K.K. Chand*
S.Pattnaik**

Data Mining as a Tool for Studying Projectile System Performance –A Case Study

Abstract

The design, development and maintenance of a projectile system generate a significant amount of information, data, knowledge and documentation for its end use. The real worth of these databases lies not only in easy data access, but also in the additional possibility of extracting the engineering knowledge implicitly contained in these data. As Data Mining (DM) or Knowledge Discovery in Databases (KDD) has been evolved into an important and active research area in theoretical as well as practical applications challenges for last few decades, therefore, the problem associated with the extracting and discovering previously unknown knowledge from large databases along with the analytical modeling and data driven techniques gain more and more importance in this context. Reflecting this trend, the paper discusses an overview on the fundamental issues of DM or KDD and then its application in projectile system performance evaluation as a case study from the armament industry.

Introduction

The Armament industry has adapted the information technology in its processes in terms of various modeling, simulation via computer aided design and drafting, manufacturing and testing systems, performance evaluation via statistical, fuzzy and neural techniques, reliability analysis etc for various decision-makings process during the lifetime of a projectile system performance development programme.

A projectile with its weapon and ammunition form a projectile-weapon or a weapon system or a projectile system, which is an integrated product of multidisciplinary and multi-technology complex system with multi-levels and multi-factors features. The growing complexity of projectile-weapon systems, as well as the rapidly

*Scientist, PXE, DRDO, Chandipur, Balasore, Orissa-756 025. **Reader, Dept. of I &CT, F.M. University, Vyasa Vihar, Balasore, Orissa-756 019.

There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information from the rapidly growing volumes of useful digital data into knowledge. These theories and tools are the subject of the emerging field of knowledge discovery in databases (KDD). It uses statistics, databases, AI, machine learning, pattern recognition, machine discovery, uncertainty modeling, data visualization, high performance computing, optimization, management information systems (MIS), and knowledge-based systems to discover and present knowledge in a form that is easily comprehensible.

his paper presents why, how and what data mining techniques can be applied on a projectile-weapon system performance evaluation and future research directions in the field. A brief overview on a case study of recent data mining in a projectile system performance evaluation application has been discussed.

Why Data Mining?

The traditional method of turning data into knowledge relies on manual analysis and interpretation.

In this method, data analysis such as spreadsheets and ad-hoc queries were not able to fit in because they can only create informative reports from data, but cannot analyze the contents of these reports. Hence, there is a significant need for a new generation of techniques and tools with the ability to automatically assist humans in analyzing a large amount of data to provide useful knowledge within the armament industry.

When the scale of data manipulation, exploration, and inference grows beyond human capacities, people look to computer technology to automate the documentation. The need to scale up human analysis capabilities to handling the large number of bytes that we can collect is both economic and scientific. The problem of knowledge extraction from large databases involves many steps, ranging from data manipulation and retrieval to fundamental mathematical and statistical inference, search, and reasoning. Data mining and knowledge discovery in databases have been attracting a significant amount of research, industry, and media attention of late.

Data Mining, is the science and technology of exploring data also popularly known as *Knowledge Discovery in Databases* (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful, valid information from data in databases. Data mining involves fitting models to or determining patterns from observed data. The fitted models play the role of inferred knowledge. Deciding whether or not the models reflect useful knowledge is a part of the overall interactive KDD process for which subjective human judgment is usually required. While data mining and knowledge discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process.

Data mining is an integration of multiple technologies. Each of these technologies: Statistics, Decision Support Systems, Artificial Intelligence, Pattern Recognition, Database Management and Warehousing, Machine Learning, Visualization, and Parallel Processing are all tools that interact and support a data mining tool to automatically extract concepts, interrelationships, and pattern of knowledge of interest from large databases. Statistics and Machine Learning continue to be developed for more sophisticated statistical techniques. The data miner is the critical interface between the syntactic knowledge or patterns generated by machines and the semantic knowledge required by humans for reasoning about the real world. The work is this paper motivated by several observations in the armament industry.

What is Knowledge Discovery and Process?

Definition

Knowledge discovery is an activity that produces knowledge by discovering it or deriving it from existing information. Indexing knowledge elements, filtering based on content and establishing linkages and relationship among the elements then organizes knowledge. Subsequently, this knowledge is made available to users for supporting their decision making process.

