ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Data Mining Based on Analyzed Road Accident Data

Uppu Nirosha
Assistant Professor,
CSE Branch.

Abstract:

Road accident statistics are collected and used by a large number of users and this can result in a huge volume of data which requires to be explored in order to ascertain the hidden knowledge. Potential knowledge may be hidden because of the accumulation of data, which limits the exploration task for the road safety expert and, hence, reduces the utilization of the database. Road traffic accident databases provide the basis for road traffic accident analysis, the data inside which usually has a radial, multidimensional, and multilayered structure. Traditional data mining algorithms such as association rules, when applied alone, often yield uncertain and unreliable results. An improved association rule algorithm based on Particle Swarm Optimization (PSO) put forward by this paper can be used to analyze the correlation between accident attributes and causes. The new algorithm focuses on characteristics of the hyper stereo structure of road traffic accident data, and the association rules of accident causes can be calculated more accurately and in higher rates. The result of which was a ten times faster speed for random traffic accident data sampling analyses on average. In the paper, the algorithms were tested on a sample database of more than twenty thousand items, each with 56 accident attributes. And the final result proves that the improved algorithm was accurate and stable.

Keywords:

PSO, Data preprocessing

INTRODUCTION:

In recent years, with the growth of the volume and travel speed of road traffic, the number of traffic accidents, especially severe crashes, has been increasing rapidly on a yearly basis. The issue of traffic safety has raised great concerns across the globe, and it has become one of the key issues challenging the sustainable development of modern traffic and transportation. Therefore, it is crucial for engineers to be able to extract useful information from existing data to analyze the causes of traffic accidents, so that traffic administrations can be more accurately informed and better policies can be introduced.Traffic conditions are a complex system due to many stochastic factors and traffic accident data has long been known to be very difficult to process. Many attempts have been made in recent years through applying different methodologies and algorithms. Association rules has captured wide attentions and careful studies because of its adoptability and the nature of being easily understood, the focus of study of which is how to increase the accuracy and efficiency of the calculation. Among the researches to date, association rules to identify accident circumstances that frequently occur together at high frequency accident locations; an adaptive regression trees to build a decision support system to handle road traffic accident analysis; and other researchers have achieved multileveled data mining of traffic accidents by means of a comprehensive application of data mining techniques. The researches above all achieved the mining of accident data on a certain level; however, the overall calculating processes are largely too complicated and cannot be applied to all types of data, especially the multi attribute ones. On the other hand, the PSO algorithm has been applied in many fields. The parameters' optimization, based on particle swarm optimizer. An association rules algorithm based on particle swarm optimization algorithm to mining the transaction data in the stock market. Moreover, others improved and applied PSO algorithm to their purpose. So far, there have been a lot of researches targeting at different types of data, and due to the "capricious" nature of real-world data, coupled with the innate shortcomings of the algorithm, the association rules still falls short of people's expectations in being less complicated, less time and space-consuming, and more efficient.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

In this paper, a new method of traffic accident data mining, based on PSO, association rules, and Information Entropy theories and through a comparative analysis of a variety of traffic accident data mining techniques, is put forward to identify the importance of different attributes and their respective values. The method is an attempt to come up with a multidimensional, all-inclusive method of data analysis to simplify existing algorithms as well as apply computational intelligence algorithms such as PSO to road traffic data analyses.

Characteristics of Road Traffic Accident Data:

Road traffic accident is under the influence of many factors, which makes it a complicated, and as far as information is concerned, an unfinished, uncertain gray system. There are different databases of traffic accident in different countries. At present, roughly 60 items of information are contained in the China "Database of Road Traffic Accident" which is used by Chinese traffic administrative agencies, spawning off approximately 130 items of single-unit information, which can be used to reconstruct the whole process of the accident in a relatively full and objective manner. It provides more than adequate the information and references for road traffic accident analyses. Road traffic accidents have their innate, random nature but are also subject to other factors. If the connections of those factors could be identified, through manual control and interference, the rates of traffic accidents could be lower. Traffic accident data is the foundation of traffic accident analysis, the form and structure of which determine the form and structure of the analysis model. From in-depth analysis of the traffic accident database operated by the Ministry of Public Security, the data could be regards as a radial, multidimensional, and multilayered structure.

