Data mining in military systems

A. A. Motta[1], A. S. Alves[1] & N. F. F. Ebecken[2]
[1]Brazilian Navy Research Institute, Rio de Janeiro, Brazil
[2]Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

The discovery of knowledge in databases has gained great importance and interest in recent years, especially in strategical areas where the access to the information is restricted many times.

There has been a substantial growth in the amount of stored information in digital media, mainly due to its increasing easiness and cost reduction. However, the large amount of stored data becomes impracticable for specialists to analyze through conventional methods. The difficulty for a more accurate analysis of these data means they may become accumulated useless information. On the other hand, it is known that in large amounts of data an enormous potential of extractable information can exist.

This work presents the application of a data mining tool in military databases. A case study is shown where the mining of rules of association and classification capable of finding useful information for the improvement of an anti-missile system in ships was carried out.

The results show that the methodology reached its main objectives, that is, rules were generated from large databases. However, it is important to remember that the generation of these rules represents only the initial stage of the system improvement with the use of data mining.

Keywords: data mining, association rules, classification rules, anti-missile systems, chaff.

1   Introduction

The finding of new knowledge from a mass of operational data is only possible if it is analyzed in a joint form. The large amount of stored data can make it impracticable for specialists to analyze them through conventional methods.

Moreover, difficulty for a more accurate analysis means that this data is transformed into an accumulation of information without utility, even though it is known that there might be a great potential of information in it.

The majority of data mining tasks, in special classification and association rules mining, suffers from strong restrictions in its implementation due the diverse factors, like, among others, the iterative character of the mining algorithms, allied to the often very large amount of data available, that limit the hardware platforms where these can be processed.

Although the generality of the majority of the algorithms developed for mining tasks of classification and association rules, it is evident that the more robust and efficient implementations are those where it is taken into account not only the available hardware, the knowledge of the data and its nature [1]. Therefore, it is believed that the biggest motivation for implementations of data mining tools for specific applications inhabits in the fact of getting greater domain of the system, thus facilitating future adaptations and additions of functionalities, as they show necessary.

Data mining systems can be used in military applications as described by Cantoni et al [2]. This technique can be especially important for electronic countermeasures (ECM) development, where the knowledge of a certain threat parameters, for which the countermeasure will be developed, is sometimes very limited. Then, the use of techniques and systems of data mining can provide important information that can contribute for the development effectiveness.

Amongst the ECM used for ships to defend themselves from fire-and-forget anti-ships missiles, it is the employment of *chaff* rockets, as described by Vakin et al [3]. *Chaff* rockets are rockets loaded with of metallized filaments that, once in suspension in the atmosphere, form a radar-reflective cloud that provide a target with the intention to confuse or seduce the missile, as illustrated in figure 1 [4].
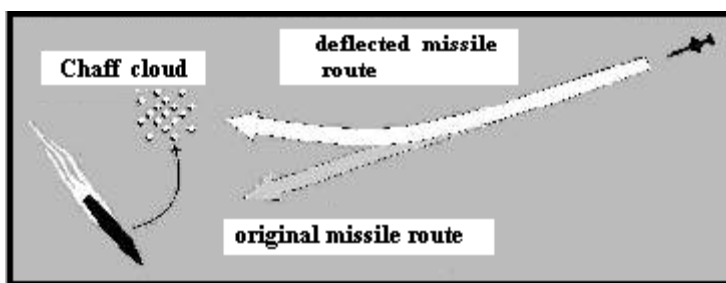

Figure 1:   Shunting line of route of the missile for cloud action of *chaff*.

Chaff clouds must be formed in key position so that incoming missiles are deviated. For this to be accomplished, a *chaff* rockets launching system must exist that determines, through the analysis of diverse parameters, the position, the load opening time, and the amount of rockets necessary to deviate the missile. The biggest difficulty in this type of countermeasure system development is the lack of necessary specific threat information, normally kept

secret. Thus, the mining of data of tactical scenarios involving adverse conditions and different threats behaviors, meteorological conditions and ships parameters can lead to the discovery of important information, hopefully useful to improve the ECM systems.

The computational data analysis possibilities, combined to military applications restrictions, has been the main motivation for the use of data mining tools to search new knowledge for the anti-missile systems improvement. The application of data mining tasks can be extended to other military systems as well, for, where there is data collected, certainly exists useful implicit information.