Knowledge discovery is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. Given a set of facts (data) *F*, a language *L*, and some measure of certainty *C*, we define a *pattern* as a statement *S* in *L* that describes relationships among a subset *FS* of *F* with a certainty *C*, such that *S* is simpler (in some sense) than the enumeration of all facts in *FS*. A pattern that is interesting (according to a user-imposed interest measure) and certain enough (again according to the user's criteria) is called *knowledge*. The output of a program that monitors the set of facts in a database and produces patterns in this sense is *discovered knowledge*.

These definitions about the language, the certainty, and the simplicity and interestingness measures are intentionally vague to cover a wide variety of approaches. Collectively, these terms capture our view of the fundamental characteristics of discovery in databases. According to Neil Fleming, as a basis for thought, the idea is that, information, knowledge, and wisdom are more than simply collections. Rather, the whole represents more than the sum of its parts and has a synergy of its own. So, the following associations can reasonably be made as shown in Figure3.

- **Data** relates to a set of facts F, a language L and some measure C;
- **Information** relates to description, definition, or perspective (what, who, when, where);
- **Knowledge** comprises strategy, practice, method, or approach (how);
- **Wisdom** embodies principle, insight, moral, or archetype (why);

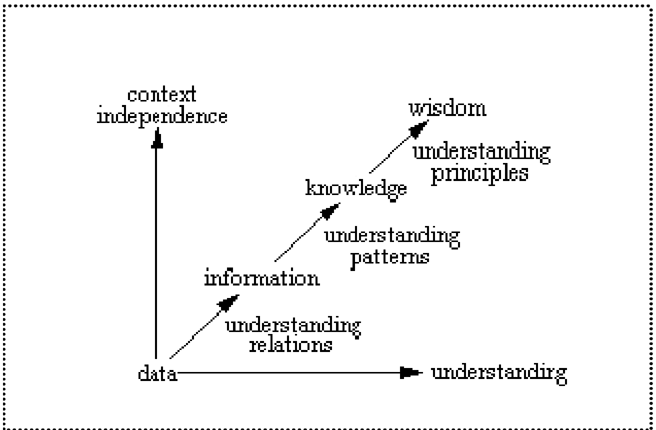

Figure-3: Knowledge Context

Knowledge Discovery refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. Data mining involves a collection of tools and techniques for finding useful patterns relating the fields of very large databases. The newest form of data mining is the linguistic summarization of data which aims at a computer generated verbal description of the knowledge implicit in a database often in the form of 'if –then' rules that resemble fuzzy knowledge granules. Text mining is to extract patterns from textual documents. A text mining technique typically involves text parsing and analysis to transform each unstructured document into an appropriate set of features and subsequently applies one or more data mining techniques for extracting patterns. The KDP process is outlined in Figure 4.

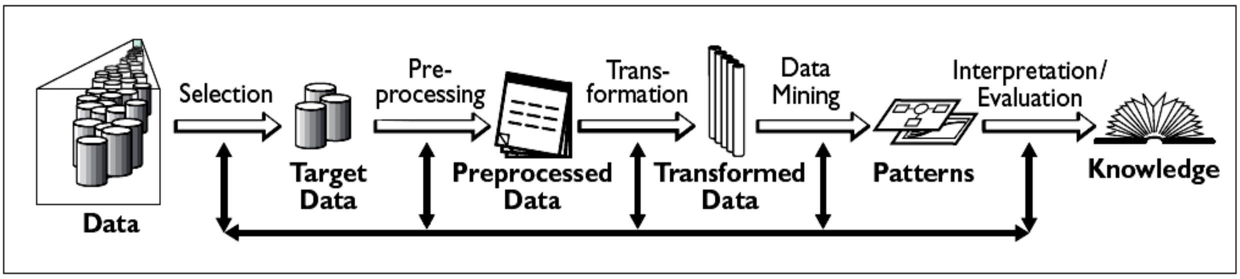

Figure 4. Knowledge Discovery Process

Data mining is part of a larger interactive and iterative process called knowledge discovery and involves the following steps:

(a) *Define the Problem:* This initial step involves understanding the problem and figuring out what the goals and expectations are of the applications, which includes relevant prior knowledge;

