Predicting and Identifying Hot Spots by Applying Neuro- Fuzzy Systems

MANSOUR HADJI HOSSEINLOU[1]
MAHDI SOHRABI MOSHTAGHIN[2]
Transportation Dept., Faculty of Civil Eng.
K.N.Toosi University of Technology
Vali_Asr St., Mirdamad Cr., Tehran, Iran, 1996715433
IRAN

Abstract: -Providing safety in roads for the purpose of protecting human assets and preventing social and economic losses resulted from road accidents is a significant issue. Identifying the hot spots of the roads provides the possibility of promoting safety quality within optimization and safety for the roads. Identifying these weak points is also related to investigate frequency and intensity of occurred accidents. Accidents are multidimensional and complicated events. Identifying the accident factors is based on applying a comprehensive and integrated system for making decisions. Therefore, applying common mathematical and statistical methods in this field has some problems. While the new research methods with abilities to infer meaning from complicated and ambiguous data seems useful. Therefore, in line with identifying the hot spots, Adaptive Neuro- Fuzzy Inference System (ANFIS) is used to predict hot spots on rural roads. In this process, a fuzzy inference system from Sugeno type is trained by applying hybrid optimization routine (back propagation algorithm in combination with a least square type of method), and accident data of Karaj- Chalus road in Tehran province. Then the system was tested by a complete set of data. Finally the stated system could predict 96.85% of accidents frequency in the studied blocks. Furthermore, the amount of effective false negative in all cases included only 0.82% of predictions, which indicated a good approximation of predictions and model credibility.

Key-Words: - Hot Spots- Accidents- Safety- Road- Neuro Fuzzy- ANFIS- Sugeno- Hybrid

1 Introduction

In this modern world, human being needs more attempts to achieve a proper social, economic, cultural, political, welfare position. Therefore, he tries to use all possibilities to create relation for the purpose of taking advantage of all possible opportunities in each geographical situation. So, roads play significant roles in economic affairs.

More than 80% of transportation is done through road transportation network in Iran, while the casualty statistics resulting from road accidents is very high and has more than 28000 victims per year with their economic and social consequences. The statistics indicates thorough attempt in planning, constructing and safe operating of country roads, while supplying safety isn't possible unless the hot spots are identified within safety inspection plans in existing road networks.

In the present research, hot spots identification methods are briefly mentioned and then the ability of applied devices, processing and their results are explained.

2 Background

Identifying traffic hot spots has been one of the clamorous discussions in transportations and traffic presses within present age. Various methods have been used in identifying the hot spots due to many problems.

May be the simplest way of locating is grading and categorizing them in descending trend according to accidents frequencies and recording accidents, which is used in our country.

Although this method has clear advantages, efficiency of real identifying hot spots has many capacities for development and promotion[1]. In this traditional method we need filed data and historical records for attribution. A main part of researches emphasize on overcoming this shortage and more development and optimization of identification methods. Therefore, many methods based on statistical mathematics, regression- to the mean, and empirical Bayesian methods have been developed [2].

Briefly speaking, a broad range of submitted techniques for Hot Spot Identification (HSID) (including simple categorization of accident rate and frequency, controlling quality rate, identifying the area) have been applied by using safety index/ number and rate methods/ accidents patterns identification methods and various application of Bayesian categorization in techniques with ability to improve identification methods of black spots according to the history of accidents and the ones expected through comparing with similar zones[3].

The performed studies in this respect showed that applying simple methods of categorization with Bayesian methods can remarkably increase the capacity of accidents data set. Therefore collecting comprehensive information about accidents history is required within research limitation and reference population[4].

It can be said about the shortages of this method: 1) They needed a macro reference population, 2) Selecting reference population was completely optional, 3) Reference population always includes required elements for identifying unsafe situations[2].

It was found in studies that there may be two types of errors in identifying hot spots which are false positive and false negative, false positive includes errors which may occur in identifying a point of research span. Here an area which indeed has safe conditions is identified for removing the faults. Since the error is for the purpose of assurance it is called false positive[2].

But false negative is an error because of which an originally hot spot area is identified as a safe one. Since this error causes omission of some areas of research span which are hot spot, it is called false negative[2].

