Experiences in Mining Aviation Safety Data

Zohreh Nazeri
MITRE Corporation
nazeri@mitre.org

Eric Bloedorn
MITRE Corporation
bloedorn@mitre.org

Paul Ostwald
MITRE Corporation
postwald@mitre.org

ABSTRACT

The goal of data analysis in aviation safety is simple: improve safety. However, the path to this goal is hard to identify. What data mining methods are most applicable to this task? What data are available and how should they be analyzed? How do we focus on the most interesting results? Our answers to these questions are based on a recent research project we completed. The encouraging news is that we found a number of aviation safety offices doing commendable work to collect and analyze safety-related data. But we also found a number of areas where data mining techniques could provide new tools that either perform analyses that were not considered before, or that can now be done more easily.

Currently, Aviation Safety offices collect and analyze the incident reports by a combination of manual and automated methods. Data analysis is done by safety officers who are well familiar with the domain, but not with data mining methods. Some Aviation Safety officers have tools to automate the database query and report generation process. However, the actual analysis is done by the officer with only fairly rudimentary tools to help extract the useful information from the data.

Our research project looked at the application of data mining techniques to aviation safety data to help Aviation Safety officers with their analysis task. This effort led to the creation of a tool called the "Aviation Safety Data Mining Workbench". This paper describes the research effort, the workbench, the experience with data mining of Aviation Safety data, and lessons learned.

1. BACKGROUND

Like other data analysis tasks, Aviation Safety data analysis is faced with the issue of finding analysis methods that are effective at finding interesting patterns in the collected data, but that can be performed with limited computational resources. In addition, such analysis methods must not require expertise in statistics or machine learning to run. Safety officers are experts in aviation and should not also need other expertise before obtaining useful results. To meet these needs we have developed a data mining workbench. Before discussing this tool in detail it is useful to understand how data is currently collected and analyzed. Currently, Aviation Safety offices collect and analyze the incident reports by a combination of manual and automated methods. Data collection consists of receiving hand-written reports and later entering them into a database. Data analysis is done by safety officers who are very familiar with the domain. Some Aviation Safety officers have tools to automate the database query and report generation process. However, the actual analysis is done by the officer with little help to extract the useful information from the data.

Figure 1 below shows the typical "life" of an incident report. After an incident occurs it is documented and submitted. At an airline it typically gets submitted to an air safety office (or equivalent organization). Most of these organizations have some form of an incident tracking system.

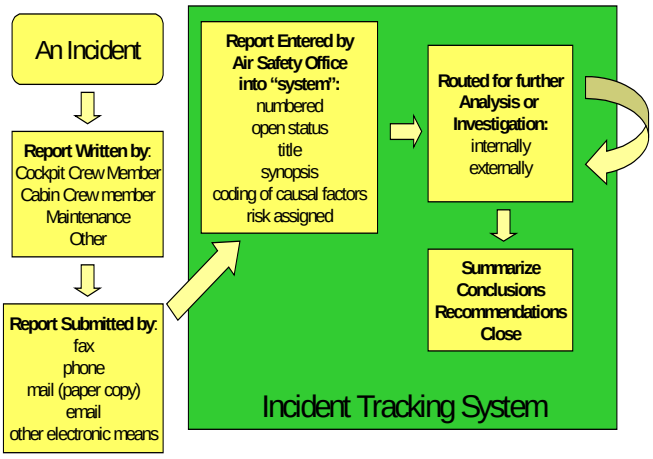

Figure 1. A typical incident report process.

An air safety office plays a key role in ensuring that an aviation organization operates in a safe manner. Incident data is used extensively to help fulfill this role. An important activity of a safety office is the internal analysis of the collection of incident reports.

The Aviation Safety Data Mining Workbench assists the Aviation Safety officer with the actual analysis of the data. Data mining finds interesting and useful information hidden in the data that might not be found by simply tracking and querying the data, or even by using more sophisticated query and reporting tools such as OLAP. It does this by searching a much larger space of patterns than is possible without automation. Of course, experienced Aviation Safety officers already know much of the important and

ACM SIGMOD 2001 May 21-24, Santa Barbara, California, USA

useful information in their data. Data mining can confirm their empirical observations and find new information as well. Data mining does not eliminate the need for safety officers to do some analysis but helps them more thoroughly analyze available data.

2. THE WORKBENCH

The "Aviation Safety Data Mining Workbench" is a collection of tools designed to help air safety officers more thoroughly analyze available safety data. It is particularly complementary to other tools often employed, such as incident tracking systems or capabilities to share or exchange data with others. The primary goal of the workbench is to find subtle patterns in the incident reports, which are indicators of safety risks before these risks lead to accidents. (Here we define an accident as an occurrence associated with the operation of an aircraft in which any person suffers death or serious injury or in which the aircraft receives substantial damage. An incident is an occurrence other than an accident associated with the operation of an aircraft, which affects or could affect the safety of operations.)