(b) *Target data set:* Selecting a data set or data samples on which discovery is to be performed;

(c) *Data Collecting, Cleaning, and Preparation:* This requires figuring out what data are needed, which data are most important and integrating the information. This step requires considerable effort, as much as 70% of the total data mining effort. It removes noise, deciding on strategies for handling missing data fields, accounting for time sequence information and known changes and mapping missing and unknown values;

(d) *Data Reduction and Projection:* Includes finding useful features to represent the data and using methods to reduce the effective number of variables under consideration;

(e) *Choosing the function(s) of data mining:* Selecting the data mining function based on data model such as summarization, classification, regression and clustering;

(f) *Choosing the data mining algorithm(s):* Includes selecting the method(s) to be used to search for patterns in the data such as statistical algorithms, visualization techniques, deviation trend analysis decision tree analysis etc. Two or more techniques can be combined depending upon the data models;

(g) *Data mining:* This model-building step involves selecting data mining tools, transforming the data if the tool requires it, generating samples for training and testing the model, and finally using the tools to build and select a model. This means it concerns with applying computational techniques to find patterns in data in a particular representational form or set of such representations. A Pattern that is interesting and certain enough can be treated as knowledge;

(h) *Validate the models*: Test the model to ensure that it is producing accurate and adequate results.

(i) *Interpretation:* Includes interpreting the discovered patterns and possibly returning to any of the previous steps as well as possible visualization of the extracted patterns, removing redundant or irrelevant patterns and translating the useful ones into terms understandable by the users;

(j) *Using the discovered Knowledge:* Includes incorporating discovered knowledge into the performance system, taking action based on the knowledge or simply documenting it for management/later use;

(k) *Monitor the model:* Monitoring a model is necessary as with passing time, it will be necessary to revalidate the model to ensure that it is still meeting requirements. A model that works today may not work tomorrow and it is therefore necessary to monitor the behavior of the model to ensure it is meeting performance standards;

Data Mining: Verification vs. Discovery

Data mining is not simply a query extraction verifying analysis tool. Decision support systems (DSS), executive information systems, and query writing tools are used mainly to produce reports about data. In this matter, these query tools serve mainly to access the records already existing in large databases. After the data are extracted, they are examined to detect the existence of patterns or other useful information that can be used in answering questions of interest. This kind of data extraction methodology that finds trends in existing data is called the verification method. Under this scheme the data miner must hypothesize the existence of information of interest, convert the hypothesis to a query, pose it to the warehouse, and interpret the returned results with respect to the decision being made. The organization using this methodology will continually reissue hypotheses using query tools that support or negate it. This is a very systematic approach in finding a potential solution to the problem posed. Little information is created using this retrieval approach.

The approach used by data mining is different. Instead, data mining uses a discovery methodology whereby based on the method used, this technology will try to detect trends and produce results about the data with very little guidance from the user. This technology is designed to find the most significant sources of data and draw conclusions from the data it selectively sifts through. In data mining, large amounts of data are inspected; facts are discovered and brought to attention of the person doing the mining. As such, data mining is a discovery tool where a more efficient mode of seeking relevant data and bringing forth this information to the data miner is accomplished.

Knowledge Discovery Techniques

The Knowledge discovery can be of two categories: descriptive knowledge discovery and predictive knowledge discovery. The former describes the data set in a concise and summary manner and presents general properties of the data; whereas the later constructs one or a set of models, performs inference on the available sets of data and attempts to predict the behavior of new data sets. The features of knowledge discovery includes:

- Large Amount of data;
- Efficiency;
- Accuracy;
- Automated Learning;
- High Level Language;
- Interesting Results;

A Case Study

Projectile Performance System

Generally, a projectile system performance evaluation includes reliability analysis, functional analysis, effectiveness analysis, cost analysis, performance analysis, etc. in consideration of its performance via whole life cycle from planning to a conceptual design, from a prototype to a real product, from deployment to retirement. Effectiveness performance is a measure of projectile system capacity for the fulfillment of a mission. Menace analysis sets mission (goal) for weapon system design. The result of function analysis is a specifically designed prototype, which will be under effectiveness analysis for acquirement of a system performance index with specific missions. Cost and reliability analyses aim at a feasible and successful product at right time and bearable expenditure. Tradeoff exists between effectiveness, cost and reliability in weapon system development, and multiple criteria decision analysis methods are helpful for comprehensive evaluation. Among all tasks in weapon system evaluation, effectiveness analysis is more important and worth more endeavors since it tells the weapon system developers if the concerned system can really fulfill a specific mission. There are many methods for effectiveness analysis, ADC (availability, dependability and capacity), SEA (system effectiveness analysis), index method, etc. while ADC method is widely used.