The remaining of this work is organized of the following form: section 2 presents the anti-missile system description, with its parameters and tactical operation modes; section 3 presents a brief conceptualization of classification and association rules; section 4 describes as the mining of data is applied to the systems databases and how it can contribute to their improvement; section 5 shows some experiments and results; and, finally, section 6 presents the final considerations.

2   *Chaff* ECM systems

The mission of an anti-missile ECM system is to defend a ship against missile attacks. In the case of radar-guided missiles, one possible countermeasure is the coordinated launching of *chaff* rockets. There are several other types of ECM available, but this work will be restricted to *chaff* employment.

2.1 Tactical employment modes

To defend a surface ship against a radar-guided anti-ship missile, a *chaff* countermeasure system can, according to a tactical software or a pre-generated table, produce fire solutions that will depend on the ship behavior, meteorological data, rocket launchers to be assigned, number and type of rockets to be launched from each launcher, and *chaff* load release time, among other factors, so that the seduction, distraction, or confusion defensive modes are applied to deviate the incoming missile.

2.1.1 Seduction
Seduction is the tactical mode in which action is taken in the moment that the missile acquires the target, in the accompaniment phase, also called final guidance phase. It can also be used against fire control radars to break their accompaniment. There are two types of seduction usually applied in these ECM systems: centroid displacement seduction and accompaniment window displacement (DUMP) seduction [5].

In centroid displacement seduction, *chaff* cloud(s) is(are), with bigger radar cross-section than the ship, formed at close distance(s) from the ship, to form a single initial radar echo that includes the ship and then, relatively, moves away from it, seducing the missile guidance radar. This mode often demands ship to maneuver to dislocate the centroid.

In DUMP seduction, a coordinated action is carried by radar tricking equipment and *chaff* rockets. Initially, a calculated radar signal is emitted to dislocate the displacement window, the direction, or both, of the missile guidance radar, producing a fake position, whilst *chaff* rockets are launched, with the lesser reaction time, in this fake position area to create a cloud that will be assumed as the target to be pursued by the missile. For this to work, a precise choice and programming of the *chaff* rockets is necessary, combined with an efficient evasive maneuver to move the ship as far as possible from that area.

2.1.2 Distraction

This tactical way is efficient only during the search phase of the missile guidance radar. It consists of launching a series of *chaff* rockets in accordance with optimized patterns for each threat type, presetted in the system, that will offer to the missile guidance radar a great number of fake targets available, with a high probability for it to acquire one of these fake targets instead of the ship, initiating its final attack phase against it.

2.1.3 Confusion

Tactical mode characterized by acting before a missile is launched. Its objective is to generate fake targets, confusing the enemy surveillance system and thus making it difficult to assign a target. The *chaff* cloud role in this mode demands their placement at the distances up to 10 km from the unit to be protected.

2.2 System requirements

In order to be possible to ECM systems to perform one or more countermeasure modes, they must comply with certain requirements. The same is observed to the ammunition, the *chaff* rockets. These systems present some intrinsic characteristics that determine and limit their applicability. Amongst others, the main characteristics are: the number of rocket launchers and their relative alignment with respect to the ship; the number of simultaneously launchable rockets per launcher; the rise angle for fixed launchers or, in the case of mobile launchers, the boundary-values of training and rise; the amount of simultaneous launchings; the reaction time between the threat detection and the *chaff* rockets launching; the time interval between launchings from the same launcher; and the ammunition characteristics (since the amount of *chaff* load, cloud opening time, the rocket range, among others characteristics, have a significant role in the system performance).

3 Data mining tasks

Amongst the diverse research areas in the data mining, one of most important and that has received much dedication is the mining of frequent items (or Frequent Items Mining, FIM) [6]. FIM is the search of subgroups that appear, with certain frequency, in a data mass, and it appears as main subproblem in other data mining tasks of great relevance, such as association and classification rules mining, stated that rules are the most expressive and intelligible forms of human knowledge [7–9].

3.1 Association rules

According to Liu et al [10], "association rule mining finds all the rules existing in the database that satisfy some minimum support and minimum confidence constraints". They represent standards where the occurrence of events in a set is high [6]. It can be said that it is the probability of a set of items to appear in a transaction, given that another set is present. An association rule is an implication of form R: X⇒Y, where X, the antecedent, and Y, the consequence, are subsets of disjoint items.

