Neural network\Decision tree\Logistic fusion			
PD\BI\BI\BI	24.25%	PD\BI\PD\BI	0%
BI\PD\PD\PD	55.98%	BI\BI\PD\PD	10.52%
BI\BI\BI\PD	0%	PD\PD\BI\BI	9.56%
PD\PD\PD\BI	0%	BI\PD\BI\PD	0%

slightly decreased to 74.50%. With the use of the neural network, accuracy decreased a little bit more. In order to obtain the classification accuracy of arcing, we use validation data. Regardless of the number of resamples and individual classifiers used, results came out almost the same as those of bagging.

Next, we use cluster algorithm which considers group data characteristics for classification. When applying the k-means clustering algorithm, we group input data into three. For each cluster, both the neural network and the decision tree were fitted for classification. The resulting classification accuracy is obtained as improved ones over fusion algorithms. Note that all calculations are done using SAS E-miner and Excel.

5. Conclusion

We compare several fusion algorithms to ensemble algorithms and a clustering algorithm in terms of classification accuracy and discrimination power using road accident data. As shown in the previous section, fusion algorithms displayed better discrimination power than a single classifier. Among them, the Bayesian procedure turned out to be the best in that sense. In terms of classification accuracy, the Dempster–Shafer algorithm appears to improve the classification performance of not only individual algorithm such as the neural network and the decision tree but also fusion algorithms including the Bayesian and logistic fusion. But this improvement was marginal. Apparently, ensemble algorithms such as bagging and arcing also showed improvement in classification accuracy but were not better than that of clustering based classification. This kind of interpretation may be limited to the case to which these algorithms were applied. That is, when the variation in observations is relatively large as we have in Korean road traffic accident data, classification based on clustering works better than ensemble. Therefore, it is suggested to cluster the accident data set first and fit a classification model for each cluster accordingly. This implies that safety policy would not be homogeneous overall. The generalization of comparison and use of multivariate algorithms as a tool for fusion are left as further study topics of extension of our current study. In addition, it will be worthwhile to compare performances of both fusion and clustering to that of a human expert.

Acknowledgements

This work was supported by grant No. (1999-1-303-005-3) from the interdisciplinary research program of the KOSEF.

References

Blanco, Y., Zhu, H., Peter, A.B., 1999. A study in the combination of two consumer credit scores. Decision Sciences Institute 5th International Conference 1, 558–561.

Breiman, L., 1996. Bagging, Boosting, and C4.5. ftp://ftp.stat.berkeley.edu/pub/users/breman
Breiman, L., 1998. Arcing classifiers. The Annals of Statistics 26 (3), 801–849.
Buede, D.M., Girardi, P., 1997. A target identification comparison of Bayesian and Dempster–Shafer multisensor fusion. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans 27 (5), 569–577.
Choi, D.B., Ko, H.S., Ahn, B.H., 1998. On multisensor data fusion using attribute association for intelligent traffic congestion information inference. 5th World Congress on Intelligent Transport Systems.
Choi, D.W., Ku, J.Y., Park, H.J., Park, J.S., 1999. On the improvement of classification accuracy using combined learners. Special Interest Group on Data Mining Seminar, Seoul, Korea.
Dar, I.M., Vachtsevanos, G., 1989. Feature level sensor fusion for pattern recognition using an active perception approach. Proceedings of International Conference on Telecommunications 4, 60–165.
Fontaine, H., Gourlet, Y., 1997. Fatal pedestrian accident in France: a typological analysis. Accident Analysis and Prevention 29 (3), 303–312.
Hall, D.L., 1992. Mathematical Techniques in Multisensor Data Fusion. Artech House, Boston, London.
Hall, D.L., Llinnas, J., 1997. An introduction to multisensor data fusion. Proceedings of IEEE 85 (1), 6–23.
Kam, M., Zhu, X., Kalata, P., 1997. Sensor fusion for mobile robot navigation. Proceedings of IEEE 85 (1), 108–119.
Kia Information System (KIS), 1997. The Development of Data Fusion Algorithm. Technical report. Aju University Traffic Research Center.
Lupton, K., Bolsdon, D., 1999. An object-based approach to a road network definition for an accident database. Computers, Environment and Urban Systems 23, 383–398.
Madanl, K., Cheblra, A., Bouchefra, K., Maurln, T., Reynaudmn, R., 1998. Hybrid neural-based decision level fusion architecture: application to road traffic collision avoidance. Society of Photo-Optical Instrumentation Engineers 37 (2), 370–377.
Merz, C.J., Pazzani, M.J., 1999. A principal components approach to combining regression estimates. Machine Learning 36, 9–32.
Oh, Y.K., Ko, Y.S., 1992. Discrimination model of impact factors on traffic accident. Journal of Korean Society of Transportation 10 (3), 173–180.
Opitz, D.W., Maclin, R.F., 1997. An empirical evaluation of bagging and boosting for artificial neural networks. International Conference on Neural Networks 3, 1401–1405.
Quinlan, J.R., 1996. Bagging, boosting, and C4.5. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence Available at: http://www.cse.unsw.edu.au/~quinlan/
Shaw, S., Garvey, T., 1992. Evidential signal processing for low-level sensor fusion. Proceedings of the 1992 International Conference on Acoustics, Speech, and Signal Processing 4, 61–64.
Sohn, S.Y., Shin, H.W., 2001. Data mining for road traffic accident type classification. Ergonomics 44, 107–117.
Vorko, A., Jovic, F., 2000. Multiple attribute entropy classification of school-age injuries. Accident Analysis and Prevention 32, 445–454.
Xufeng, P., Lien, G., Xiaolei, L., 1997. The research of automobile transmission system fault diagnosis based on vibration signal. The 9th International Pacific Conference on Automotive Engineering, pp. 117–123.