INTEGRATION OF APRIORI ALGORITHMS WITH CASE-BASED REASONING FOR FLIGHT ACCIDENT INVESTIGATION

Nan-Hsing Chiu
Department of Information Management, Ching Yun University, Jhongli city, Taoyuan, Taiwan, R.O.C.

Pei-Da Lin
Occurrence Investigation Division, Aviation Safety Council, Taipei, Taiwan, R.O.C.

Chang En Pu
Investigation Bureau of MOJ, Taipei, Taiwan, R.O.C.

Keywords: Accident Investigation, Decision Support Systems, Case-based Reasoning, Apriori.

Abstract: The analysis of flight accidents has been demonstrated to be a crucial tool for improving flight safety. The utilization of visual decision support systems potentially assists investigators in quickly and accurately identifying the underlying causes of accidents. This study aims at supplying a visual decision support system, based on the Apriori and case-based reasoning approaches, for assisting investigators in analyzing human injuries in flight accidents. We demonstrate our approach using the aircraft configuration of flight CI611. The experimental results show that the proposed approach provides support for quick decisions by investigators on the basis of a visualization system.

1 INTRODUCTION

With modern aviation technology, airplanes have become one of the safest means of transportation. There were about 21,380 flight incidents and accidents between 1989 and 1999 (Weng et al., 2003). It is important in any flight accident to ensure that human injuries are as few as possible. The investigation of aircraft accidents is focused on the circumstances of the accidents, including analyzing all available information in order to draw conclusions. The objective of the investigation of flight accidents is to enhance air traffic safety by introducing recommendations intended to prevent accidents in the future.

An investigation of an aircraft accident engages a variety of personnel: specialists, experts, legal authorities, and accredited representatives, including representatives for agents of design, manufacture, operation, and so on (Milosovski et al., 2008). Investigators need to pursue a comprehensive examination of the accident site, wreckage, witness information, any recorded media, component examinations, tests, simulations, and other evidence in order to determine the cause of an accident. This shows how an aircraft accident investigation is a complex issue involving many different factors.

A detailed analysis of flight circumstances is an important basis of investigation and is essential in identifying the underlying factors leading to accidents. The Apriori algorithm is a well-known association rule approach that is able to derive frequent itemsets from a variety of datasets (Cristian & Mitica, 2007). The case-based reasoning (CBR) approach also demonstrates high reliability on similar measurements among diverse data sources (Chiu & Huang, 2007). Therefore, this study aims to combine these two approaches for visually investigating and analyzing cases of human injury during flight accidents.

Chiu N., Lin P. and Pu C. (2010).
INTEGRATION OF APRIORI ALGORITHMS WITH CASE-BASED REASONING FOR FLIGHT ACCIDENT INVESTIGATION.
In *Proceedings of the 5th International Conference on Software and Data Technologies*, pages 487-490
DOI: 10.5220/0002922204870490

2 LITERATURE REVIEW

The improvement of flight safety involves attending not only to factors that increase the likelihood of crashes occurring but also to factors that increase the likelihood of both fatal and non-fatal injury in the crashes that do occur. In 2003, O'Hare et al. (2003) utilize integrated database information to identify risk factors for both fatal and non-fatal outcomes for all civil aviation crashes. Their case study shows that the most significant factors were post-crash fire, off-airport crash location, aerobatic flight, lack of an airworthiness certificate, and aircraft classified as microlight. Taneja & Wiegmann (2003) performed a descriptive study of fatal injuries based on autopsy reports. Their results demonstrate that the most significant factors of fatal injury are post-crash fire and lack of shoulder harness use. Each flight accident usually involves unique reasons for fatal and non-fatal outcomes. Given the diversity of aircraft and autopsy reports, a user-friendly tool would play an important role for investigators to be able quickly to identify the key factors in human injury in flight accidents.

Providing investigators with a summary of information from an accident's injury data plays a crucial role in determining the probable causes of injury. Agrawal et al. (1993) introduce mining association rules based on the Apriori algorithm. This algorithm finds frequent itemsets and is the most popular algorithm for association rules discovery. Chiu et al. (2006) utilized the benefits of an approach using the Apriori algorithm for identifying the malfunction of electronic ballasts in an aircraft. They compared the Apriori algorithm with a simple genetic algorithm. Their experimental results demonstrate that their approach achieves higher accuracy in distinguishing and classifying malfunctions in aircraft.

Because human injuries with the exact same type are usually rare occurences within the overall dataset, providing investigators with a visualization of similar cases is a useful approach for understanding frequent itemsets and their distribution in a given aircraft accident. CBR is a data-intensive method that matches an input pattern with all the information in the database. It searches the repository for all existing datasets that have attributes similar to those of the new dataset and retrieves the nearest measurement value as estimates. Paul et al. (2001) present a CBR approach for quantifying confidence in the probability that a program is free of specific classes of defects. Their results show that it is possible to analyze Y2K defects based on defect data. Chiu & Huang (2007) investigate the effect of CBR on the improvement of estimation accuracy. Their empirical results show that their proposal is a feasible approach for improving the estimation abilities.