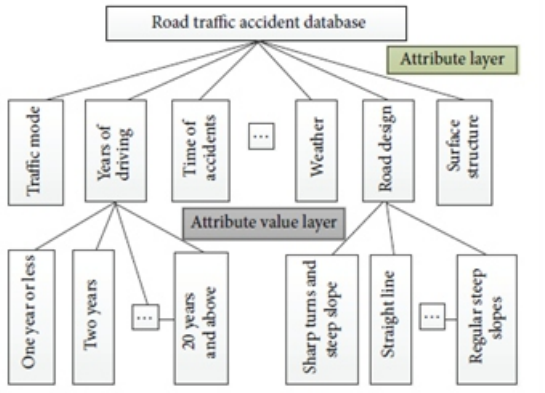

Fig:-Structure of the traffic accidents data.

The structure of the data determines the structure of the causation-analysis model. This paper designs a double-layered analysis model and provides an improved algorithm according to the hyper stereo structure of the data. The purpose is to analyze the importance of each value on the attribute value layer with the association rules method and to compare the importance of each attribute on the attribute value layer with the Information Entropy method. Characteristics of Data Analysis Algorithms.

Association Rules:

Association rules is a data mining method for investigating the associative property of different events, which can be used in traffic accident data mining to mine the importance of attributes, that is, the associative relationship of events with certain types of accident. Its basic idea is to treat each characteristic as an item. Accident site, number of death, and so on can all be called an item.

The higher the association, the more likely one event is directly linked to the cause of a certain type of accident. To decide how related two items are, we need to identify how many times some characteristics appear at the same time in a large number of similar events. If items show up at the same time frequently, indicating that there is a statistic pattern behind it, we can start to believe that the items are relevant.

PSO Algorithm:

PSO was intended for simulating social behavior, which has many advantages such as higher convergence rates and being more applicable than other algorithms. However, it generally has a lower accuracy than genetic algorithms and under certain initial conditions it can only reach optima in a subset of the problem.PSO is a population-based search algorithm and is initialized with a population of random solutions, called particles.

Unlike in the other evolutionary computation techniques, each particle in PSO is also associated with a velocity. Particles fly through the search space with velocities which are dynamically adjusted according to their historical behaviors. Therefore, the particles have the tendency to fly towards the better search area over the course of search process. The PSO was first designed to simulate birds seeking food which is defined as a "cornfield vector."

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Definition of Information Entropy:

Traffic accidents are events with strong randomness, while entropy is the mathematical method for analyzing an event's uncertainty. From the perspective of statistical mathematics, entropy is a measure of system randomness. Information Entropy is a value for characterizing the statistical characteristics of random variables. It is a measure of the average uncertainty of random variables, an objective description of statistical characteristics of the population.In the traffic accident data gathering process, due to the influence and limitations from many factors, the number of traffic accident data items is usually insufficient, which cannot be used to analyze the statistical characteristics. The use of Information Entropy as a statistical measure of the uncertainty information does not require the probability distribution of the data to be known and does not require the distribution to be single humped; that is, no prior information is needed. So Information Entropy is very suitable for testing the degree of discreteness of the population. As far as Information Entropy is concerned, the decrease of uncertainty means the reduction of entropy value.

The Method of Traffic Accident Data Preprocessing:

Due to human factors, in the actual accident data gathering process, some data items may be unavailable, affecting the integrity of the data and causing the results of the causation analysis to be considered not convincible. Data preprocessing is an essential step in any of the data mining processes. Researches on data preprocessing techniques mainly focus on the preprocessing of data which follows obvious patterns, the widely used method of which is to find the patterns, characteristics, and properties so that data can be preprocessed in a certain way. In contrast, it is rarely seen that data in non digital form is processed in the same manner. The process includes the following parts: data washing, data filling, data integration, and data transformation. This paper adopts the data-washing methodology according to the characteristics of traffic accidents to build a traffic accident database using a form which resembles how "antivirus" software works, which has an "antivirus" definition library of its own, that is, an error database that is artificially created, and to predefine a set of rules for the "software" to use in order to hunt down all the "viruses." However, before the data-washing can be initiated, a database of all the errors needs to be created.