   The support of a rule R in a database D is defined as sup(X ∪ Y). Thus, such rule can be understood as a prediction, since if a transaction T supports X, it will also support Y, with certain probability, which is called rule reliability and denoted as conf (R):

$$\operatorname{conf}(R) = P\left(Y \subseteq T \mid X \subseteq T\right) = \frac{P\left(Y \subseteq T \wedge X \subseteq T\right)}{P\left(X \subseteq T\right)} = \frac{\sup\left(X \cup Y\right)}{\sup\left(X\right)} \qquad (1)$$

   The confidence of a rule discloses how much this rule is applicable, whereas its support indicates how much this rule is trustworthy. Therefore, for a rule to be relevant, it is necessary to have sufficient support and confidence, and it is said that a rule is relevant or strong if its support and confidence are above some predetermined minimum values [7].

3.2 Classification rules

According to Liu et al [10]," classification rule mining aims to discover a small set of rules in the database that forms an accurate classifier". It is the process of class prediction for records present in a database. Considering a standard as a set of attribute values, a classification rule is the implication of a standard with a class. It must be capable of making a record prediction depending on if its antecedent is a subset of this record. A rules-based classifier is composed of a set of rules set ordered by some anteriority norm and a standard class (its default).

   The problem of classification involves the discovery of relationships (or dependences) between class variables other variables, using such relationships to classify unknown cases. These relationships are stored as classification models, in form of rules [10, 11].

4 Data mining employment in ECM

ECM systems possess specialized software responsible for the execution of its tactical modes. This tactical software receives, from the others systems and sensors in the ship, the necessary parameters for the scenario creation, or threat instancing. From the interpretation of this scenario, the tactical mode designated, ammunition availability, and the status of the system units, the software calculates a solution that consists, basically, of the selection of the number of

chaff rockets to be launched from each launcher and the opening time of the loads (or fuze temporization for each rocket).

There might have cases in which, due to reaction time requirements, lack of available ammunition in one or more launchers, or even other diverse failures, it is not possible to calculate or to execute a solution, and, for these cases, these systems are endowed with tables (*look-up tables*) composed of diverse tactical scenarios with their respective standard solutions, representing an alternate way to find a possible defensive solution. Computer simulation is intensively used to validate to tactical software and to elaborate the *look-up tables*.

4.1 Data simulation

Simulation is a fast and non-expensive method to test a system behavior. However, it suffers from an important limitation: the validity of the employed models. It is extremely difficult to model any complex system, its difficulty growing with its complexity, and it must be established a commitment between accuracy and feasibility.

The simulator that generates the data to be used by the countermeasure system simulates the attack of a missile against a ship, in which the attacking missile radar tracks the target, while the ship searches to defend herself from the incoming threat through the use of *chaff* rockets. For that, diverse factors are considered, including surrounding meteorological data; missile-target geometry; attacking missile characteristics (e.g., physical, aerodynamical, sensors, and control system); own ship characteristics (e.g., physical, electromagnetic, hydrodynamic, propulsion and maneuver systems); countermeasure system characteristics (e.g., launchers type and positioning in the ship, launching control system); rockets characteristics (e.g., physical, aerodynamics, propulsion); and rockets payload characteristics (e.g., physical, electromagnetic, aerodynamics).

The modeling of the entities that composes the simulation environment is the great challenge in the simulation process, because, to ensure useful results, it is necessary to well model these entities (e.g., own ship, threat, *chaff* cloud, etc). For this, some usually unavailable data are necessary, such as the ship radar cross-section for all possible angles and the exact characteristics of the threat search radar, among others.

4.2 Data mining contributions

Considering the limitations of simulation and the amount of involved variables in the problem, it is easy to conclude that any analysis for simple data inspection hardly leads to some type of contribution, what motivates the application of data mining tasks in this problem.

Diverse data mining data can be applied to simulation results and contribute, somehow, for the countermeasures systems improvement. In this work, association and classification rules are applied to improve the survivability chances of a ship equipped with a *chaff* system under attack by an anti-ship missile.

4.2.1 Association rules mining

Association rules constitute a tool of non-supervised computational learning. It is based on observations and discoveries, searching for occurrences of standards, trends and generalizations in the data mass without user intervention and the application of this task can lead to surprising discoveries.