Because even simple tools like query capabilities and histograms can help focus the intensive data mining efforts, the workbench has a wide range of capabilities. These include the ability to load a set of incidents, select a subset for further analysis, produce some simple reports, and provide data for some more complex analysis. Additionally, three data mining techniques are currently incorporated into the tool.

The first data mining technique is called FindSimilar and is designed to search a collection of incidents to find those most similar to a selected incident. This is useful in determining if similar incidents have occurred before, and if so, how were they addressed. FindSimilar uses both structured and text fields to calculate the similarity between two records.

The second technique is called FindAssociations. This technique searches the collection of incidents to find subsets that have an interesting correlation. For example, this tool can identify a set of incidents that share a common location and incident type. Knowing that such an association exists between incident types, locations, and aircrafts may help in determining what action to take to reduce or eliminate those incidents in the future.

The third technique is called FindDistributions [5]. This technique focuses on a selected field or attribute of the incidents. It first determines an overall distribution for this field. Subsets of the data are then obtained and the distribution of the selected field is calculated for each subset. Those subsets that differ most from the overall distribution are identified as the most interesting. This technique helps in identifying anomalies that may be candidates for action.

All of these techniques are automated and help the air safety officer do certain analyses that are either difficult or resource intensive to do otherwise. Each technique is discussed in more detail below.

2.1 FindSimilar

FindSimilar searches a collection of incident records to find those most similar to a selected target incident record. The technique developed for FindSimilar is a hybrid approach motivated by a need we found when discussing data mining with airline safety officers. One task that they are repeatedly called on to perform is to find records of incidents that are similar to those that just recently occurred. If the new event is found to be similar to some past records, it may be part of a larger, more serious pattern. When this is the case, past actions taken to prevent this incident may have to be reviewed and updated. If, on the other hand, the incident is anomalous, it may be noted and closed, or simply announced to the relevant departments as a warning. Surprisingly this determination of record similarity was not well supported by the tools available to the safety officer. Officers could perform queries on both the structured and free-text fields, but these only respond with exact matches. Similarity of match between entire incident records was not supported.

To provide safety officers with a tool that found similar records from mixed free-text and structured data, we took a hybrid approach. In this hybrid approach, a match between two records is evaluated as the weighted match between each of the available fields within those records. When doing this match, we use methods that are appropriate for the type of attribute. More precisely this means:

$$Sim(record_i, record_j) = w_1*match(a_{1i},a_{1j}) + w_2*match(a_{2i},a_{2j}) + \ldots + w_n*match(a_{ni},a_{nj})$$

Our system supports three different types of matching: 1) strict Boolean, 2) ordinal, or 3) vector-based. In strict matching, which is appropriate for nominal typed attribute such as location, the match function takes the following form:

$Match(a_{ni},a_{nj}) = 1$ if $a_{1i} = a_{1j}$
else = 0

When the data are ordered, the system requires information from the user concerning the size and ordering of the domain. This matching is appropriate for any ordinal or interval type of data from numeric (e.g. Number_hours_flown) to string-based (Phase_of_flight). Given |Domain a|, the match function is:

$$Match(a_{ni},a_{nj}) = 1 - (a_{ni} - a_{nj}) / |Domain\ a_n|$$

When the data are textual, vector space matching is used. There are a number of different weighting schemes that could be supported, but by default the tool uses a tf-idf (term frequency inverse document frequency) method. In this approach, a vector with length equal to the size of the vocabulary is built for each field. The value at position x represents the ratio of the number of times that word appears in the document (term frequency or tf), and the number of times that word appears in the collection (document frequency or df). Geometrically speaking the overall document match is the distance in this large dimensional vector space between these two vectors, or the sum of the products over the square root of the sum of the squares.

$$\text{Match}(a_{ni},a_{nj}) = \sum_{x=1 to V} \frac{weight_{nix} * weight_{njx}}{\sqrt{(weight_{nix})^2 * (weight_{njx})^2}}$$

where:
V = size of vocabulary, $weight_{nix}$ = (weight of word x in field n of record i) and the default weighting method is tf.idf =*(term frequency$_{ix}$ / document frequency$_x$)*

The tool currently supports stemming, three different weighting schemes, the use of a stop word list, and the use of a thesaurus file for matching synonymous words. In stemming, words that are the same except for different endings (morphological variants e.g. engineered, and engineering) all map to the same base term (in this case engineer). Stop word lists are used to filter out words that are unlikely to add any additional meaning to the text. Examples of stop words are "and", and "the".

