

Investigation of Traffic Factors in Accident Zones in Iranian Railways (RAI) by Applying Data Mining Techniques

Masoud Yaghini¹, Shabnam Sharifian²

^{1,2}Iran University of Science and Technology (IUST), School of Railway Engineering (SRE)

²Khatam Institute of Higher Education

²Metra Consultant Engineers Co.

¹yaghini@iust.ac.ir, ²shabnam.sharifian@gmail.com

Abstract

Demand to travel by rail is ever increasing due to its benefits for both passenger and freight; therefore it is of utmost importance for railway administrators to carry passenger and freight safely in all aspects to their destinations. Undergoing safety procedures and developing safety systems requires awareness of what is causing unsafe conditions, which can be done by learning from the past. This research has been performed to analyze past accidents data of the Iranian Railway (RAI) by applying data mining, one of the most revolutionary developments of the next decade according to the online technology magazine ZDNET News. Data mining is the process of selection, exploration, and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database. Data mining has been an active analytical technique in many scientific areas for years. One of these areas is Transportation and many researchers have figured out the useful role data mining plays in dealing with a mass of transportation data, and the advantages of applying data mining to retrieve or analyze the data. By the application of Cross-Industry Standard Process for Data Mining (CRISP-DM) Reference Model as the data mining methodology and utilizing Clementine 12.0 as the software tool, the mentioned objectives of this paper were fulfilled. For this research, some 6500 accident records were selected from the RAI accidents database from years 1996 to 2005. In this paper railway stations and blocks have been clustered separately based on two criteria: 1) total accident injuries and loss of life per traffic amount and 2) total damage to properties per traffic amount in order to find accident zones. By comparing the results of K-Means and TwoStep clustering algorithms, stations are ultimately clustered into three and blocks into four clusters. Traffic related factors are studied in these zones as follows: stations clusters where further investigated by considering type of freight carried, type of rolling stock attached or detached to or from the train, and the rolling stock's row in the train, and block clusters have been analyzed based on train type, train length, train weight and breached weight to determine their relationship with the accident type by applying decision tree technique. Decision tree models have been built with CHAID and C5.0 algorithms and results have been compared to choose the most accurate tree. The ultimate relations discovered can be utilized to develop regulations and rules based on accident conditions and relations discovered in order to prevent their repetition in future.

Key Words: Data Mining, Clustering, Decision Tree, CRISP Reference Model, Iranian Railways (RAI), Railway Accidents, Accident Analysis