To remove these shortages, lots of attempts have been done which lead to superiority of imperial methods resulting in Bayesian pattern.

On the other hand, studying methods based on new mathematical science such as neuro fuzzy system indicate that such means with their unique abilities to infer meanings from ambiguous and deficient data can be useful in many transportation problems such as identifying hot spots[5].

3 Methodology

According to what has been previously stated and concerning defective and confused records of accidents in Iran we hope that we can find successful results in identifying hot spots by collecting existing data in adaptive neuro- fuzzy inference system and taking advantage of the capabilities of such system.

Fuzzy and Neuro systems are the best approximate reasoning means and optimization in soft computing.

Soft computing means, a series of new calculative methods in computer, artificial intelligence and other applied sciences. Complicated research, modeling and analysis are required for all these fields for which the accurate scientific methods in past weren't successful with their easily, analyzing and full solutions.

- ✓ In soft computing, data research there is no need to presupposition. Data research algorithm in the means of soft computing, search the relation between data in 2 ways of supervision and without supervision (non automatic). Moreover, data research means can use various data and not only the accurately classified number ones[6].
- ✓ Soft computing goes on by accepting inaccuracy and emphasizing on human mind. The conducting principle here is taking advantage of inaccuracy to control the problem and reduce the solution expense[7].
- ✓ Contrary to hard computing methods in which accuracy and full modeling of a fact is concentrated, the soft method emphasizes on minor facts, insufficiency and unassurance. The difference is tough and hard conditions for hard computing to imply and guarantee the reply[3].
- ✓ Professor Zadeh believes: Despite traditional method of hard computing, there is a tolerance in soft computing for any ambiguity, unassurance and relative facts in order to achieve cheaper solutions and better relation between facts[6].
- ✓ Neuro systems besides fuzzy logic are one of the means of soft computing in inferring meanings of complicated and ambiguous data and extracting patterns understanding of which is important for human and other computerized complicated techniques. An adaptive neuro fuzzy system can be considered as an expert in informative category for analysis. This expert can be applied for estimating new desirable situations and replies to "what if …." Questions[7].

Now, regarding the complicatedness and ambiguity of a series of factors involving in accidents, modeling such process by common methods of mathematics (which have been broadly applied during the last 50 years) is very difficult by existing accidents information in our country. Therefore developing neuro fuzzy systems is more acceptable in such situations[8].

Consequently by using MATLAB software and self organized algorithms in ANFIS of fuzzy

inference system (Sugeno type) were used in present research by applying optimization routine back propagating algorithm in combination with a least square type of method.

4 Collecting and processing data

To study the function of adaptive neuro fuzzy system in identifying hot spots, achievable data about occurred accidents for Karaj- Chalus two lane major road was selected which is one of the main roads connecting the capital to the north of Iran.

This road is among beautiful roads throughout the world, and has a high volume of commutation; meanwhile it has instinctive talent for accidents due to its mountainous situation.

However after 70 years since its construction it is not much adapted with modern loading standards.

Accident frequency and intensity was selected for identifying hot spots of this road. In this regard the data related to occurred accidents during 3 months were collected by considering one of the crowded months. That were related to 122 points of this road. Finally 10 items which led to proper inference as well as proper covering form number point of view were selected for identifying process. Table. 1. shows the collected data together with used labels in introducing models. These data include 7 related ones to geometrical situations and place status. They were adopted from topographical and geometrical drawings of the road including the situation of horizontal plan, height, vertical slope, horizontal sight push, road width, bridge and tunnel effect and application of studied area.

Table. 1. The collected data together with used labels in introducing models

Collected data	Used labels
Horizontal Align co.	1
Width diffrence	2
Dever diffrence	3
Slop co.	4
Tunnel & Bridge co.	5
Sight Distance diffrence	6
Residental co.	7
total Vehicle / Day	8
total Vehicle / Hour	9
Day & Hour co.	10
Accident Amplitude	11

The last 3 data also indicate the amount of commutation of vehicles per day in the day of accident occurrence, the amount of commutation per hour in the day of accident, the time of accident as per effect of 24 hours and week days in accident frequency for the said road. Therefore the collection of these data presents a reliable interpretation of road situations in the time of accident occurrence for all recorded accidents and their occurrence blocks.

