تحلیل داده های سوانح راه آهن ج.ا.ایران با استفاده از تکنیک داده کاوی Association Rules

احمد میرآبادی[۱]، شبنم شریفیان[۲]

[۱] دکترای مکاترونیک، مؤسسه تحقیقات و توسعه فناوری ریلی ایران، mirabadi@iust.ac.ir

[۲] کارشناس مهندسی راه آهن، کارشناس ارشد مهندسی صنایع، مؤسسه تحقیقات و توسعه فناوری ریلی ایران، shabnam.sharifian@gmail.com

چکیده

از بین سیستمهای حمل و نقلی راه آهن اقتصادیترین و ایمنترین ابزار جا به جایی بار و مسافر محسوب می شود که ترافیک را کاهش داده و حداقل آلودگی را برای محیط زیست به دنبال دارد. یکی از کلیدیترین راههای حفظ و ارتقاء جذابیت جا به جایی با استفاده از راه آهن برای مسافران و صاحبان بار، ارتقاء سطح ایمنی است. طراحی و پیاده سازی سیستمهای ایمنی نیازمند آن است که از شرایط خاصی که در سطح شبکه راه آهن ما ناامنی ایجاد می کند آگاهی کافی وجود داشته باشد. بسیار اتفاق می افتد که سوانح به دلیل عدم توجه به شرایط مشابه در گذشته به وقوع می پیوندد. لازم است مسئولان ایمنی این صنعت با بهره جستن از تجارب حاصل از سوانح گذشته، زمینه تکرار آنرا در آینده از بین ببرند. استفاده از تکنیکها وابزارهای جدید و به روز، می تواند دیدی متفاوت با آنچه تا به حال از تکنیکهای آمار توصیفی توسط متصدیان ایمنی راه آهن ایران ارائه گردیده است، جهت ارتقاء سطح ایمنی بوجود آورد. مقاله حاضر حاصل تحقیقاتی است که بر روی داده های سوانح ریلی راه آهن ج.ا.ایران بدین منظور صورت پذیرفته است. در این تحقیق با استفاده از تکنیک association rules، که یکی از تکنیکهای کارآمد داده کاوی[1] محسوب است، به تحلیل داده های سسالهای ۷۵ تا ۸۵ سوانح راه آهن ایران پرداخته شده است تا روندها، ارتباط بین فاکتورهای سوانح و الگوهای تکرار شونده ای که در نگاه اول و با استفاده از تکنیکهای آماری پنهان باقی می مانند، استخراج گردند. همچنین لازم به ذکر است جهت انجام تحقیقات از متدولوژی CRISP-DM و نرم افزار Clementine استفاده شده است. در انتها با تکیه بر دانش و تجربیات مستند شده حاصل از تحلیل داده ها، دستورالعملهایی ارائه گردیده است که با پیاده سازی آنها از تکرار الگوها، روندها، و ارتباطات شناسایی شده موجود بین فاکتورهای سوانح در آینده، پیشگیری شود.

کلمات کلیدی: تحلیل سوانح، راه آهن ایران، داده کاوی، association rules

[1] Data Mining

Application of Association Rules in Iranian Railways (RAI) Accidents Data Analysis of the

Ahman Mirabadi[1], Shabnam Sharifian[2]
[1] Ph.D Mechatronics, Iran Institute of Railway Research and Development (IRRD), mirabadi@iust.ac.ir
[2] MSc. Industrial Engineering, BSc. Railway Engineering, Iran Institute of Railway Research and Development (IRRD), shabnam.sharifian@gmail.com

Abstract
Railway is meant to be an excellent means of transport to reduce environment pollution and avoid traffic congestion, a safe and economic way to reach destination for both passenger and freight. For this reason it is of utmost importance for railway administrators to carry passenger and freight safe and economic to their destinations. Undergoing safety procedures and developing safety systems requires awareness of what is causing unsafe conditions. Many accidents happen because lessons that could have been learned from the past have been neglected. Literature review on accidents analysis reveals that little attention has been given to this matter in the railway industry. This project has been defined to analyze past accidents data of the Iranian Railway (RAI) by applying association rules data mining technique, in order to discover relations and regularities among the data that were unknown at first. By the application of CRISP-DM and Clementine as the data mining methodology and tool, the objective of the project was fulfilled. The ultimate outcomes of the research are regulations and rules witch have been suggested according to conditions of accidents and relations discovered among the most common accidents factors (human error, wagon and track) with other fields of the database in order to prevent their repetition in the future. To fulfill this research the accidents database from the years 1996 to 2005 with some 6500 records were used.

Key Words: Data Mining, Association Rules, CRISP Reference Model, Iranian Railways (RAI), Railway Accidents, Accident Analysis

Introduction

Today, we are surrounded by an ocean of all kinds and ever increasing amount of experimental data (i.e., examples, samples, measurements, records, patterns, pictures, tunes, observations, etc.) produced by various sensors, cameras, microphones, pieces of software and/or other human made devices. The first obvious consequence of such a fact is that humans can't handle such massive quantity of data which are usually appearing in the numeric shape as the huge (rectangular or square) matrices [1].

To understand the term 'data mining' it is useful to look at the literal translation of the word: to mine in English means to extract. The verb usually refers to mining operations that extract from the Earth her hidden, precious resources. The association of this word with data suggests an in-depth search to find additional information which previously went unnoticed in the mass of data available [2].

This terminology was first formally put forward by Usama Fayaad[2]. He used to refer to a set of integrated analytical techniques divided into several phases with the aim of extrapolating previously unknown knowledge from massive sets of observed data that didn't appear to have any obvious regularity or important relationships. As the term 'data mining' slowly established itself, it became a synonym for the whole process of extrapolating knowledge [2]. Here is a more complete definition of data mining:

> *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* [2].

According to the online technology magazine ZDNET News, data mining is predicted to be "one of the most revolutionary developments of the next decade". In fact, the MIT Technology Review chose data mining as one of 10 emerging technologies that will change the world. "Data mining expertise is the most sought after" among information technology professionals, according to the 1999 Information Week National Salary Survey. Since many companies have implemented a data warehouse strategy, they are now starting to look at what they can do with all that data [3].