Due to complexities in calculation and model connection in the field of weapon system evaluation, application of tremendous achievements of computer technologies is a great appeal. With popularity of decision support system (DSS) and its various applications at different fields to develop a DSS for a projectile system evaluation is a natural idea for armament people.

System Performance

System performance is realized through designed-in system *capabilities* and *functions*. In this context, the term *capabilities* refers to the various desired performance attributes and measures of the system, such as maximum speed, range, altitude, or projectile's delivery accuracy. The term *functions* refer to the desired objective capabilities and objective scenarios that the system must be capable of executing in an operational environment. For example, a projectile for a given weapon system may have the capability to fly at Mach 1.0, but its ability to function at that speed in a real-world objective is dependent upon many other factors. Therefore, factors of reliability, durability, maintainability - overall sustainment - are inherent in achieving optimum system functionality.

Desired capabilities are determined by *priorities. Priorities* reflect the users requirement that drives the inevitable tradeoffs that the system design must undergo, balancing performance, availability, operations and supports. The level of operational effectiveness achievable is predicated upon the allocation of resources towards these priorities. Performance cannot be considered separate from the other elements of operational effectiveness – they are inextricably linked. The system capabilities and functions represent the desired mission capabilities as a total. The knowledge-discovery module for a weapon system is one component of the intelligent expert system for monitoring and controlling the performance evaluation system. The projectile performance system consists of six modules:

- the expert-system module;
- the scheduling, coordinating and planning module;
- the trajectory simulation module;
- the testing-management module;
- the trajectory computation module, and;
- the decision making module;

(a) The expert system consists of a knowledge base for identifying various stages of initial productions for a weapon system developmental programme.

(b) The scheduling, coordinating and planning module identifies the effective planning programme suggested by the expert system.

(c) The trajectory simulation module predicts the various expected trajectory trends using trajectory-modeling techniques via numerical techniques.

(d) The testing-management module plans for measurement of various parameters along with trajectory computation through dynamic testing.

(e) The trajectory computation module predicts various trajectory trends for a particular combination of weapon system for the pre-plan testing-management module.

(f) The decision-making module studies various measured parameters using statistical, fuzzy techniques, case-based reasoning for evaluation of system performance for verification and validation purposes as per requirements.

Looking at the performance affecting a projectile system and its analysis, six main categories of data and information can be differentiated as:

(a) *Parameters directly affecting the initial conditions or Fire Control:* gun position, gun height, muzzle velocity, azimuth and elevation or launch angle;

(b) *Parameters indirectly affecting the initial conditions:* propellant temperature and barrel abrasion or wear;

(c) *Parameters directly affecting the projectile:* in flight containing all atmospheric properties depending on height (temperature, air density, air pressure, wind, speed of sound) and all projectile depending coefficients (drag, lift, Magnus and Coriolis forces, spin and yaw);

(d) *Parameters indirectly affecting the projectile:* case thickness, case strength and explosive weight;

(e) *Parameters directly affecting the target:* impact angle, velocity, location, and angle-of-attack;

(f) *Parameters indirectly affecting the target:* thickness and strength;

From a measurement point of view, all these parameters are error afflicted and therefore, every parameter can be quantified by its mean and standard deviation. When we develop mathematical or simulations models we sometimes deal with systematic uncertainties, which can originate from biases in measurements or biases in solutions methods.

In general, for a particular projectile system performance model is shown in Figure 5.