Its most immediate application in the current problem (launching of *chaff* rockets to defense a ship under an anti-ship missile attack) is the use of the mined information as an evaluation instrument for the simulation outcome. They can disclose, for instance, that a solution or even a simulation behavior associated to events characterize a wrong trend of a model, showing that the model must be reviewed and corrected or improved and that certain assumptions might not be applicable, what can improve the ECM system effectiveness, since the data are used to produce a defensive solution or an entry in the *look-up table*.

Association rules can also reveal occult characteristics in the problem (such as, for instance, unnoticed details regarding the missile search pattern) that can contribute not only to improve the defense system, but lead to tactical or doctrine re-evaluation on *chaff* employment. In addition, based in the knowledge of the available data, some user interventions can be applied, like the use of some type of hierarchy between the records, and possibly leading to other useful discoveries.

Association rules can be used to create parameters to help to understand threats behavior, to identify affinities between solutions and tactical scenarios, to foresee solutions for new scenarios, to perform behavior analysis for fault detection, etc, demonstrating that their use in such databases can contribute for the ECM systems improvement, and that the application possibilities are innumerable.

4.2.2 Classification rules mining

Liu et al [10] showed that it is possible to integrate association and classification rules mining, so that it is possible, starting from association rules, to perform classification tasks implementing a few algorithm changes that will allow the employment of both mining rules without performance loss. It must be pointed, however, that classification is a supervised method, that is, it is learned with the use of examples, being applied for prediction.

In this work, classification of tactical scenarios was applied in classes of defense sectors. The area near the ship is divided in sectors in accordance with the launcher to be used in a scenario solution, so that, in ships equipped with four launchers, four classes were created. Neglecting the combined solutions (those where the use of more than one launcher for the defense of the ship is necessary), the result of the analysis of this process can be used, for instance, in the reloading politics procedure for situations where there is not plenty of ammunition available.

There are instances of the problem (scenarios) for which it is not possible to find a solution, and the grouping of these instances in two classes (with solution and without solution) can help in the *chaff* rockets application doctrines development. In addition, a real time scenario classification can allow the ship to

engage in other defense method. Other parameters can be used for the scenarios separation in classes, such as, for instance, the necessity or not of maneuvering by the ship. Even in cases for which missiles do not directly hit the target, nearby explosions often cause expense damage to the ship systems, especially in the detection and communications antennas.

Like the association mining rules, classification mining rules can collaborate for the ECM systems improvement and its use doctrine.

5   Case of study and *speed up*

One of the biggest challenges in the mining rules tasks use is the algorithm implementation. Due to their iterativity aspect, to the great number of generated combinations, and mainly the usually very large size of the databases, implementations must be constructed carefully, using optimized data structures so that computational costs do not make them unfeasible.

A tool has been developed in C programming language, capable of effective execution of the rules generation task for application in the databases generated for a *chaff* rockets employment environment simulator. Its initial task is a data pre-processing, carried through the databases produced by the simulator, in which this data is converted to a proper format, suitable for the mining tasks implementations, some parameters discretization, data cleanness, ordering, and attributes choice.

The produced databases did not present *a priori* problems for the serial mining rules tasks execution (in one machine). After verification that the rules were being correctly mined and to ensure that a reduced set of classification rules was extracted, as suggested by Liu et al [11], a parallel version was developed, to reduce to overall consumed time with direct impact in its feasibility.

The experiments were carried in a PC cluster with 16 Pentium IV HT 2.8GHz with 1GB of RAM computers, 8 Pentium IV 2.4GHz with 1GB of RAM computers, and 16 AMD Dual-Core Opteron with 2GB of RAM. In the executions, the node selection was intentionally set as automatic, so that there is no identification of the used processors in each run. The implementation includes the development of a parallel implementation of CBA algorithm [11], including associations-based classification – successfully strategy already in use in serial implementations. Figure 2 shows the *speed up* obtained with the use of up to 10 processors and databases varying from small to large sizes. It can be observed that, as the database size increases, the *speed up* approaches the theoretical value. Due to confidentiality issues, it is not possible to disclosure further details regarding the database.