Railway is nowadays an excellent means of transport to reduce environment pollution and avoid traffic congestion, a safe and economic way to reach destination for both passenger and freight. As a result, passengers and freight owners prefer railway to other transportation moods. Therefore; to retain this conception and achieve competitive advantage, railway transportation administrators should work full justice to raise the

[2] The "First International Conference on Knowledge Discovery and Data Mining", Montreal, 1995.

level of safety and reduce accident causing factors.

Developing safety systems and undergoing safety procedures requires complete awareness of the nature of previous accidents on the railway network and understanding where the real problems are laying. On the other hand many accidents occur due to ignorance of learning from past accidents or failures. For this reason it is of utmost importance to analyze accident data in order to extract hidden knowledge among huge amounts of data. Effective analysis of data from a database can help in development of knowledge that can support safety management strategies and reducing future accidents.

The objective of this article is to discover meaningful new correlations, patterns and trends among rail accidents data of the Iranian Railways (RAI). Once unnoticed and hidden relations are clarified, the outputs are further analyzed in order to present suitable solutions according to needs of RAI.

In this article, we considered safety as *"being protected against loss of life and property in all aspects and dimensions related to the railway industry for both railway and non-railway individuals"*. Therefore; the objective of this paper is to analyze data on accident factors which have resulted in *loss of life and property* by applying Association Rules, a data mining technique.

Literature Review

Data mining has been an active analytical technique in many scientific areas over years [4]. 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 [5].

Literature review reveals that recent studies on the adoption of data mining in Transportation Engineering, covers the following areas:

Traffic environment evaluation [6], traffic flow [7; 8; 9; 10; 11], pavement management [12; 13], accident analysis [14; 15; 16; 4; 17; 18; 19], roadway condition [20], demand prediction [21; 22; 23; 24], travel time estimation [25, 26], logistics [27], GIS [28; 29; 30; 31; 32; 33], project management and costs [34; 35; 36], routing problem [37; 38; 39], traffic signal optimization [40; 41], etc.

Very little is known about the usefulness of applying data mining in railway traffic and transport related areas, although there are numerous applications of data mining in other transportation fields, witch the most popular is road transportation. In railway transportation the following topics have been covered in the literature review:

Damage Detection [42], Safety Prediction [43], Mechanical Behavior of Ballast [44], Railroad Operation Monitoring and Control [45; 46; 47, 48; 49], Relating Track Geometry to Vehicle Performance [50], Speed Estimation [51; 52], Vehicle Identification [53], Maintenance [54; 55], Routing [56], Accidents [57; 58; 59], Safety at Level Crossings [59; 60; 61].

Demand for traveling by rail for both passenger and freight has increased significantly over the past years and is predicted to increase steeper than before. Railway administrators are planning to respond to the future growing demands with a wide range of software and hardware procedures. No mater what these procedures are, they all lead to increase in mobility and accidents are a consequence of this increase.

Application of data mining in railway accidents analysis in the literature review is concentrated on accidents at level crossings, while other aspects of rail accidents have remained unconsidered and some how left out of the researchers' attention.

Methodology and Tool

The methodology of this article is based on the CRISP-DM reference model. This model (Figure 1) consists of six phases, their respective tasks, and the relationships between these tasks. Moving back and forth between different phases is always required. The outcome of each phase determines which phase, or particular task of a phase, has to be performed next [62].

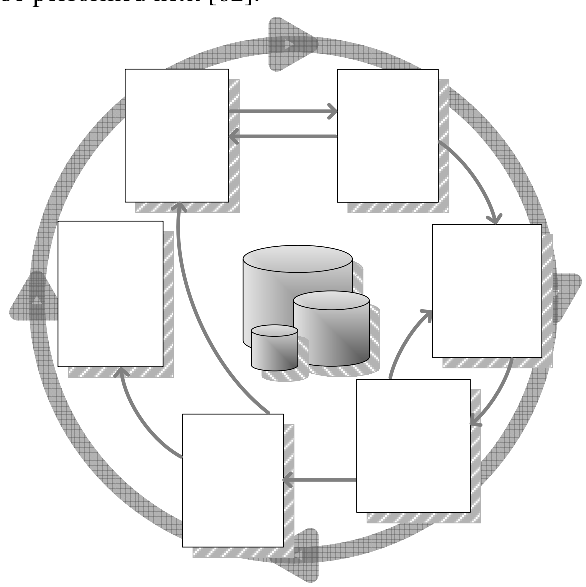

Figure 1: Phases of the CRISP-DM Reference Model [62]

A convenient way to adopt data mining analysis is to use a software program that hosts facilities to mine the data in a variety of ways. Clementine 12.0, a product of Integral Sol., Ltd., due to its visual interface, algorithm breadth and following CRISP model made it suitable for the purpose of mining data in this research.

Results

In this section, results of the six phase of the research methodology are reviewed.

Business Understanding: Safety and accident analysis in RAI

The Iranian Railways (RAI) is the national state-owned railway system of Iran. The Iranian Ministry of Roads and Transportation (R&T) is the state agency that oversees the RAI.
In 2008, Iran with an area of 1,648,195 Sq km and nearly 70 million populations, RAI operated 11,106 km of rail network over 14 districts [63]. The railway network expands by about 500 km per year according to the Ministry of R&T.

In 2006, RAI reported its facilities as follows [63]:

1. Locomotives (diesel-electric, electric and shunting), numbering 565
2. Wagons (covered, low sided, high sided, flat, well-wagon, tank wagon, mineral materials carrying wagon, bulk carrying, ballast, gas, refrigerator, etc), numbering 16,330
3. Different kind of passenger coaches, numbering 1,192
4. Main Stations, numbering 429
5. Bogie change installation at Jofa and Sarakhs international border stations witch changes about 200 bogies each 24 hours based on two working shifts.
6. Free Trade International stations, numbering 16

The Gen. Dept. of Movement Safety is held responsible for safety issues in RAI. One of the functions performed by this Dept is gathering and analysis of accidents data. The analyses fulfilled on the data stored in the database are simple descriptive statistics and comparison of the statistics with previous periods. For this purpose an annual bulletin has been published every year since 2001, to reports the number of accidents categorized by accident type, grade, cause, etc in railway districts.