2.2 FindAssociations

FindAssociations finds co-occurrences of different data values. This tool uses only the structured fields in the database (narratives are not used). The technique used for the FindAssociations tool is the Apriori algorithm [2] which finds associations among data values. For example, one finding or association rule might be:

Birdstrikes → March 5%: 74%

meaning there is an association between the birdstrikes and the month of March. The values at the end of this association rule detail the strength of that association. The first value (5%) describes the support of this rule. Support is calculated as the number of times the two (or more) values on the left hand side must co-occur before they are considered for a candidate association rule. The second value (74%) describes the confidence of this rule. The confidence is calculated as the ratio of how often the entire rule is true over the frequency of the left-hand side of the rule. In this example, confidence is 74%, meaning the ratio of Birdstrike incidents in March over the total number of Birdstrike incidents is 74%. The user provides minimum thresholds on both confidence and support to guide the efficient search for associations.

Although only the associations that meet the minimum support and the confidence are generated by the algorithm, the number of rules returned by the tool is still very large. To further filter the resulting association rules, we applied the Lift measure [7]. Lift is a measure of the influence of occurrence of left-hand-side itemset on the likelihood that the right-hand-side will occur. For the association A => B, Lift measure is calculated as (Confidence of A => B) / (Frequency of B).

2.3 FindDistributions

FindDistributions [5] focuses on a user-specified attribute to find exceptions to the common distribution. This tool uses an attribute-focusing technique [3] and works on the structured fields in the data only.

For a particular attribute, A, of some structured flat-file database, the attribute-focusing algorithm calculates the overall distribution of A in the database. Then it compares this distribution with the distribution of A in various subsets of the database. If a subset has a statistically different distribution of A, then the condition that defines the subset is marked as interesting. Note that the overall distribution is our baseline rule, and the distributions for the subsets are the potential exceptions.

3. DATA MINING OF AVIATION SAFETY DATA

We applied the above techniques to three different sets of Aviation Safety data. Our goal was to use these reports to characterize those situations in which incidents occur. One set of data was a version of the Aviation Safety Reporting System (ASRS) database. This database is a source of anonymous, voluntary incident reports related to flight safety. The data contained structured fields (fields with pre-determined values) and free text fields (variable length narratives).

The ASRS data is extremely de-identified and general. As a result, our tools found general association rules such as:

> 50% of incidents in the "CLIMB" phase involved "MED_LARGE_TRANSPORT" air carriers.

and distributions such as:

> focusing on the MONTH attribute, number of incidents in the "TAXIING" phase in winter months was more than the expected given the distribution.

The FindSimilar tool found incidents that were similar in their structured field values as well as the words used in their narratives. Here is an example three different matches found for the same target record [4].

Target record synopsis field:

> PVT PLT RUNS INTO WX ON HIS XCOUNTRY TO HOUSTON. GETS LOST, USES A MAYDAY CALL FOR AN EMER DIVERSION TO DWH, NEAR HOUSTON. PLT HAD ALLOWED A 3 HR ELAPSED TIME PERIOD FROM WX BRIEFING TO TKOF.

The first match was found using both structured and text fields for the above target record, and here is the synopsis field for the matched record:

> A: C150 PLT FAILS TO CHK WX, GETS TRAPPED IN IMC ENRTE AND HAS TO CALL ATC FOR HELP. HE WAS NOT IFR CERTIFIED. ATC VECTORED HIM TO A VFR ALTERNATE WHERE HE WAS CLRD FOR A VISUAL APCH.

This is a good match; both records describe problems with bad weather, a call for help and an emergency diversion. Below is the synopsis field for the match found when only the text fields were used:

> B: PVT PLT IN A C172 GETS LOST AFTER RUNNING

INTO BAD WX AT NIGHT. NO FLT PLAN HAD BEEN FILED. PLT LOW TIME, NOT IFR RATED. WAS ON HIS WAY HOME TO ZZV FROM FL. HE HAD FAILED TO UPDATE ANY ENRTE WX.

This is also a good match. Both records (the target and the match) describe getting lost after running into bad weather. There is an important difference between the two matches found. The match on the text-and-structured fields (A) found a record of an incident that occurred in the daylight (like the target record) while the match on text-only fields (B) found an incident that occurred at night. The fixed field "lighting" captured this information. This example shows how information from both text and fixed fields can be exploited to provide a better overall match between incident reports than using just text or just fixed fields alone.

The third match was found using the structured fields only. The incident in this record involved runway transgressions/ unauthorized landings, not weather problems (like the target incident). This is not a good match. This matched incident doesn't even have a narrative so it is difficult to know more about what happened. The match between the target and this record is based on matches on the fixed attributes "acft_type", "engine_type", "flt_condition", "lighting", "num_engines", and "quarter_day".

Other sets of data we used were incident reports from safety offices of two European airlines, one of medium size and the other of small size. Both sets of data contained structured fields (fields with pre-determined values) and free text fields (variable length narratives). We had data from 1991 to 2000 for the medium sized airline, and data from 1997 to 2000 for the smaller airline.