The error database comprises a gathering of area-specific knowledge and common sense, through which data that contains errors are marked through comparison. Of course, the data washing only could correct and delete the error data which is detected in logic, not all error types. Due to the fact that data format and structure in data mining are not entirely identical with those in the database, data needs to be transformed before it can be mined, so that existing data can be changed into proper format or form to be mined through data mining techniques. The current data inside the traffic accident database happens to be coded; only part of the statistical attributes is in coded format which is required by the association rules; therefore, the data in the database should be kept as close to their original form as possible. Attributes whose attribute values are not coded but sequential ought to be scattered to a certain extend and coded in orders.

Fundamental Principle of the Algorithm:

Focusing on the characteristics of road traffic accident data, causation analyses need to examine the double-leveled structure. Unfortunately, although they can analyze the causes of accident from different angels and each method has its advantages, none of the data mining methods currently being widely used can accomplish an overall, multi angled, multilayered data mining task on its own. With the accumulation of traffic accidents database, the data quantity is more and more huge, so how to obtain the effective knowledge, hiding rules, and fundamental causes is changing into one key issue for road traffic administration. To meet the demand of better accuracy and efficient analysis of traffic accident causes, this paper combines the binary PSO algorithm to improve the association rules. The reason is that the speed of the PSO algorithm does not decrease with the increase of the number of datasets. To solve the problem that accident data needs to be analyzed in different layers, this paper introduces the Information Entropy theory into road traffic accident analysis, with the help of the Association Rules theory, and puts forward the concept of Association Entropy and its algorithm. With the introduction of PSO and Association Entropy, traffic accident causes can be analyzed from all angles and on all layers, satisfying the requirement that the association rules have to be within a certain support level. In the meanwhile, causes on different levels can provide references for different traffic administrations at different levels, so that more effective preventative measures can be taken.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

CONCLUSION:

This paper aims at the hierarchically structured characteristics of road traffic accident databases, mixed using the method of associated rules, PSO, and Information Entropy to analyze the degree of importance of traffic accidents. Through a modification of traditional methodologies and algorithms, a PSO algorithm and associated entropy model are built for calculating the degree of importance of road accidents. Through applying the improved algorithm on both the attribute and the attribute value layers, respectively, each accident-triggering factor's influence on the severity of accident is calculated. The algorithm this paper introduced has the advantage of better accuracy and higher mining rates over the traditional association rules and PSO algorithms, the result of which is quite different from what the experts concluded, which indicates when facing a large amount of random information, people's experiences and how people perceive things are limited. Of course, traffic accident data as a type of data possesses certain physical meanings, whether there really exist connections between certain types of data, and that certain types of data were manually gathered so they may not be error-free, as well as whether they can be fully applied to the models and algorithms of this paper in their entirety questions as those that are yet to be discussed in future researches. However, approved by real applications and tests of effectiveness, this type of data mining method which is based on traffic accident database provides yet another powerful tool to quantify data in traffic accident analysis, which is going to be helpful to accident experts and traffic administrative agencies to clarify how much of role different factors play in investigations of traffic severity.

REFERENCES:

1. Savolainen P, Mannering F, Lord D, Quddus M. The factual investigation of parkway accident harm severities: a survey and evaluation of methodological options. Accid Anal Prev. 2011;43:1666–76.

2. Depaire B, Wets G and Vanhoof K. Auto collision division by method for idle class bunching, mishap investigation and aversion, vol. 40. Elsevier; 2008.

3. Karlaftis M, Tarko A. Heterogeneity contemplations in mischance displaying. Accid Anal Prev. 1998;30(4):425–33.

4. Mama J, Kockelman K. Crash recurrence and seriousness displaying utilizing bunched information from Washington state. In: IEEE Intelligent Transportation Systems Conference. Toronto Canadá; 2006.

5. Jones B, Janssen L, Mannering F. Investigation of the recurrence and term of turnpike mischances in Seattle, mishap examination and aversion, vol. 23. Elsevier; 1991.

6. Miaou SP, Lum H. Displaying vehicle mischances and expressway geometric outline connections, mishap investigation and aversion, vol. 25. Elsevier; 1993.