As previously mentioned the objective of this project is to discover correlations and trends that lead to loss of life and property related to railway and non railway individual and ultimately develop solutions to break the identified accident patterns toward safety.

Data Understanding: Gaining Insight on Accidents Data

The data stored in RAI Accidents Database (DB) was used to fulfill this research. This DB was in two different parts: 1) 1996-2005 and 2) 2006 and later. These two parts were developed with different software developing languages and contained different types of information fields; even similar information fields were stored in different formats or left unfilled due to changes in users of the system and lack of personnel training in using the new system. Therefore, because of the larger amount of data in the 1996-2005 DB, this part with some 6500 records was chosen for mining and analysis. Studying fields of each table, 38 out of 63 information fields were found proper for data mining. Appendix 1 (Table 5) holds titles and definitions of these fields.

Others fields were ether descriptive, not filed regularly through out these years, or anyhow irrelevant for data analysis. There was also some information fields witch were extracted during the modeling phase from those mentioned in table 2, according to requirements of the models developed.

Data Preparation: Getting the data fit for mining

It is quite common users of the system enter data with errors or unusual values, even data might be stored in an inconsistent format with analysis objectives, and therefore data must be prepared before mining. In other words, the data wished to be analyzed in the real world by data mining techniques are incomplete (lacking attribute values or certain attributes of interest, or containing only aggregate data), noisy (containing errors, or outlier values that deviate from the expected), and inconsistent (e.g., containing discrepancies in the department codes used to categorize items) [64].

In this research missing values, noises and inconsistency data were dealt with as follows:

- Missing data: replaces by their mod (for set data type), mean value (for range data types), and in some cases they were remover or ignored (depending on the field or amount of missing data).
- Noisy data: omitted due to the small number of such records or lack of information to clean others.
- Inconsistent data: numerical and alphabetical peace of information became coded in suitable numerical sets, ranges or flag type.

Modeling: Association Rules

Association rules were developed to underline groups of items that typically occur together in a set of transactions. Association rules are probably the most well-known local method for detecting relationships between variables. They can be used to mine very large data sets, for which a global analysis may be too complex and unstable [2] witch take the form "If antecedent, then consequent," along with a measure of the support and confidence associated with the rule [3].

Among the association rule algorithms supported by Clementine 12.0; witch are Generalized Rule Induction (GRI), Apriori, and CARMA; GRI is utilized for association rules extraction. For this purpose, the share of accident factors during the years 1996 to 2005 were examined for more effective rule extractions. According to figure 2 among all factors, human error, wagon and track are the most accident producing ones. Therefore in this paper, the relationship between these factors and other fields of the database are examined.

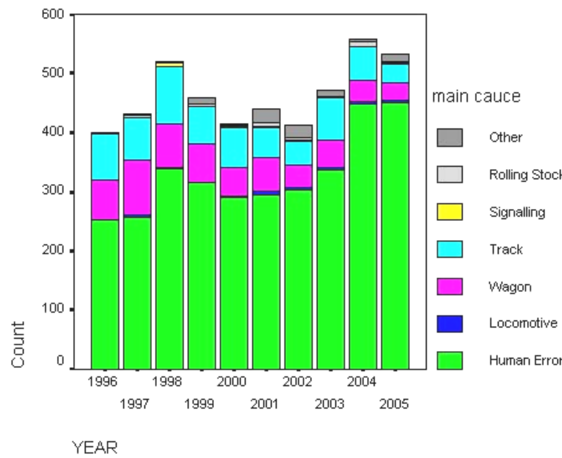

Figure 2: Share of Accident Factors from 1996 to 2005

Table 1: Relations Examined within the Accident Database

Factor	Examined Relation
Human Error	– Job - Accident Type
	– Job - Site (Station/ Block)
	– Job - Years of Service - Districts - Accident Type
	– Job - Years of Service - Districts - Accident Grade
	– Job - Years of Service - Day of Week - Accident Type
	– Job - Years of Service - Day of Week - Accident Grade
	– Employment Condition - Level of Education - Districts - Accident Type
	– Employment Condition - Level of Education - Districts - Accident Grade
	– Job - Age - Districts - Accident Type
	– Job - Age - Districts - Accident Grade
	– Job - Age - Day of Week - Accident Type
	– Job - Age - Day of Week - Accident Grade
	– Job - Districts - Day of Week - Accident Type
	– Job - Districts - Day of Week - Accident Grade
	– Job - Level of Education - Districts - Accident Type
	– Job - Level of Education - Districts - Accident Grade
Wagon	– Train Type - No. of Wagons (Train Length) - Districts - Accident Type
	– Train Type - No. of Wagons (Train Length) - Districts - Accident Grade
	– Train Type - Wagons Type - Districts - Accident Type
	– Train Type - Wagons Type - Districts - Accident Grade
	– Wagons Type - Tunnel - Districts - Accident Type
	– Wagons Type - Tunnel - Districts - Accident Grade
	– Wagons Type - Curve Radius - Districts - Accident Type
	– Wagons Type - Curve Radius - Districts - Accident Grade
Track	– Train Type - Gradient - Districts - Accident Type
	– Train Type - Gradient - Districts - Accident Grade
	– Train Type - Curve Radius - Districts - Accident Type
	– Train Type - Curve Radius - Districts - Accident Grade
	– Train Type - Tunnel - Districts - Accident Type
	– Train Type - Tunnel - Districts - Accident Grade
	– Train Type - Point - Districts - Accident Type
	– Train Type - Point - Districts - Accident Grade
	– Curve Radius - Speed - Districts - Accident Type
	– Curve Radius - Speed - Districts - Accident Grade

General	– Accident Factor - Districts - Accident Type – Accident Factor - Districts - Accident Grade – Accident Factor - Day of Week - Accident Type – Accident Factor - Day of Week - Accident Grade – Accident Factor - Day of Week – Accident Factor - Accident Type – Districts - Accident Type – Accident Factor - Site – Accident Type - Site – Districts - Site – Accident Factor - Speed – Speed - Accident Type – Districts - Accident Grade

In order to study meaningful relations, the experience of the researchers along with priorities of railway safety experts of RAI, resulted in a list of candidate relations for studying association rules. Table 1 holds the fields examined in relation to one another by considering accident factors. Among the association rules examined, those contributing to loss of life and property are considered.