The experience with the airlines' data yielded more interesting results. The additional data allowed us to find patterns that were not visible with the ASRS data. Unfortunately the details of these patterns cannot be revealed due to privacy concerns, but the flavor of these results can be revealed. For example we could find anomalies in departmental investigations, and associations between plane types and specific locations or incident types.

While the safety officers had an explanation for most of our findings, there were findings that were "interesting" or surprising. As it might be expected, there were less surprising findings for the safety officer of the smaller airline. This was due to the fact that the officer was well experienced in his field, and (more importantly) the amount of data was small enough for the officer to "remember" details of many previous incidents. Because of the small number of incidents and his wealth of experience, this officer could compare and analyze the incidents data in some depth without the help of any tools. In the case of both airlines, however, the "interesting" findings helped to draw the safety officers' attention to issues that were not addressed before.

In the case of the medium sized airline, our workbench was more helpful. It not only found interesting similarities, associations, and distributions that helped the officers to analyze their data, but also was used by them to help with their periodically generated reports.

4. LESSONS LEARNED AND FUTURE DIRECTIONS

The following are based on observation and lessons learned from applying our data mining workbench to the Aviation Safety data.

- Data that is overly general will not provide many interesting results. The ASRS data has been extensively de-identified to protect pilots, but this de-identification results in the loss of the details needed to find subtle patterns. By putting a large number of different aircraft types into the "MED_LARGE_TRANSPORT" class, for example, we miss peculiarities in the operation of specific plane types. Subtle patterns can get lost in generalities.

- Our experience with the Aviation Safety data indicated that this domain could certainly benefit from data mining. One characteristic of the data in this domain is that it is in both formats, structured fields and free text narratives, and neither field should be ignored. The information found in the structured fields is not recorded in the text fields, and the text fields discuss the important aspects of the incident not captured by the structured fields. For the similarity analysis, we could exploit all of the available data since the FindSimilar tool in our workbench uses both structured and the text fields.

- In the case of finding associations and distribution patterns in the data, we experienced the problem of having many results returned by the tools. This leaves the analyst with the task of going through a large amount of findings and separating the ones that are relevant, unusual, surprising, or in any other way "interesting". While statistical measures like lift are useful here, they cannot eliminate those patterns that are expected by the knowledgeable safety officer. More work is needed to eliminate or at least reduce the "uninteresting" results by incorporating available domain knowledge data mining (e.g. Liu et al, 1997; Adomavicius and Tuzhilin, 1997).

- Linking the incident reports to other sources of safety-related data, such as aircraft maintenance and weather data, could help finding better causal relationships. Ideally, a data warehouse of all related aviation safety data sources would be very beneficial. While we did not perform any experiment in this area, we did put together some sample databases filled with test data to show one of the safety officers how they could benefit from having an integrated data environment.

- Another condition that we noticed to be of great importance is the ease of using the tools. Safety officers (or other users) do not want to generate data samples for every new analysis tool. They want to run data mining tools, perform the data analysis, generate reports, run OLAP, and so on, all from one source. Data mining methods need to be incorporated into the safety process so that they aren't even immediately seen as data mining. The difficulty here is to resist the urge to put in more parameters, but to think about reasonable defaults.

- Although we developed and applied our workbench tools around the Aviation Safety data, the tools can be applied to data (structured and text) from other domains as well.

5. ACKNOWLEDGMENTS

Our thanks to Tom Curran of AerLingus for his expert feedback on the design and usefulness of the workbench. Also, thanks to Tom McLaughlin of MITRE Corporation for his help with implementation of the workbench.

6. REFERENCES

[1] Adomavicius, G., and Tuzhilin, A. Discovery of Actionable Patterns in Databases: The Action Hierarchy Approach, Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (1997), pages 111-114.

[2] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A.I. Fast Discovery of Association Rules. Advances in Knowledge Discovery and Data Mining (1996), AAAI Press/The MIT Press, pages 307-328.

[3] Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., and Ramanujam, K. AdvancedScout: Data Mining and Knowledge Discovery in NBA Data (1996). Data Mining and Knowledge Discovery, Volume 1, #1, Kluwer Academic Publishers, pages 121-24.

[4] Bloedorn, E. Mining Aviation Safety Data: A Hybrid Approach (March 2000), Third Federal Data Mining Symposium, Washington, DC.

[5] Harris, E. Jr., Bloedorn, E, Rothleder, Neal J. Recent Experiences with Data Mining in Aviation Safety (1998), SIGMOD98 DMKD Workshop, Session 5.

[6] Liu, B., Hsu, W., and Chen, S. Using General Impressions to Analyze Discovered Classification Rules (1997), Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, pages 31-36.

[7] Two Crows Corporation. Introduction to Data Mining and Knowledge Discovery (1998), Second Edition, ISBN 892095-00-0.