

# Use of Data Mining on Satellite Data Bases for Knowledge Extraction

Lance Self

Air Force Research Laboratory  
Space Vehicles Directorate  
Kirtland AFB, NM 87117  
Lance.Self@vs.afrl.af.mil

## Abstract

Extracting knowledge from existing sources of information is a key development area to unlock previously unknown relationships between specific data points. Data mining is a technique that uses artificial intelligence methods to extract previously unknown relationships. It becomes a factor when large volumes of data, such as in the case of clusters of satellites, need to be analyzed. Data mining, and how it can be applied to satellite databases, is the subject of this paper.

## Introduction

More and more large businesses are turning to the use of data mining techniques. They are doing so for several reasons, among them the need to be responsive to changing customer needs, and business analysis. The same tools which businesses use can also be applied to satellite technology. Analyzing satellite telemetry data through data mining promises to garner much of the same advantages which are prevalent in the business community.

## Data Warehouse

Data warehouses are very large databases where massive amounts of records are stored. A data warehouse is defined as:

“an intelligent store of data that can manage and aggregate information from many sources, distribute it where needed, and activate business policies”[1]

Business records contain historical information about the business, customers, and transactions and are usually associated with business such as banks, large department stores,

credit cards, insurance, telecommunications, health care, and government agencies. Businesses such as these accumulate tens of thousands of records each day. These records are in the form of operational data such as point-of-sale, order entry, cataloging, and pricing.

Satellites also accumulate records, but in the form of telemetry data points. Several hundreds to thousands of data points are formatted in telemetry frames every time step. These frames, once telemetered to the ground station, are stored in databases to be analyzed. The data points form the basis of historical data which accumulates and can be mined for relationships and information.

## Data Mart

A smaller, focused form of a data warehouse is called a data mart. One definition of a data mart is:

“ a data store that is subsidiary to a data warehouse of integrated data”[2]

A data mart is made up of operational data that is designed to serve a particular or specialized group of users. The data mart is designed to meet the demands of those users, where those demands range from analysis to summarized presentations. Each data mart is organized by subject. There may be a data mart for finance, another for inventory control, and yet another for sales. A data mart may be specific to order entry, for example, where the users of that order entry data mart would primarily be field sales representatives.

## Data Mining

Data mining is the act of extracting information or forming knowledge from raw data points held within data warehouses or data marts. In business, previously undiscovered customer patterns, or buying habit information related to customers, are typical outputs of

data mining techniques. Data mining can be used to predict behavior based on historical patterns and one way it does this is by identifying new relationships. When these factors appear data mining becomes an essential decision making tool for the organization.

Historically, statistics were employed to do data mining techniques. Statisticians used various mathematical techniques such as regression analysis to create models which would predict customer behavior. Data mining is meant to be used by the end-user, however, and not statisticians. Data mining is made available to the user through a series of GUI's which guide the user.

At first, these may seem similar to standard structured query language (SQL) queries. SQL queries require the user to know what queries to send. Data mining GUI's remove that requirement by guiding the user toward a series of algorithms. Along with these algorithms, various AI techniques are also available to data mining, among them fuzzy logic, neural networks, genetic algorithms, and the nearest neighbor technique.



Figure 1 Business use of Data Warehouse

### Mining Satellite Data

In a standard business use, as Figure 1 shows, raw data from various sources are input to the data warehouse. That data is cleansed, normalized, reformatted, and distributed to various data marts for use. As stated earlier, data warehouses are massive storage facilities. With satellite data, these massive facilities, and their costs, are not needed. Better suited for satellite mining is a scaled down version of the architecture, keeping the quality, redundancy, and openness of the design but eliminating the need for high end, high cost

hardware and software. With such a scaled down version the return on investment would yield a higher return.

Mimicking the business design, but altering it to suit satellite telemetry data yields one possible design in figure 2.

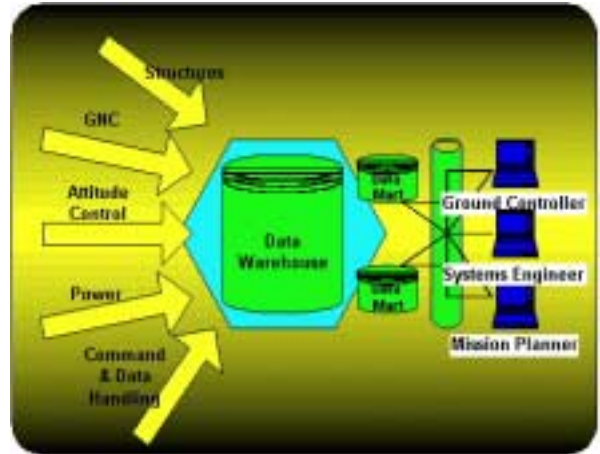


Figure 2 Satellite Use of Data Warehouse

In this design, instead of input being from sales regions or inventory, it is replaced by the telemetry stream input. Parameters specific to GNC would be identified, as would parameters from the other major satellite control functions such as attitude control, power, etc.. This would not mean an adjustment in the design of the telemetry stream is needed however. The parameters will be stored in a database. The identification and categorizing of the parameters takes place within the data warehouse where the data is prepared (normalized and reformatted).