Loss of life and property are weighted as follows:

- Non-RAI people killed = 5, RAI personnel killed = 4, Non-RAI people injured = 3, RAI personnel injured = 2.
- Loss of 0-50 million Rials = 5, Loss of 50-500 million Rials = 4, Loss of 500-1500 million Rials = 3, Loss of 1500-2000 million Rials = 2, Loss of 2000 million Rials and higher = 1

Appendix 2 (Tables 6 to 18) hold samples of some association rules extracted from GRI algorithm. Each table holds confidence and support for each relation, and losses of life and property along with their confidence and support.

Evaluation: Evaluating Model outcomes

Regardless what some software vendor advertisements may claim, data mining software just can't be purchased and installed, and you can't sit back and watch it solve all the problems. Data mining is not magic. Without skilled human supervision, blind use of data mining software will only provide you with the wrong answer to the wrong question applied to the wrong type of data. The wrong analysis is worse than no analysis, since it leads to policy recommendations that will probably turn out to be just and expensive failure. Therefore results of software must be evaluated by human experts.

In this phase, further review on related accident records archived in RAI Safety Dept and consultation with safety experts were applied to evaluate the rationality of the results found. Development in the next phase is based on outcomes of Evaluation Phase.

Development: Safety Regulations Determination

If a system would be considered as a chain and its links as the system elements, once the chain is put under pressure, it tars from the weak link. The railway system is somehow like a chain and when accidents happen, we have got a weak link at that part of the chain. Once we recognize these points, they can be replaced with stronger links and help the system perform more efficiently in the future.

With the application of data mining, we have recognized such links by discovering repetitive patterns within the past accidents data. Up to this point, we have recognized areas ware problems have lied in a general point of view. This section covers rules witch have been suggested according to past findings to prevent past trends to be repeated. Some of these rules might already exist in RAI's rule books, and some may not. Following is an emphasis on those rules witch exists and a suggestion on those witch don't. There are three main areas for the mentioned rules: 1) human resource (Table 2), 2) track (Table 3), 3) wagons and freight (Table 4).

Table 2: Safety Rules Related to Human Resource

Category	Rules
Workers Uniforms	– Different working groups must have uniforms witch differ from other groups in style and color. – Wearing close other than uniforms on duty is banned. – Design and material of close must be customized to the seasons of the year, weather conditions of the district, type of work to ease activities performed by the workers (for example preparing suitable hats, work gloves, coat, pots, etc). – Workers must use phosphoric covers over their close out doors at night. – Designing proper pockets on the close for workers to ease carrying and accessibility to working equipment.
Work and Rest Hours	– Least working hours of station personnel is 8 hours per day. – Most continual working hours of station personnel is 12 hours. – Most continual working hours of station personnel should be no more than 18 hours. – Least obligatory resting hours for station personnel is 8 hours per day. – Spreading station operations threw out the week as equally as possible.
Training	– Workers who engage permanently or periodically in a work must have passed the related course or apprenticeship on that specific task or working area and received the certificate. – Training workers on a specific task or working area can be done before full employment or frequently after employment, depending on the duties. – Workers training must consist of sufficient theoretical and practical hours and this must be mentioned in the issued certification. – All personnel must pass occupational safety, accident encountering, and crises handling courses related to the tasks they perform and working area and receive the issued certification. – Passing required courses according to condition of accidents involved in, in the past.
Attaining Qualified and Promotion	– Personnel must embody sufficient physical, sense of sight, hearing, and mental condition. – Examine personnel's records for bad record or addiction every two years. – Giving priority to domestic work forces to occupy operational job possessions. – Having passed necessary apprenticeship or training courses according to tasks and area of activity, and receiving issued certification. – Having passed occupational safety, accident encountering, and crises handling courses according to their job and working area. – Tacking part in the periodic exams to extend validity of safety certifications with acceptable point. – Being in the age range defined for the job. – Possessing required level of education, intelligence, and social skills according to job. – Personnel promotion is to happen by support of required knowledge, experience, and results of personnel appraisals threw out the working years.
Reward and Welfare	– Periodically rewarding personnel with safe performance or causing the least accidents in compliance with their job. – Planning shorter working periods for those who work in intermediate stations, especially in areas with bed climates, in a way that workers with similar conditions take similar shifts. – Considering years of experience, marital situation, experience, endemic, etc in human resource planning. – Equipping bad climatic areas with CTC to reduce the number of workers and remote controlling stations with higher accuracy. – Intermediate stations, especially those in bad climatic areas, with cooling, heating, comfort, hygienic, lamination, communication, safety and security facilities. – Providing houses for workers near their working area or monetary contribution for them to buy one on their own. – Providing transport service for workers to get to work and return home.
Other Facilities	– Providing facilities in big stations for workers whose duties require getting around the stations a lot, in order to facilitate work. – Documenting personnel functions accurately in their records.

Table 3: Safety Rules Related to Track

Category	Rules
Inspections	– Track parts, turnouts and crossings must be inspected in standard periods defined according to technical recommendations, area and operational conditions. – Providing machines, handheld devices, etc for inspections according to track inspection and measurement standards. – Inspections must not be postponed for the sake of traffic congestion.

	– Results of inspections delivered to related units at once for actions to be planned and taken place. – Inspection results must be documented and entered in related databases for future analysis.
Mainten-ance	– Performing periodic services and maintenance in blocks, in tunnels, on bridges, and in stations according to their valid technical standards and avoid postponing for the sake of traffic congestion. – Using safety maintaining devices at both ends of section under repair to avoid collision with rolling stocks while working. – Using Speed reduction signs or warnings in adjacent lines and about the area under repair. – To make sure signals and points are in a condition that shows that segment occupied.