CBR is a well-known approach for solving problems on the basis of similar measurements. Thus, for quickly identifying similar cases for investigators, , this study utilizes an approach based on the visualization of aircraft configuration in order to illustrate integrated information. This integrated information is derived by combining the association rule approach of the Apriori algorithm with the similar measurements used in CBR for further improvement in the aircraft accident investigation process.

3 METHODOLOGY

Figure1 shows the framework for visual analysis of human injuries in flight accident investigations and is based on the integration of the Apriori and CBR (ACBR) approaches. The flight accident database includes three data sources: a passenger database, a family member database and an autopsy report database. The passenger database includes all of the basic data for the flight crews and passengers. The family member database includes deoxyribonucleic acid (DNA) data for family members of the flight crews and passengers in order to match DNA. The autopsy reports show DNA and relative information about the bodies of the dead. They also describe relative information for passengers with slight or serious injuries. The DNA matching stage ensures correct connections among passenger database entries, family member database entries and autopsy reports.

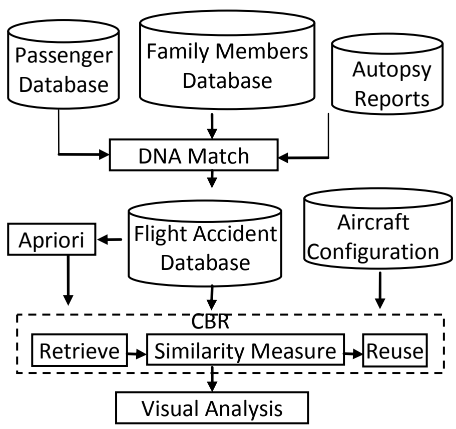

Figure 1: A framework for ACBR.

The Apriori algorithm is adopted for deriving frequent itemsets from the flight accident database as cases for analysis. Using the Apriori algorithm, human injuries with categorical attributes are used to construct association rules for determining the frequent itemsets. Association rules are statements of the form “if antecedent(s), then consequent(s).” It is an implication of the form $a \rightarrow b$, where a is the conjunction of conditions and b is the result of their association. Support refers to the percentage of records in the training data for which the antecedents (the "if" part of the rule) are true. Confidence is based on the records for which the rule's antecedents are true and is the percentage of those records for which the consequent(s) are also true. Given a set of human injury datasets, we wish to generate all the association rules that have greater support and confidence than the user-specified minimum support and confidence.

The CBR approach for visual analysis of human injuries includes three stages: retrieval, similarity measure and reuse. For a case under analysis, this approach retrieves the human injury cases from the flight accident database. The distance between the case under analysis and other human injury cases in the flight accident database is measured. Among all human injury cases, the most similar one will show the lowest distance from the case under analysis. As each human injury case has a respective distance indicating similarity, the visual analysis system plots similar cases using different colors on the aircraft configuration image. On the basis of visual recognition, investigators can easily and quickly understand similar cases from a large number of diverse data sources.

The similarity measure of the CBR approach is shown in Equation (1), where C_i is the case being clustered and C_h is the human injury case from the flight accident database. For each case being clustered, there are n feature values in conjunction with feature weight w. Investigators are able to assign feature weights for different features, depending on their requirements. Using the values of features normalized between 0 and 1, the similarity measurement determines the straight-line distance between two features of a human injury case. The distance decreases as similarity increases. The sum of the squares of the distance for each feature is the square of the distance between two human injury cases. Therefore, the closest human injury case to the case C_i is the case with the minimum distance.

$$\delta(C_i, C_h) = \sqrt{\sum_{f=1}^{n} W_f \times (C_{if} - C_{hf})^2} \quad (1)$$

4 EXPERIMENTS

An aircraft configuration matching flight number 611 (CI611) from China Airlines was adopted to construct our system. The aircraft was a Boeing 747-200 on a regularly scheduled flight from Chiang Kai Shek International Airport (now Taiwan Taoyuan International Airport) in Taiwan to Chek Lap Kok International Airport in the Hong Kong special administrative region. This aircraft broke into pieces in mid-air and crashed while in flight on 25 May 2002. All 225 people aboard the flight were killed in this accident.

Given the right to privacy of flight crews and passengers, the simulated data of 225 cases are used to demonstrate the visualization system. Every human body is divided into 14 sections for analysis. In addition to the sections of head and body, the left-hand side and right-hand side of the body are accounted for in six parts consisting of the arm, elbow, palm of the hand, thigh, shank and palm of the leg. In the association results from the Apriori analysis of this dataset, the most frequent itemset is the head and body. There are eleven cases with frequent injuries to the head and body in seats from 25D to 25G, 26E to 26G and 27D to 27G.