After collecting data, they were categorized into 2 training and testing groups. The former are latter groups were produced for adapting the model with road situations and for studying the credibility of prediction results from ANFIS in estimating expected frequencies respectively.

As stated before neuro- fuzzy system adopts the ambiguous and invisible concepts and through which the required parameters and components can be adjusted for predicting other data not being confronted with yet. Then the adapted system can predict the frequency of accidents related to the situations of input data introduced in the beginning of this section.

1- First step, the relation between data and accident frequency is found by using the means for identifying proper input data in neuro fuzzy system. According to direct correlation with target function, the data area: representative of horizontal plan, representative of sight distance, bridge and tunnel effect, height difference, zone user, longitudinal slope, representative of width difference, commutation of car equivalent per day, commutation of car equivalent per hour, representative of accident time situation.
2- At next step it was found through studying a broad range of components that the positions (123456) as mentioned above lead to the best result among all components with more than 6 inputs.
3- Then it was specified that data such as the amount of commuting car equivalent per hour (input 9) and time representative (input 10) haven't' proper overall situations due to insufficient frequency and high scatteredness comparing with other inputs.
4- As per above results, three combination of input data which have proper cover on road and contain optimized components from previous stages were studied. The results showed that the combination (1234568) is the best combination for predicting the amount of accident frequency of Karaj- Chalus road. The components of this combination include the horizontal plan, road width, height difference, longitudinal slope, bridge and tunnel effect, sight push, and the amount of car equivalent per day. This combination could produce the accident frequency in 96% with a very high

approximation, of course the other combination could estimate target function but the above combination provides the best result.

5 The results identifying process with best quality

To clarify the quality of predictions, it is necessary to make a statistical review on the amount and type of occurred errors. The summary of results from applying adaptive neuro fuzzy inference system in predicting accident frequency in Karaj- Chalus road is shown in figures 1, 2 and 3. by using the (1234568) data combination. Figure 1 and 2 indicate a very proper approximation of accident frequencies in ANFIS system by putting the expected and predicted amount of accident frequency beside or over each other.

Table 2, shows the data distribution in various error amounts. According to table 2., nearly 83% of data lead to accurate prediction. Moreover 7.8% of predictions have less than 5% error rate and this error in 76% of cases produce less than one unit of error. Therefore, in 90.4% of predictions, the occurred error is very low. 4.1% of cases there is an error of 5 to 15% which has been occurred due to the system that is acceptable.

In 5.6% of remaining cases, the occurred error is more than 15%. This 5.6% (including 33 data items) needs more discussion in view of type and reason of produced error is less than unit and only because of very small frequency they are subject to percent mutation. This group of expansion (including 42%) is ignorable in practice.

If we consider the prediction of second position in view of error type, 15 cases of 33 errors were form the type of false negative (without overlooking predictions with less than unit error and confusion leading to negative frequency). Of course 8 predictions among these cases are about frequency one unit. 3 cases of these 8 predictions and generally 5 cases of 15 errors are included in less than unit error and in all 33 cases is more than 15% equals 14 numbers of total data.