Table 4: Safety Rules Related to Wagons

Category	Rules
Carrying Dangerous Goods	– Before moving dangerous goods wagons, air pipes must connected, air breaks must be in right position and container wagon's doors must be closed. – Brakes of dangerous goods wagons stabled in stations must be fastened and drag shoes must also be used. – Will moving a dangerous goods wagon, it's recommended a shunting man be ready at wagon's brakes. – Such wagons must be coupled to avoid unwanted movements. – Other wagons must not hit dangerous goods wagons will shunting. – Dangerous goods wagons must be 12 axles away from the locomotive. – Dangerous goods wagons must be provided with relative signs and warnings. – Holding and parking tracks for dangerous goods wagons must be separated and far from other wagons. – The speed dangerous goods wagons hit other wagons or locomotives must be no more than 3 km/h. – Wagons containing fast flammable goods must not be carried with cotton in one train. – Workers involved in shunting must be aware of the dangerous goods within wagons. – Stabling dangerous goods wagons else ware their final destination must not happen. – If dangerous goods wagons are to be stopped along the way in intermediate stations in case of emergencies, they must be kept far form buildings, brakes fastened and safety procedures taken care of. – Extra space must be considered in tank wagons so if the liquid inside expands, it would not leak or deform the tank. – All stations and trains must be equipped with fir extinguishers. – In case of big fires, announcement facilities to fire departments must be available. – Material used in different parts threw out stations and trains must be made of substances with do not catch fire or hard to burn, so speed of fire would be as slow as possible. – Training station and train workers to guide people away from the dangerous area. –

Conclusion

This data mining project had been defined with the objective of determining hidden relationships of the most common accidents factors, witch are human error, wagon and track, with other fields of the RAI accidents database. Following CRISP-DM reference model, some 6500 records and 38 fields of data from RAI's accident database throughout the years 1996 to 2005 were applied for mining. Association rules was chosen as the data mining technique. Ultimately safety rules are defined to prevent the discovered patterns to happen in future. These rules have been categorized in three main areas as follows:

1. Human Resources
    - Workers Uniforms
    - Work and Rest Hours
    - Training
    - Attaining Qualified and Promotion
    - Reward and Welfare
    - Other
2. Track
    - General
    - Inspections
    - Maintenance
3. Wagons
    - Stabling Wagons in Stations
    - Carrying Dangerous Goods

Discussions

Studies definitely does not stop hear. There are other data mining techniques witch their application would be useful. One of them is time series. A time series is an ordered collection of measurements taken at regular intervals. The measurements may be of anything that interests you. Methods of modeling time series assume that history repeats itself - if not exactly, then closely enough that by studying the past, you can make better decisions in the future. By this technique you predict the values of one or more series over time. For example, you may want to predict the expected occurrence of specific accidents threw out the year in a particular district or on the entire network.

Studying the past behavior of a series will help you identify patterns and make better forecasts. When plotted, many time series exhibit one or more of the following features:

- Trends
- Seasonal and non-seasonal cycles
- Pulses and steps
- Outliers

Because planning decisions take time to implement, forecasts are an essential tool in many planning processes. The continuation of this project will concentrate on discovering seasonal and non-seasonal cycles, pulses and steps, and Outliers.

References

[1] Huang, T.M., Kecman, V., Kopriva, I., "Kernel Based Algorithms for Mining Huge Data Sets," Springer Berlin Heidelberg New York, 2006.

[2] Giudici, P, "Applied Data Mining: Statistical Methods for Business and Industry," John Wiley & Sons Inc., 2003.

[3] Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining," John Wiley & Sons Inc., 2005.

[4] Chang, L.Y., Wang, H.Y., "Analysis of traffic injury severity: An application of non-parametric classification tree techniques," Accident Analysis and Prevention 38 (2006) 1019–1027, 2006.

[5] Lee, D.H, Jeng, S.T., Chandrasekar, P., "Applying Data Mining Techniques for Traffic Incident Analysis," Journal of The Institution of Engineers, Singapore, Vol. 44 Issue 2, 2004.

[6] Tao, Y., Xinmiao, Y., "Fuzzy Comprehe-nsive Assessment, Fuzzy Clustering Analysis and its Application for Urban Traffic Environment Quality Evaluation," PII: S1361-9209(97)00021-7, 1998.

[7] Nagatani, T., Emmerich, H., Nakanishi, K., "Burgers equation for kinetic clustering in traffic flow", Physica A 255, P. 158-162, 1998.

[8] Vlahogianni, E., I., Karlaftis, M., G., Golias, J., C., "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," Transportation Research Part C 13, P. 211–234, 2005.

[9] Pande, A., Abdel-Aty, M., "Assessment of freeway traffic parameters leading to lane-change related collisions," Accident Analysis and Prevention 38, P. 936–948, 2006.

[10] Celikoglua, H.B., Cigizoglub, H.K., "Modeling public transport trips by radial basis function neural networks," Mathematical and Computer Modeling 45, P. 480–489, 2007.

[11] Celikoglu, H.B., Cigizoglu, H.K., "Public transportation trip flow modeling with generalized regression neural networks," Advances in Engineering Software 38 (2007) 71–79, 2007.

[12] Amado, V., "Expanding the use of Pavement Management Data. In: Transportation Scholars Conference 2000," University of Missouri, 2000.

[13] Attoh-Okine, N.O., "Combining Use of Rough Set and Artificial Neural Networks in Doweled Pavement Performance Modeling – A Hybrid Approach," In Proceedings 80th Annual Meeting of TRB, Washington, D.C., 2001.

[14] Chong, M.M., Abraham, A., Paprzycki, M., "Traffic Accident Analysis Using Decision Trees And Neural Networks," IADIS International Conference on Applied Computing, Portugal, IADIS Press, Pedro Isaias et al. (Eds.), ISBN: 9729894736, Volume 2, pp. 39-42, 2004.

[15] Chong, M., Abraham, A., Paprzycki1, M., "Traffic Accident Analysis Using Machine Learning Paradigms," Informatica 29 (2005) 89–98 89, 2005.

[16] Wang, H., Parrish, A., Smith, R.K., Vrbsky, S., "Improved variable and value ranking techniques for mining categorical traffic accident data," Expert Systems with Applications 29 (2005) 795–806, 2005.

[17] Solomon, S., Nguyen, H., Liebowitz J., Agresti, W., "Using data mining to improve traffic safety programs," Industrial Management & Data Systems, Vol. 106 No. 5, 2006, pp. 621-643, q Emerald Group Publishing Limited, 2006.