It is important to point out that, in very dense databases, there is a risk of combinatory explosion in association or classification rules mining applications, making its usage impracticable quite often. However, the results showed that the methodology worked fine, being possible to extract rules from the experimental databases and that, from the results of the *speed up* study, there is an indication that its application in bigger databases is not only possible, but feasible.
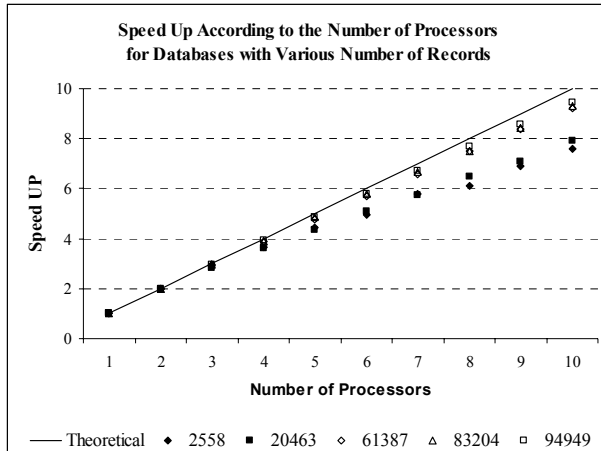
**Speed Up According to the Number of Processors
for Databases with Various Number of Records**

Figure 2:   *Speed up* according to the number of processors for databases with various numbers of records.

6   Final remarks

The association or classification rules mining from simulated databases represents the initial stage of this work. For the attainment of results that come to effectively contribute for the ECM systems improvement, it is necessary to give continuity to the work. It is necessary to establish more precisely the parameters of minimum support and minimum confidence, which directly influence the methodology effectiveness. Very high values of minimum support can prevent the mining of important rules or reduce too much the classifier precision. Thus, after *tuning* the rules generator tool, it will be necessary to perform an analysis of the disclosed information by the data mining tasks so that the new knowledge can be added to the application.

In addition, it is possible to include tasks execution improvements, such as data hierarquization or the inclusion of other data mining tasks, among others, extending, in this way, the possibilities of discovery of new knowledge in this area of application. Finally, this work introduces the application of data mining tasks with the purpose to improve anti-missile systems for ships. However, the methodology is not restricted to the present case study, rather non-military problems can also benefit from this methodology, and the methodology just needs to be applied to other individual problems.

References

[1]   Adamo J-M., Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms, Springer, New York, pp 66-74, USA 2000.

[2] Cantoni V., Lombardi L., Lombardi P., Challenges for Data Mining in Distributed Sensor Networks. In The 18th International Conference on Pattern Recognition. Proc. of the ICPR'06, pages 1000-1007, China, Hong Kong, August 2006.

[3] Vakin S. A., Shustov L. N. & Dunwell R. H., Fundamentals of Electronic Warfare. Artech House, INC. Norwood, MA, USA 2001.

[4] http://en.wikipedia.org/wiki/Chaff_%28radar_countermeasure%29, last accessed on January 30, 2007.

[5] http://www.sew-lexicon.com, last accessed on January 30, 2007.

[6] Agrawal R, Imielinski T. & Srikant R., Mining Association Rules between Sets of Items in Large Databases, Proc. of the ACM SIGMOD Intl. Conf. on Management of Data, pages 207–216, Washington, USA, 1993.

[7] Agrawal R., Mannila H., Srikant R., Toivonen H. & Verkamo A. I., Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, Fayyad U. and et al, 1995.

[8] Agrawal R. & Srikant R. Mining sequential patterns. In Proc. 11th Int. Conf. Data Engineering, ICDE, pages 3–14. IEEE Press, 1995.

[9] Beeferman D., BERGER A., & LAFFERTY J. Statistical models for text segmentation. Machine Learning, 34(1-3):177.210, 1999.

[10] Liu B., Hsu W., & Ma Y.. Integrating classification and association rule mining. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), pages 27–31, 1998.

[11] Liu B., Ma Y., & Wong C. Improving an association rule based classifier. In 4th European Conference on Principles and Practice of Knowledge Discovery in Databases PKDD, pages 504–509, 2000.

[12] Girish K., Palshikar A, Mandar S., Kale B & Manoj M., Association rules mining using heavy itemsets, In Tata Research Development and Design Centre, Engineering and Industrial Services, Tata Consultancy Services Ltd., Pune 411001, India, 2006.