Once in data marts the data is available for use organized by subject matter. GNC is available for the navigation engineers, health & status parameters by the ground controllers. Once in the proper format AI techniques can be applied to garner information and knowledge from these categories.

Artificial intelligence techniques such as genetic algorithms can be the same standard techniques already used in the business world. These techniques are meant to draw information for use by decision makers. As a group they are normally referred to as Decision Support Systems (DSS) or Executive Information Systems (EIS) and are designed to give the decision maker insight into the raw data relationships. These are the same goals of satellite decision makers. They also have a need to discover

new information, but related to the satellite bus instead of a company's profit/loss margin.

An important factor in any design is scalability or the ability of the design to grow as the need arises. Since this design mimics business design it is scalable. As more parameters are ingested into the system they can be diverted to their own data mart and become available to a new group of users. For example, if a group of users, such as those concerned with orbit determination, needs to investigate certain areas then related raw data could be stored into a data mart. The raw data points need not be exclusive, that is a subset of data points located in other data marts may also be sent to the orbit data mart.

Another advantage of this design is modularity. Data marts are islands of information. These islands are re-configurable, and may be joined with other tables or data marts if the need arises. Joining with other data marts does not destroy the original data but instead spawns a new data mart.

Data marts can start off small and grow as more experience and confidence in their usefulness is gained. In addition to data marts for satellite operational functions there can be data marts for various payloads. This is a natural progression since much of the groundwork has already been laid.

### **Nearest Neighbor Technique**

Berson and Smith [2] discuss five data mining techniques in their book. No one technique is considered "best", in fact it's assumed that a combination of one or more would be used for any given application. All five are applicable to satellite telemetry data analysis. One of the techniques, nearest neighbor, can be immediately applied to attempt to predict satellite sub-system behavior based on historical data points.

Given that telemetry data points are organized by sub-system such as GNC, ACS, power, and structures, these sub-systems become the data mart organizations. Once in this format, the sub-systems, and sub sub-systems can be mined using nearest neighbor technique to predict their behavior.

An example of a small sub-system within the power system are the batteries. Historical behavior of batteries, their charge rate, ability to hold charge, temperature, etc. are all entries within the data mart. A predictor of weather or not a battery will exceed temperature limitations may be a combination of those

entries. Each entry is given a weight factor based on historical knowledge of that entry affecting, in this case, battery temperature. If, for example, historical data reveals that batteries that have been overcharged 3% - 5% fail their life expectancy by 30%, the overcharge data entry is given a weighted value accordingly. How near the battery in question is to the cluster center of mass of the historical batteries measured gives an indication of how that battery will behave given it too is overcharged.

One of the challenges of using the nearest neighbor technique is grouping of the data. There may be many other factors which are affecting a failing battery which are unaccounted for. There may be environmental factors or factors from other sub-systems which have a direct affect on the data within the power sub-system. These relationships, intra-relationships between data points within a sub-system, and inter-relationships between satellite sub-systems, are easier to exploit when the data is organized using data marts.

### **Related Work**

Under an Air Force Research Laboratory SBIR Phase II contract [3], AbTech Corp. has researched the use of data mining techniques for use on satellite telemetry data. The Phase II resulted in a product called ModelQuest Satellite Telemetry Anomaly Resolution (MQ STAR) which analyzes telemetry data using a mixture of a statistical modeling approach and case based reasoning. MQ STAR operated on historical data sets to detect unanticipated anomalies as well as perform trend analysis to predict failures.

Another company which completed a SBIR contract with the Air Force Research Laboratory is Charles Rivers Analytics [4]. Their expert system, Mining Agent for Spacecraft Telemetry (MAST), generates rules from satellite telemetry data for knowledge discovery. A learning software agent takes in telemetry data along with learning parameter rules. These rules are passed to an interface agent which presents these rules (candidates) to the user. Validated rules are then implemented by an inference engine. Events are stored in an event database, and validated rules are stored into a rules database.

Each company has taken a different approach to solving the problem of detecting events, anomalies, and failures with large volumes of data before they become a major problem.

## Conclusion

More and more large businesses are turning to the use of data warehouses and data mining to gain insight into how to make their business more profitable and responsive. The techniques described are needed for businesses to increase profit margins, reduce customer turnover, and take advantage of new or unknown opportunities. Similar methods can be used by satellite engineers to explore and discover previously unknown information and knowledge. Data mining techniques will grow in importance as the volume of data from individual satellites, and clusters of satellites, grow.

## References

- [1] Orfali, R., Harkey, D., Edwards, J. (1996). *The Essential Client/Server Survival Guide*. 2<sup>nd</sup> Edition. New York, John Wiley & Sons, Inc.
- [2] Berson, A., Smith, S. (1997). *Data Warehousing, Data Mining, & OLAP*. New York, McGraw-Hill.
- [3] Gifford, C., Montgomery, G., *ModelQuest Satellite Telemetry Anomaly Resolution*, Final Report, AFRL Contract No. AFRL-VS-TR-1999-1068, 1999.
- [4] Gonsalves, P., Ivanov, A., *Autonomous Learning Agent System for Spacecraft Telemetry Processing*, Final Report, AFRL Contract No. AFRL-VS-TR-1999-1094, 1999.