[18] Xie, Y., Lord, D., Zhang, Y., "Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis," Accident Analysis and Prevention 39 (2007) 922–933, 2007.

[19] Factora, R., Mahalela, D., Yair, G., "Inter-group differences in road-traffic crash involvement," Accident Analysis and Prevention 40 (2008) 2000–2007, 2008.

[20] Graves, S.J., Rochowiak, D., Anderson, M., "Mining and Analysis of Traffic Safety and Roadway Condition Data," University Transportation Center for Alabama (UTCA), 2005.
[21] Keuleers, B., Wets, G., Aremtze, T. and Timmermans, H., "Using Association Rules to Identify Spatial-Temporal Patterns in Multi-Day Activity Diary Data," In Proceedings 80th Annual Meeting of TRB, Washington, D.C., 2001.
[22] Wet, G., Vanhoof, K., Aremtze, T., Timmermans, H., "Identifying Decision Structures Underlying Activity Patterns: An Exploration of Data Mining Algorithms," In Proceedings 79th Annual Meeting of TRB, Washington, D.C., 2000.
[23] Sheu, J.B., "A fuzzy clustering approach to real-time demand-responsive bus dispatching control,' Fuzzy Sets and Systems 150 (2005) 437–455, 2005.
[24] Yao, X., "Where are public transit needed – Examining potential demand for public transit for commuting trips, Computers," Environment and Urban Systems 31 (2007) 535–550, 2007.
[25] Choi, K., Chung, Y.S., "Travel Time Estimation Algorithm Using GPS Probe and Loop Detector Data Fusion," In Proceedings 80th Annual Meeting of TRB, Washington, D.C., 2001.
[26] Zhang, X., Rice, J.A., "Short-term travel time prediction," Transportation Research Part C 11 (2003) 187–210, 2003.
[27] Regan, A.C., Song, J., "An Industry in Transition: Third Party Logistics in the Information Age," In Proceedings 80th Annual Meeting of TRB, Washington, D.C., 2001.
[28] Clementini, E., Felice, P.D., Koperski, K., "Mining multiple-level spatial association rules for objects with a broad boundary," Data & Knowledge Engineering 34 (2000) 251-270, 2000.
[29] Mas, J.F., Puig, H., Palacio, J.L., Sosa-Lopez, A., "Modeling deforestation using GIS and artificial neural networks," Environmental Modeling & Software 19 (2004) 461-471, 2004.
[30] Ermini, L., Catani, F., Casagli, N., "Artificial Neural Networks applied to landslide susceptibility assessment," Geomorphology 66 (2005) 327–343, 2005.
[31] Zhang, B., Valentine, I., Kemp, P., Lambert, G., "Predictive modeling of hill-pasture productivity: integration of a decision tree and a geographical information system," Agricultural Systems 87 (2006) 1–17, 2006.
[32] Jin, X., Davis, C.H., "Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks," Image and Vision Computing 25 (2007) 1422–1431, 2007.
[33] Yoo, C., Kim, J.M., "Tunneling performance prediction using an integrated GIS and neural network," Computers and Geotechnics 34 (2007) 19–30, 2007.
[34] Papernik, D.K., Nanda, D., Cassada, R.O. Morris, W.H., "An Information Strategy to Enable the Analysis of Performance Measures," In Proceedings 79th Annual Meeting of TRB, Washington, D.C., 2000.
[35] Williams, T.P., "Bidding ratios to predict highway project costs, Engineering," Construction and Architectural Management Vol. 12 No. 1, 2005 pp. 38-51, 2005.
[36] Chou, J.S., O'Connor, J.T., "Internet-based preliminary highway construction cost estimating database," Automation in Construction 17 (2007) 65–74, 2007.
[37] Li, M., Fan, Y., Chen, J., Gao, L., Di, Z., Wu, J., "Weighted networks of scientific communication: the measurement and topological role of weight," Physica A 350 (2005) 643–656, 2005.
[38] Avineri, E., Prashker, J.N., "Sensitivity to travel time variability: Travelers_learning perspective," Transportation Research Part C 13 (2005) 157–183, 2005.
[39] Dondo, R., Cerda, J., "A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows," European Journal of Operational Research 176 (2007) 1478–1507, 2007.
[40] Y.K. Wong, W.L. Woon, "An iterative approach to enhanced traffic signal optimization," Expert Systems with Applications 34 (2008) 2885–2890, 2008.
[41] Barai, S.K., "Data mining applications in transportation engineering," Transport. Vilnius: Technika, Vol. 18, No. 5, p. 216-223, 2003.
[42] Hou', Z.G., Gupta', M.M., "A Rail Damage Detection and Measurement System Using Neural Networks," CIMSA 2004 - EEE international Conference on Computational Intelligence for Measurement Systems and Applications Boston, MA, USA. 14-16 July 2004.

[43] Nefti, S., Oussalah, M., "A Neural Network approach for Railway Safety Prediction," 2004 IEEE International Conference on Systems, Man and Cybemetics, 2004.

[44] Shahin M.A., Indraratna, B., "Modeling the mechanical behavior of railway ballast using artificial neural networks," Can. Geotech. J. 43: 1144–1152 (2006) doi: 10.1139/T06-077 © 2006 NRC Canada, 2006.

[45] Sekine, S., Imasaki N., Endo, T., "Application of Fuzzy Neural Network Control to Automatic Train Operation and Tuning Of Its Control Rules, 0-7803-2461-7/95/$4,00 0 1995 IEEE, 1995.

[46] Rubaai, A., "A Neural-Net-Based Device for Monitoring Amtrak Railroad Track System," IEEE Transactions On Industry Applications, Vol. 39, No. 2, March/April 2003.

[47] Sacchi, C., Regazzoni, C., Gera, G., Foresti, G., "Use of Neural Networks for Behaviour Understanding in Railway Transport Monitoring Applications," 0-7803-6725- 1/0 2001 IEEE, 2001.

[48] Sekine, S., Nishimurat, M., "Application of Fuzzy Neural Network Control to Automatic Train Operation," 0-7803-2461-7/95/$4.00 0 1995 IEEE, 1995.