7. Miaou SP. The relationship between truck mishaps and geometric outline of street sections–poisson versus negative binomial relapses, mischance examination and anticipation, vol. 26. Elsevier; 1994.

8. Poch M, Mannering F. Negative binomial examination of crossing point mischance frequencies. J Transp Eng. 1996;122. 9. Abdel-Aty MA, Radwan AE. Demonstrating car crash event and association. Accid Anal Prev Elsevier. 2000;32.

10. Joshua SC, Garber NJ. Assessing truck mishap rate and associations utilizing straight and poisson relapse models. Transp Plan Technol. 1990;15.

11. Maher MJ, Summersgill I. A complete strategy for the fitting of prescient mishap models. Accid Anal Prev Elsevier. 1996;28.

12. Chen W, Jovanis P. Technique for recognizing components adding to driver-harm seriousness in car accidents. Transp Res Rec. 2002:1717.

13. Chang LY, Chen WC. Information mining of tree based models to break down interstate mischance recurrence. J Saf Res Elsevier. 2005;36.

14. Tan PN, Steinbach M, Kumar V. Prologue to information mining. Pearson Addison-Wesley; 2006.

15. Abellan J, Lopez G, Ona J. Analyis of car crash seriousness utilizing choice tenets by means of choice trees, vol. vol. 40. Master System with Applications: Elsevier; 2013.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

16. Rovsek V, Batista M, Bogunovic B. Recognizing the key danger elements of auto collision damage seriousness on Slovenian streets utilizing a non-parametric arrangement tree, transport. UK: Taylor and Francis; 2014.

17. Kashani T, Mohaymany AS, Rajbari A. An information mining way to deal with recognize key elements of activity damage seriousness, promettraffic and transportation, vol. 23; 2011.

18. Han J, Kamber M. Information Mining: Concepts and Techniques. USA: Morgan Kaufmann Publishers; 2001.

19. Fraley C, Raftery AE. Model-based grouping, discriminant investigation, and thickness estimation. J Am Stat Assoc. 2002;97(458):611–31.

20. Sohn SY. Quality capacity sending connected to localtraffic mischance decrease. Accid Anal Prev. 1999;31:751–61.

21. Ng KS, Hung WT, Wong WG. A calculation for surveying the danger of auto collision. J Saf Res. 2002;33:387–410.

22. Pardillo-Mayora JM, Domínguez-Lira CA, Jurado-Pina R. Experimental alignment of a roadside peril record for Spanish two-path country streets. Accid Anal Prev. 2010;42:2018–23.

23. Vermunt JK, Magidson J. Dormant class bunch examination. In: Hagenaars JA, McCutcheon AL, editors. Progresses in dormant class investigation. Cambridge: Cambridge University Press; 2002.

24. Oña JD, López G, Mujalli R, Calvo FJ. Examination of auto collisions on rustic parkways utilizing Latent Class Clustering and Bayesian Networks, mishap investigation and aversion, vol. 51; 2013.

25. Kaplan S, Prato CG. Cyclist-driver crash designs in denmark: a dormant class bunching approach. Activity Inj Prev. 2013;14(7):725–33.

26. Chaturvedi A, Green P, Carroll J. K-modes bunching. J Classif. 2001;18:35–55.

27. Goodman LA. Exploratory inert structure investigation utilizing both identifiable and unidentifiable models. Biometrica. 1974;62.

28. Agrawal R, Srikant R. Quick calculations for mining affiliation rules in expansive databases. In: Proceedings of the twentieth International Conference on substantial information bases; 1994. pp. 487–99.

29. Akaike H. Element investigation and AIC. Psychome. 1987;52:317–32.

30. Raftery AE. A note on Bayes elements for log-direct possibility table models with obscure earlier data. J Roy Stat Soc B. 1986;48:249–50.

31. Fraley C, Raftery AE. What number of groups? Which bunching strategy? Answers by means of model-based bunch investigation. Comput J. 1998;41:578–88.

32. Wong SC, Leung BSY, Loo BPY, Hung WT, Lo HK. A subjective appraisal system for street security approach methodologies. Accid Anal Prev. 2004;36:281–93