A decision making system to automatic recognize of traffic accidents on the basis of a GIS platform

S. Savaş Durduran *

Selcuk University, Engineering of Faculty, Department of Geomatic Engineering Campüs/Konya, Turkey

A R T I C L E   I N F O

A B S T R A C T

The prediction of traffic accidents is one of most important issues in our life. In the prediction of traffic accidents, a GIS platform to extract the important features including day, temperature, humidity, weather conditions, and month of occurred traffic accidents has been used. In this study, a decision making system (DMS) based on correlation-based feature selection and classifier algorithms including support vector machine (SVM) and artificial neural network (ANN) has been proposed to predict the traffic accidents identifying risk factors connected to the environmental (climatological) conditions, which are associated with motor vehicles accidents on the Konya–Afyonkarahisar highway with the aid of geographical information systems (GIS). Locations of the motor vehicle accidents are determined by the dynamic segmentation process in ArcGIS 9.0 from the traffic accident reports recorded by District Traffic Agency. In this DMS, firstly the number of dimension of traffic accidents dataset with five features (ay, temperature, humidity, weather conditions, and month of occurred traffic accidents) has been reduced from 5 to 1 feature by using correlation-based feature selection (CFS). In CFS method, the correlation coefficients between five features and outputs (the cases of without accident or with accident) has been calculated and chosen the feature that has highest correlation coefficient. Secondly, the traffic accident cases with one feature have been classified as without accident or with accident using SVM and ANN models. The proposed DMS has obtained the prediction accuracy of 61.79% with ANN classifier and achieved the prediction accuracy of 67.42% using SVM with RBF (radial basis function) kernel. These results have indicated that the proposed DMS could be used on prediction of real traffic accidents.

1. Introduction

The quantity of traffic accidents whose moral and material losses have reached to unbearable levels, increase day by day and the locations of these accidents are very important (Müge & Şenkal, 1999). Today, the most negative result of developing transportation systems is traffic accidents with injuries and loss of lives. The tremendous social and economic costs associated with traffic accidents have led many road authorities and researchers to establish safety management programs that aim to continually improve the safety performance of highways (Sawalha & Sayed, 2006). So, traffic safety is the most critical issue in agencies' transportation strategy. The identification of safety deficient areas on the highways is aimed to implement precautionary measures and provisions by researchers and traffic officials (Erdogan, Yilmaz, Baybura, & Gullu, 2007).

Geographical information system (GIS) technology is becoming an increasingly popular tool for visualization and analyses of accident data in highways. GIS has the ability to hold a vast amount of data that can be easily stored, shared analyzed and managed. It provides a platform for spatial data analyses and visualization to explore relationships between spatial and non-spatial data (Erdogan et al., 2007).

The success of safety improvement programs in reducing accident occurrence depends on the methods used in the accident analyses. Today, many researchers were used different deterministic and statistical methods in the studies that aiming to determine the high rate accident locations and safety deficient areas on the highways (Erdogan et al., 2007; Levine, Kim, & Lawrence, 1995; Loo, 2006).

The purpose of the road analysis is to determine the dangerous road sections, to take precautions that are suitable for those sections in order to prevent the accidents and diminish losses. The first stage in diminishing losses due to accidents is to determine the section where precaution needs to be taken. The suitable precaution can be taken only after the correct determination of this section. However, in order to make this determination correctly, a determination method that is suitable to the traffic conditions should be used. Every country uses a method that is suitable to

* Tel.: +90 332 223 1909; fax: +90 332 241 0635.
E-mail address: durduran@selcuk.edu.tr

itself in order to determine the dangerous sections (Müge & Şenkal, 1999). Many of these researches have explored the relationships between traffic accidents and geometric design and operation of road segments. However, data relating to accidents are widely available, but have received surprisingly little analyses with respect to weather. The complexity involved in establishing the exact cause-and-effect relationship in traffic accidents acts as an obstacle, because road accidents are the results of an intricate driver–vehicle–environment matrix (Andreescu & Frost, 1998). So, it is aimed to decrease the accidents determining the effects of weather and some environmental conditions on traffic accidents in Konya–Afyonkarahisar highway with the aid of GIS and artificial intelligence.

A novel decision making system based on correlation-based feature selection and classification algorithms including support vector machine and artificial neural network for predicting the traffic accidents Konya–Afyonkarahisar highway in Turkey with the aid of GIS. The used traffic accidents dataset comprises five features including day, temperature, humidity, weather conditions, and month of occurred traffic accidents and comprises 378 data points (179 without traffic accident and 179 with traffic accident). In order to select the significant feature from dataset and to reduce the complexity of classification algorithms, the correlation-based feature selection method has been applied to traffic accidents dataset and selected the first feature of dataset that is most related to class cases of traffic accidents dataset in the end of this process. After feature selection process, the traffic accidents dataset with one feature (day attribute) has been assigned to either without accidents or with accident via SVM or ANN classifiers. Thanks to this decision making system, the case of traffic accident according to obtained information from a GIS could be determined.