[49] Martinelli D.R., Teng, H., "Optimization of Railway Operations Using Neural Networks, Transpn," Res.-C, Vol. 4, No. 1, pp. 3349, 1996, Couvrieht @I? 1996 Elsevier Science Ltd., Printed'& GreatBritain. All rights reserved, 1996.

[50] Li, D., Meddah, A., Hass, K., Kalay, S., "Relating track geometry to vehicle performance using neural network approach," Proceedings of the Institution of Mechanical Engineers; Sep 2006; 220, F3; ProQuest Science Journals pg. 273, 2006.

[51] Lee, J., Han, S., "A neural network approach to estimate the top speed and trainset configuration," Proceedings of the Institution of Mechanical Engineers; Mar 2004; 218, 1; ProQuest Science Journals pg. 41, 2004.

[52] Kerh, T., Ting, S.B., "Neural network estimation of ground peak acceleration at stations along Taiwan high-speed rail system," Engineering Applications of Artificial Intelligence 18 (2005) 857–866, 2005.

[53] Chow, T.W.S., Shuai, O., "Feedforward Neural Networks Based Input–Output Models for Railway Carriage System Identification," Neural Processing Letters 5: 127–137, Kluwer Academic Publishers. Printed in the Netherlands, 1997.

[54] Liu, J.N.K., Sin, K.Y., "Fuzzy Neural Networks for Machine Maintenance in Mass Transit Railway System," IEEE Transactions on Neural Networks, Vol. 8, No. 4, July 1997.

[55] Holmgren, M., "Maintenance-related losses at the Swedish Rail," Journal of Quality in Maintenance Engineering, Vol. 11 No. 1, 2005 pp. 5-18, 2005.

[56] Glickman, T.S., Erkut, E., Zschocke, M.S., "The cost and risk impacts of rerouting railroad shipments of hazardous materials," Accident Analysis and Prevention 39 (2007) 1015–1025, 2007.

[57] Oha, J., Washington, S.P., Nama, D., "Accident Prediction Model For Railway-Highway Interfaces," Accident Analysis and Prevention 38 (2006) 346–356, 2006.

[58] Lobb, B., Harre, N., Suddendorf, T., "An evaluation of a sub-urban railway pedestrian crossing safety programme," Accident Analysis and Prevention 33 (2001) 157–165, 2001.

[59] Saccomanno, F.F., Park, P.Y.J., Fua, L., "Estimating countermeasure effects for reducing collisions at highway–railway grade crossings," Accident Analysis and Prevention 39 (2007) 406–416, 2007.

[60] Lobb, B., Harré, N., Terry, N., "An evaluation of four types of railway pedestrian crossing safety intervention," Accident Analysis and Prevention 35 (2003) 487–494, 2003.

[61] Davey, J., Wallace, A., Stenson, N., Freeman, J., "The experiences and perceptions of heavy vehicle drivers and train drivers of dangers at railway level crossings," Accident Analysis & Prevention, Volume 40, Issue 3, May 2008, Pages 1217-1222, 2008.

[62] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., CRISP-DM 1.0: Step-by-step data mining guide," SPSS Inc., 2000.

[63] RAI, http://www.rai.ir/, 2006.

[64] Han, J., Kamber, M., "Data Mining: Concepts and Techniques," Second Edition, Elsevier Inc., 2006.

Appendix 1

Table 5: Chosen Fields for Data Mining

#	Field Title	Definition
1.	Accident key	Unique key for each accident
2.	Date	YY/MM/DD on which accident happened
3.	Time	HH/MM on which accident happened
4.	Day of Week	Day of which accident happened starting from Sat.
5.	District	RAI District on which accident happened in
6.	Station Before	One end of the block accident happened on
7.	Station After	One end of the block accident happened on[3]
8.	Kilometer	Kilometer from Tehran on which accident happened in
9.	No. of RAI Personal Killed	Number of railway personnel killed in the accident
10.	No. of RAI Personal Injured	Number of railway personnel injured in the accident
11.	No. of People Killed	Number of non-railway personnel killed in the accident (passengers, etc)
12.	No. of People Injured	Number of non-railway personnel injured in the accident (passengers, etc)
13.	Losses of Track Properties	Financial Damage to Track (in Rials[4])
14.	Losses of Locos Properties	Financial Damage to Locomotives (in Rials)
15.	Losses of Wagons Properties	Financial Damage to Wagons (in Rials)
16.	Losses of Other Properties	Other Financial Damages (in Rials)
17.	Accident Type	Type of the accident happened including: collision of rolling stocks with other rolling stocks, collision of rolling stocks with other obstacles, derailment, harm to passenger, harm to RAI personnel, fire accident (in locomotives, wagons, coaches), other.
18.	Accident Factor	Cause of the accident consisting of: Human Error, Loco, Wagon, Track, Signaling, Rolling Stock, Other.
19.	Accident Grade	Severity of the accident consisting of: I1, I2, D1, D2, D3 (from worst to least).
20.	Accident Class	Class of the accident consisting of: Railway, Non-railway, Fatality and injury.
21.	Gradient	Steepness of the track at the accident location
22.	Curve	Curve Radius at the accident location
23.	Tunnel	If there is or not a tunnel at the accident site
24.	Point	If there is or not a point at the accident site
25.	Train Type	Type of the train consisting of: passenger, freight, mixed, service, locomotive, shunting, other
26.	Train No.	RAI assigned unique number to the crashed train
27.	Wagon No.	RAI assigned unique number to the crashed wagon
28.	Locomotive No.	RAI assigned unique number to the crashed locomotive
29.	Train Weight	Weight of the crashed train
30.	Braked Weight	Braked weight of the train
31.	No. of Wagons	Number of the wagons involved in the accident
32.	Train Speed	Speed of the train before accident
33.	Age	Age of the accident culprit
34.	Years of service	Years of service of the accident culprit
35.	Marital Situation	Marital situation of the accident culprit including single or maries.
36.	Level of Education	Level of education of the accident culprit including: illiterate, elementary school, guidance school, high school, graduate, AA, BA, MA, PhD.
37.	Job	Job of the accident culprit (witch includes 52 job titles)
38.	Emp. Condition	Employment condition of the accident culprit consisting of: governmental official, daily-paid, railway official, hourly, contracted

[3] Accidents which happened in stations have the same station before and after.
[4] Rial is Iran's currency.