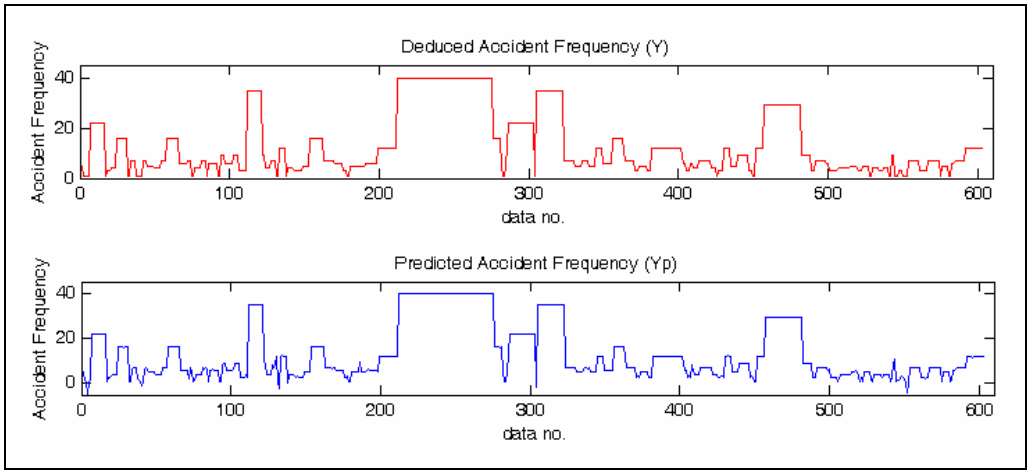

Figure.1. Indicate a very proper approximation of accident frequencies in ANFIS system by putting the expected and predicted amount of accident frequency beside each other.

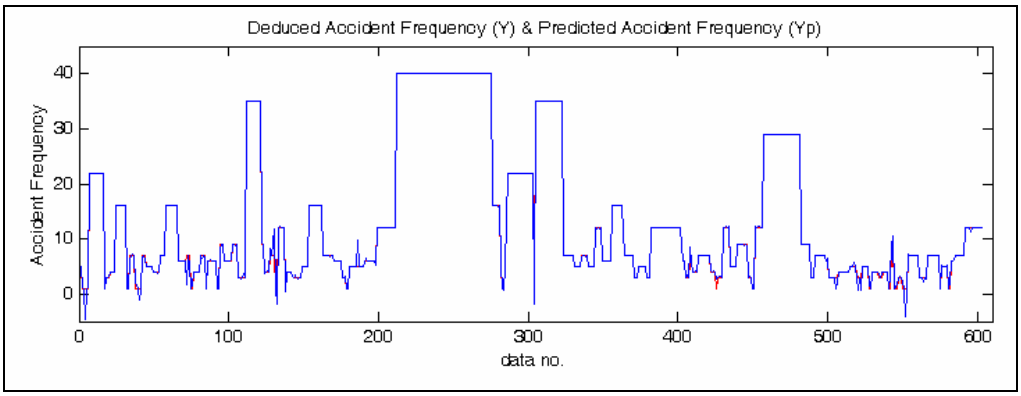

Figure.2. Indicate a very proper approximation of accident frequencies in ANFIS system by putting the expected and predicted amount of accident frequency over each other .

Finally by adding up to the predictions and omitting less than unit error in more than 15% of effective errors constitutes 3.15% of total predictions. It means the adapted system has had acceptable predictions in more than 96.85% of total ones. The summary of table. 2. results is shown in figure. 3.

In other words, root means square error is 6.58% in this model. In fact each of the predictions is as mean with 6.58% error which is very low and acceptable.

6 Conclusion

In conclusion, if we want to discuss the credibility of presented Model, the following results will be explicitly obtained:

a) It should be considered first that in this identification method, it is attempted to present a method for predicting accidents frequency by applying a statistics of accidents in a certain road. So the model was adapted for predicting the series of tests. Therefore, the model could write the training data as the similar accidents frequency. The result showed that the model can estimate the accident frequency in more than 96% of cases with good quality and proper coverage over peripheral conditions. In fact, the ability of the model to produce real data with average error of 6.58% evidences the assurance of processing results, while considering average accident frequency (equals 13.4), each data items was identified and -0.88% error was averagely found.

b) On the other hand, the credibility of results are studied according to predictable usages for the model. It should be taken into consideration that such identifications may be used for future trends of accidents in the roads without written record. So the results shall be free of any false negative as much as possible. However as previously stated, there was only 5 false negative with more than 1 or 1 error unit. This amount of false negative equals 0.82% of predicted cases. Such a result also shows the credibility of applied method.

c) As expected before, in identifying hot spots regarding Karaj- Chalus road, a reliable success was obtained in view of producing amounts similar to reality in accident frequency. These results confirm the application of such systems in analyzing complicated traffic problems.