The rest of the paper is arranged as follows. The material is described in the next section. Section 3 presents the proposed decision making system. The experimental data and results to present the effectiveness of proposed method are given in Section 4. The conducted conclusions and discussions are given in Section 5.

2. Materials

2.1. Traffic accident dataset

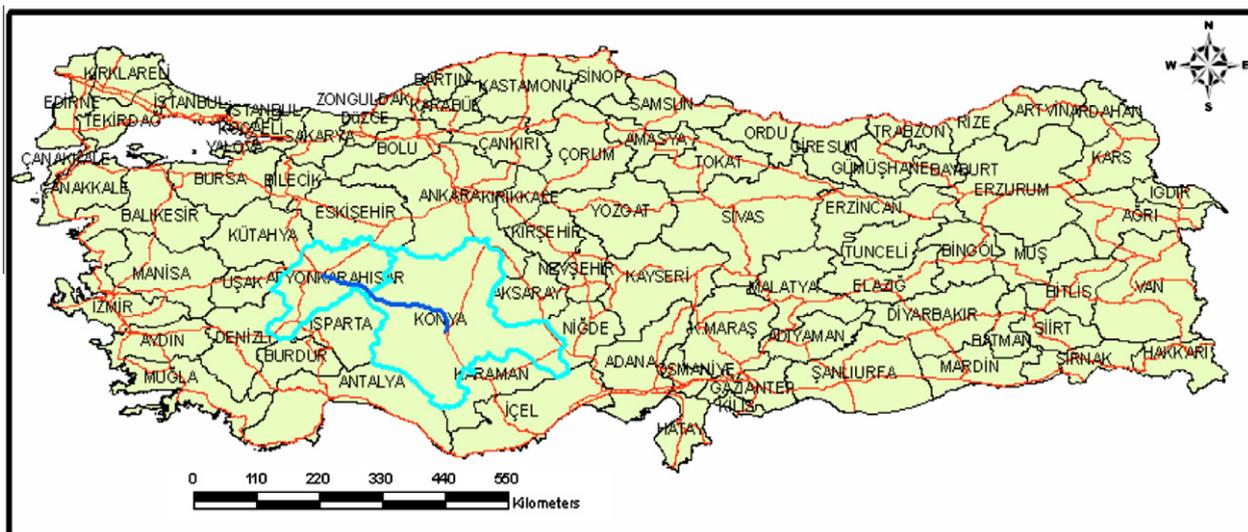Afyonkarahisar–Konya highway is a junction region in Turkey connecting the industrial, tourism and agricultural areas to each others. Especially in winter, weather related crashes happen frequently because of continental climate in the region (see Fig. 1). The length of the highway is approximately 240 km 65% of the road has two lanes. Traffic accident reports belong the highway are obtained on paper form by District Traffic Agency Officers in Turkey. These records include collected accident parameters such as the date, hour/minutes, kilometer of crash, code of highway, age, sex and alcohol consumption of driver, weather conditions, lighting conditions, vehicle type, and number of persons injured/killed. Table 1 presents the statistical values of attributes of raw traffic accident dataset. The important thing here is to select the correct database and keeping the accident data correct, updated and complete. In the database, the effects of road situation, environmental conditions and the vehicle status, on the accident will be determined. Besides, coordinates will be determined by GPS and the spatial situation will be correlated with the map.

Using this accident data a GIS-based study performed for the purpose of reducing the number of the accidents by determining the effects of environment and weather conditions on the phenomenon of accidents. First, 179 accident and 179 not accident records belong to 2006 year and were collected from the District Traffic Agency and input into an MS Access database. Meanwhile, highway was digitized at a scale of 1:1,000,000 with ArcGIS 9.0 software. The location of the accidents positioned with the "kilometer of crash" data on the route of highways using the "linear referencing" toolset in ArcGIS 9.0. Making the accident analysis in the region by taking advantage of the database created by the correlation of regional numerical map and accident data, provides a good infrastructure for the prevention of the accidents.

Table 1
The statistical values of attributes of raw traffic accident dataset.

Attribute number	Minimum	Maximum	Mean	Standard deviation
1	1	7	4.23	1.97
2	−4	29	14.36	8.31
3	20	91	53.53	21.74
4	1	5	2502	1.29
5	1	12	7.74	3.16

Fig. 1. Konya–Afyonkarahisar highway in Turkey.