Appendix 2: samples of some association rules extracted from GRI algorithm

Table 6: Job – Years of Service – Districts – Accident Type

Job	Years of Service	Districts	Accident Type	Supp-ort	Confi-dence	Loss of Life	Support/ Confidence		Loss of Property	Supp-ort	Confi-dence
Shunting man	16 – 20	Hormozgan	fire accident	0.26	5	1	0.11	100	4 5	0.75 0.99	91.18 93.33
Assistant driver	6 – 10	Khorasan	fire accident	0.31	57.14	1	0.26	83.33	4 5	0.88 1.59	77.5 88.89

Table 7: Job – Years of Service – Districts – Accident Grade

Job	Years of Service	Districts	Accident Grade	Supp-ort	Confi-dence	Loss of Life	Support/ Confidence		Loss of Property	Supp-ort	Confi-dence
Assistant driver	6 – 10	Khorasan	D3	0.55	56	1	0.26	100	4 5	0.88 1.59	100 95.83

Table 8: Job – Age – Districts – Accident Grade

Job	Age	Districts	Accident Grade	Supp-ort	Confi-dence	Loss of Life	Support/ Confidence		Loss of Property	Supp-ort	Confi-dence
Driver	21 – 30	Hormozgan	D3	0.49	90.91	4	1.92	78.16	4 5	0.75 0.99	91.18 100

Table 9: Train Type – No. of Wagons (Train Length) – Districts – Accident Type

Train Type	No. of Wagons	Districts	Accident Type	Supp-ort	Confi-dence	Loss of Life	Support/ Confidence		Loss of Property	Supp-ort	Confi-dence
Fraught	35	Azar-bayejan	collision of RS with RS	0.4	61.11	14	0.7	100	3	0.13	83.33
Other	0	Tehran	fire accident	0.42	100	4 5	1.99 2.05	86.67 97.85			
Loco-motive	0	South east	collision of RS with O	0.46	100				4	1.48	71.64
Freight	25	North west	collision of RS with RS	0.33	66.67				3	0.26	83.33
Loco-motive	0	Esfehan	collision of RS with RS	0.44	60				3	0.13	66.67
Shunting	0	Lorestan	derailment	0.99	51.11	2	0.04	100			

Table 10: Train Type – Tunnel – Districts – Accident Grade

Train Type	Tunnel	Districts	Accident Grade	Supp-ort	Confi-dence	Loss of Life	Support/ Confidence		Loss of Property	Supp-ort	Confi-dence
Freight	Yes	Lorestan	D3	0.49	54.55	5	0.73	100	3	0.15	57.14

Table 11: Train Type – Tunnel – Districts – Accident Type

Train Type	Tunnel	Districts	Accident Type	Supp-ort	Confi-dence	Loss of Life	Support/ Confidence		Loss of Property	Supp-ort	Confi-dence
Fright	Yes	Lorestan	collision of RS with RS	0.44	65				3	0.15	100

Table 12: Accident Factor – Districts – Accident Type

Accident Factor	Districts	Accident Type	Supp-ort	Confi-dence	Loss of Life	Support/ Confidence		Loss of Property	Supp-ort	Confi-dence
Human error	Arak	fire accident	0.46	100	3 5	0.15 0.26	100 100			
Human error	Hormozgan	fire accident	1.81	100	3 4 5	0.11 0.75 0.99	100 91.18 93.33	1	0.11	100
Human error	Esfehan	fire accident	0.97	100	4 5	0.31 0.71	100 90.62			
Human error	Tehran	fire accident	3.89	100	4 5	1.99 2.05	86.67 97.85			
Human	South	fire	0.93	100	4	0.22	90			

error		accident			5	0.73	100			
Human error	South east	derailment	0.77	100	2 6	0.07 0.11	100 100			
Human error	Khorasan	fire accident	2.12	100	4 5	0.88 1.59	77.5 88.89	1	0.26	83.3
Wagon	Arak	collision of RS with O	0.6	77.78				4	0.15	100
Wagon	Hormozgan	collision of RS with O	0.82	94.59				3	0.13	100
Wagon	Tehran	collision of RS with O	1.92	73.56				3	0.09	100
Wagon	South	collision of RS with O	0.57	80.77				3	0.07	100
Wagon	South east	collision of RS with O	1.81	65.85				4	1.48	71.64
Wagon	Khorasan	collision of RS with O	0.38	70.59				4	0.29	84.62
Wagon	North	collision of RS with O	0.84	84.21				4	0.29	100

Table 13: Accident Factor – Day of Week – Accident Grade

Accident Factor	**Day of Week**	**Accident Grade**	**Supp-ort**	**Confi-dence**	**Loss of Life**	**Support/ Confidence**		**Loss of Property**	**Supp-ort**	**Confi-dence**
Human error	Sat.	D1	2.3	56.73				1	0.2	77.78
Human error	Sun.	I1	1.06	91.67				1	0.46	95.24
Human error	Sat.	D3	9.71	65.46	2 5	5.08 0.42	61.74 52.63	3 4	0.13 1.02	83.33 76.09

Table 14: Job – Accident Type

Job	**Accident Type**	**Support**	**Confidence**
Driver	Collision of RS with O	24.42	51.36

Table 15: Job – Site

Job	**Site**	**Support**	**Confidence**
Driver	Station	24.42	73.42
Assistant driver	Station	7.51	70.29
Shunting man	Station	7.42	95.83

Table 16: Accident Factor – Day of Week

Accident Factor	**Day of Week**	**Support**	**Confidence**
Human error	Sat.	14.08	68.97
Human error	Sun.	16.14	72.23

Table 17: Accident Factor – Accident Type

Accident Factor	**Accident Type**	**Support**	**Confidence**
Human error	collision of RS with O	46.29	59.9
Human error	derailment	6.36	100
Human error	fire accident	15.54	98.58
Wagon	collision of RS with O	13.64	69.9

Table 18: Accident Factor – Site

Accident Factor	**Site**	**Support**	**Confidence**
Human error	Station	73.25	73.96
Wagon	Station	13.64	59.39