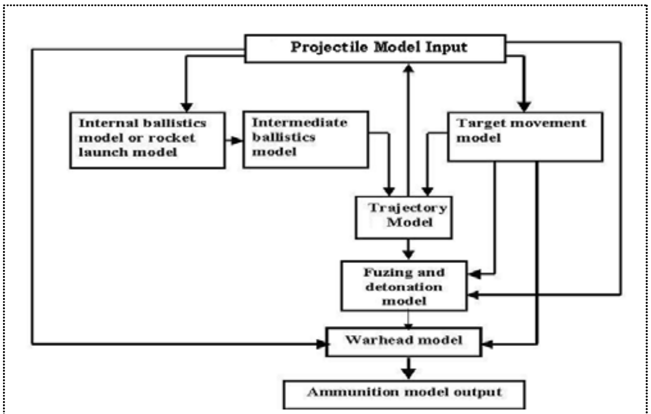

Figure 5: A Projectile Performance System Model

Conclusions

Data mining works by simply learning from data it can. Data mining is a technology that has emerged to provide organizations whether large or small the opportunity to discovery-hidden trends and patterns in their data. This realization has come about as a result of the increasing loads of data being stored in organization's databases. There is no dependence on the skills or intuition of a programmer to synthesize a model. Data Mining has evolved and continues to evolve from the intersection of research fields such as machine learning, pattern recognition, and database statistics, artificial intelligence etc. Due to the high potential payoffs of DM applications there is a need to incorporate its utilities in the armament industry areas in particular projectile system performance evaluation. It is the responsibility of ballisticians, scientists, researchers and practitioners in armament field to ensure that the potential contributions of DM or KDD are not overstated and that users understand the true nature of the contributions along with their limitations. For this a new generation of computational techniques and tools is required to support the extraction of useful knowledge from the rapidly growing volumes of data. These techniques and tools are the subject of the emerging field of DM or KDD and so on.

References

1. Kecman Vojislav, Ch-1 Learning and Soft Computing, (1st Edition), MIT Press, , 2001.

2. Pujari K Arun, Data Mining: Techniques-, (1st Edition) University Press India, , 2001,

3. Fayyad U., Piatetsky-Shapiro G., Smyth P., Uthurusamy R., Advances in Knowledge Discovery and Data Mining, Foundation Issues of Knowledge Discovery (1st Edition), Cambridge, MIT Press , 1996.

4. Han, J. and Kamber, M., Data Mining: Concepts and Techniques (2nd Edition), San Francisco: Morgan Kaufman,. 2003

5. Leondes Cornelius T., Knowledge-Based Systems Techniques and Application (Vol-2)-, San Diego, Academic Press, 2000. ISBN 0-12-443877-6.

6..Holsapple C.W. Hand Book on Knowledge Management (Vol:1& 2) Berlin:Springer, 2003.ISBN 81-8128-038-5.

7. Text of Ballistics and Gunnery-I and II, London, Her Majesty's Stationary Office, 1987.

8. Chand KK, Understanding Trajectory Modelling via Error & Uncertainty Analysis-, National Conference on Advances in Armament Technology, 21-22 November, 2008

9. Ballistics Tools For the 21st Century, 23rd International Symposium on Ballistics, 16-20 April 2007.

10. Driels M. R, Weaponeering: Conventional Weapon System Effectiveness-, (1st Edition), AIAA Series, Inc, Virginia, 2004.

11. Fayyad Usama et.al. *From Data Mining to Knowledge Discovery in Databases.* American Association for Artificial Intelligence, AI Magazine FALL 1996, Vol.3, , , pp.37-54. (http://www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/fayyad96.pdf)

12. Graves G.H, Higgins J.L., Applications of Simulation in Logistics Combat Developments Proceedings of the Winter Simulation Conference 2002, p-911-916. (http://www.informs-cs.org/wsc02papers/119.pdf)

13. Wright Preggy Knowledge Discovery In Databases: Tools and Techniques.ACM, 1998. (http://www.acm.org/crossroads/xrds5-2/kdd.html)

14. Bellinger Gene The Effective Organization-(http://www.systems-thinking.org/teo/teo.htm)

15. Bellinger Gene.The Knowledge Centered Organization- (http://www.systems-thinking.org/tkco/tkco.htm)

16. Joshi Karuna Pande , Analysis of Data Mining Algorithms.. (http://userpages.umbc.edu/~kjoshi1/data-mine/proj_rpt.htm).