For the proposed model, or ACBR approach, the user interface for assigning features and weights to different sections of human body is shown in Figure 2. In addition to the sections of the head and body, this user interface provides the 12 options described above for retrieving similar cases from the relational databases. The two sections of the right arm and left arm are near to the head and body and are selected with lower weights than are the head and body. Two other sections, the right elbow and left elbow, are also assigned lower weights than the previous four parts, sections 1, 2, 3 and 6. These weights are determined based on the requirements for analysis and the knowledge of the investigators.

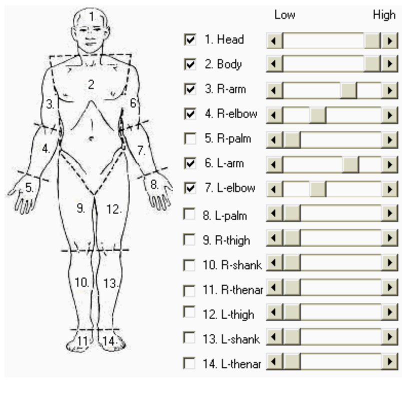

Figure 2: User interface for feature weights.

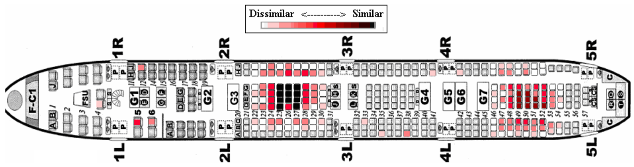

Figure 3: The visulization results of ACBR.

Figure 3 shows the visualization results of the the most frequent itemset used to measure similarities to all the other cases of the ACBR approach. There are 108 cases showing lower similarities to higher similarities to the case of the most frequent itemset involving head and body injuries. This visual system shows the relative similarity between a particular set of cases and all the other cases from the most similar to the least similar. It also reveals the group from seat rows 49 to 52 who had similar injuries, which provides further directions for investigators to explore the causes of injury.

5 CONCLUSIONS

An investigative result of a flight accident cannot reveal corresponding accidents if the relative information is inadequate. The ability to demonstrate visually the various data could be very helpful for investigators in understanding the relevant information. The present paper proposes an improvement of detailed analysis of accident data using the integration of the Apriori and CBR approaches for visually analyzing flight accidents. The frequent itemsets representing human injuries are identified using the Apriori approach, and the similar injury cases are confirmed based on the CBR approach.

The experimental results are encouraging when the association rules of the Apriori and CBR approaches are both used in the visualization system using the aircraft configuration and accident data from flight CI611. The frequent itemsets are quickly retrieved using the Apriori approach, and it provides objective itemsets, rather than human injuries that are subjectively identified by experts or investigators. We are encouraged by the results of the present study and are interested in investigating whether the use of different configurations (e.g., fuzzy logic or grey relational analysis) would result in further enhancements of this method in the future.

ACKNOWLEDGEMENTS

This research was supported by National Science Council, Taiwan, Republic of China, under the contract number NSC 98-2410-H-231-006-, NSC-97-3114-P-707-001-Y and NSC 98-3114-Y-707-001.

REFERENCES

Agrawal, R., Imielinski, T., and Swami, A., 1993. *Mining association between sets of items in massive databases*, In Proceedings of the ACM SIGMOD International Conference on Management of Data. pp. 207-216.

Cristian, A., Mitica, C., 2007. *Grid implementation of the Apriori algorithm*, Advances in Engineering Software. Vol. 38, pp. 295-300.

Chiu, C., Hsu, P. L., and Chiu, N. H., 2006. *Combining Apriori algorithm and constraint-based genetic algorithm for tree induction for aircraft electronic ballasts troubleshooting*, Lecture Notes in Computer Science. pp. 381-384.

Chiu, N. H., Huang, S. J., 2007. *The adjusted analogy-based software effort estimation based on similarity distances*, Journal of Systems and Software. Vol. 80, pp. 628-640.

Milosovski, G., Bil, C., and Kosevski, M., 2008. *Application of expert systems to aircraft accident investigation*, 46th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, Jan. pp. 7-10.

O'Hare, D., Chalmers, D., and Scuffham, P., 2003. *Case-control study of risk factors for fatal and non-fatal injury in crashes of civil aircraft*, Aviation, Space, and Environmental Medicine. vol. 74, pp. 1067-1072.

Paul, R. A., Challagulla, V. U. B., Bastani, F. B., and Yen, I. L., 1993. *A memory-based reasoning approach for assessing software quality*, Computer Software and Applications Conference.

Taneja, N., Wiegmann, D. A., 2003. *Analysis of injuries among pilots killed in fatal helicopter accidents*, Aviation, Space, and Environmental Medicine. vol. 74, pp. 337-341.

Weng, C. T., Ho, C. S., and Lan, C. E., 2003. *Aerodynamic model estimation and analysis for a jet transport in a landing accident*, AIAA Atmospheric Flight Mechanics Conference and Exhibit, Austin, Texas, August.