7 Efficiency expansion

The first interpretation of the above 3 results is that according to abilities of such models we can expect that such systems shall be used in various inspections of under loading or ready to load roads. It should be added that the system can be trained by extracted data from the road with completer accident history and used for the roads with similar conditions. This matter can be very useful in studying the behavior of roads which haven't been opened yet and predicted to operate under certain loading.

Other application of such processing is the ability of this system to study the amount of factors effective in accident occurrence. Therefore by using the results of neuro fuzzy inference system, the hidden relation among the data can be found. Such result can be applied for giving priority in cases where a limited budget is available and one situation with several hazardous reasons is under investigation.

Table.2. The data distribution in various error amounts

Error amounts percental	Number of data in error amounts	Percentage of data in error amounts
100%<	10	1/7%
[70% 100%)	5	0/8%
[60% 70%)	3	0/5%
[40% 60%)	2	0/3%
[30% 40%)	4	0/7%
[20% 30%)	4	0/7%
[15% 20%)	5	0/8%
[10% 15%)	5	0/8%
[5% 10%)	20	3/3%
[2% 5%)	34	5/6%
[1% 2%)	13	2/2%
<1%	498	82/6%

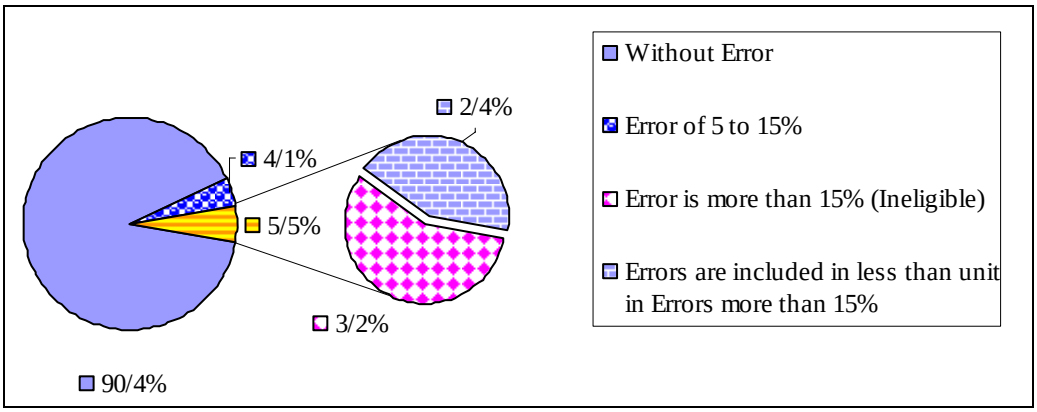

Figure.3. The summary of results from applying adaptive neuro fuzzy inference system in predicting accident frequency in Karaj- Chalus road.

References:

[1] Kononov, J. and B. N. Jason. Diagnostic Methodology for the Detection of Safety Problems at Intersections. In Transportation Research Record 1784, TRB, National Research Council, Washington, D. C. , 2002.

[2] Simon Washington and Wen Cheng, HIGH RISK CRASH ANALYSIS-Final Report 558. Department of Civil Engineering .University of Arizona.Tucson, AZ 85721- December 2005

[3] Hauer, E. , D. W. Harwood, F. M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. In Transportation Research Record 1784, TRB, National Research Council, Washington, D. C. , 2002.

[4] Davis, G. A. and S. Yang. Bayesian Identification of High-Risk Intersections for Older Drivers via Gibbs Sampling. In Transportation Research Record 1746, TRB, National Research Council, Washington D. C. , 2001.

[5] DusÏ an TeodorovicÂ, Fuzzy logic systems for transportation engineering, Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Yugoslavia- Transportation Research Part A 33 (1999)

[6] Mahdi Ghazanfary, Fundamental of Expert systems, Neuro-Fuzzy network and Theory of Fuzzy sets, Iran University of Since & Technology , 2004

[7] Wang, Lie-Xin, A Course in Fuzzy Systems and Control, translate by Mohamad Teshne Lab, Nima Saffar Pour and Dariush Afuoni, K.N.Toosi University of Technology, 1999

[8] Fuzzy logic systems for transportation engineering: the state of the art by DusÏ an TeodorovicÂ Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Yugoslavia- Transportation Research Part A 33 (1999) –pp. 337-364