Discovering Service Inventory Demand Patterns from Archetypal Demand Training Data

EUGENE A. BEARDSLEE AND THEODORE B. TRAFALIS

Department of Industrial Engineering, University of Oklahoma, USA

Abstract

A critical element of establishing inventory control parameters and consumption forecasts is determining the nature of demand generated by the processes supported by the inventory. Often these processes are stochastic and generate demand for inventory items that cannot be easily characterized using stationary probability distributions. The research presented here describes a method of classifying demand patterns using several data mining algorithms including support vector machine, C4.5 decision tree, Bayesian networks and Naïve Bayes classification methods. Using simulation to model various demand source processes, archetypal demand time series were generated for use as training data input for the data mining analysis. Once the data mining models were trained, actual inventory transaction sets were classified into four types of demand: periodic, seasonal, level shift and sparse. These classified demand series were then validated against a set of hand classified time series, and through the application of the Box-Jenkins method for characterizing time-series information. Results show that, using this method, managers of service part inventories could identify demand patterns with sufficient reliability and flexibility to improve inventory management.

Keywords: Non-stationary Demand, Inventory Control, Data Mining, Transaction Time Series

Introduction

With the advent of increased computing resources available to support inventory and supply chain management, the opportunity to leverage the power of advanced optimization method applications is now more promising than ever. Among the more difficult problems addressed by numerous research efforts over the last century is that of determining the optimal inventory control policy that will minimize the cost of inventory while still providing a high level of support to the community served by the inventory material. The specific inventory we examine in this paper is the inventory supporting repair or maintenance services. In general, this type of inventory contains multiple stock keeping units (SKUs) in varying quantities with varying individual item values.

The best design of resupply networks and the optimal allocation of inventories within service parts supply chains is of unquestionable importance to the economical maintenance of equipment (Muckstadt, 2005). Specifically with regard to service part or maintenance type inventories, as a group the repairable items comprise the largest part of the [U.S. Air Force] spares budget; in 1990 the Air Force had over $31 billion invested in repairables (Sherbrooke, 2004). Customers have become more demanding and require customized products delivered in a consistently timely manner. As competition intensifies, product shortages and stock-outs significantly affect companies' reputations (Paschalidis et al, 2004).

One of the key difficulties related to service part inventories arises from the fact that much of the demand for these inventory items is driven by system or equipment failure. In addition, regularly scheduled maintenance can also be a source of inventory item demand. Along with scheduled maintenance, additional system problems often are identified for repair during the actual scheduled maintenance task, spawning additional demand for service parts from the inventory. Forecasting the inventory item demand generated by these interacting processes is usually an unsuccessful endeavour. This is especially evident when considering the unknown affected interrelated system components, each requiring different sets of parts to complete each maintenance task. The combination of these factors, including non-scheduled and non-recurring events such as vehicle recalls or military weapon system upgrades, contribute to the description of a highly complex inventory management scenario. A significant aid to this process would be the ability to identify the type of demand structure dominating a given SKU inventory requirement.

In conjunction with the complex demand structure of the problem, the value of the individual items in a maintenance inventory varies greatly. Individual items may vary from as little as $0.001 to $10,000. The inventory may also be comprised of many thousand distinct parts supplied from a variety of vendors, each with their own replenishment lead time.

The remaining sections of this paper are organized as follows. Section 2 discusses the definition and background of the problem under analysis and provides an examination of related research. Section 3 presents the methods of defining the demand patterns through simulation. Section 4 provides a description of the data mining methods examined. The following section 5 describes the experimentation methods and provides a compilation of the results. The final section gives conclusions drawn from the analysis of the results.

Background and Problem Definition

Attempting to determine the optimal inventory policy for a given problem through the use of simulation is not a new approach. In research focusing on a multi-item inventory, (Ramirez Cerda and Espinosa de los Monteros F., 1997) evaluated a new inventory policy which tries to compensate for orders placed for multiple items every day. They introduced two parameters to the typical reorder level policy, one defining the period of evaluating the applied policy and a *can order* level to improve stock position and order efficiency through "opportunistic" ordering. They relied upon the Chi-Square goodness of fit test to "classify" the demand distributions, which were then used in simulation to examine their new inventory policy. Determining a more accurate method of characterizing inventory demand also led to the use of a distorted normal distribution which exploits the orthogonal polynomial of the normal distribution called the Hermite polynomial (Sugita and Fujimoto, 2005). However, research on inventory optimization of slow moving items led (Grange, 1998) to conclude that a) identification of distributions from typically available demand history can be extremely difficult, and b) misapplication of a demand distribution will yield unsatisfactory inventory optimization results.

Approaching the problem of selecting an optimal inventory control policy is intimately connected to the problem of forecasting or characterizing the demand for the inventory items. One of the shortfalls found in much of the research focused on inventory control optimization is that of attempting to fit a stationary probability distribution to an inherently non-stationary data stream. Our research goal is to define a small set of typical or archetypal inventory demand pattern classes, and show that employing training data generated from simulations can effectively be used to classify actual inventory demand transactions. With the ability to classify the type of demand an inventory item is displaying, determining an adequate forecasting methodology and subsequently an optimal control policy becomes a more attainable goal. For example, while the Box-Jenkins is a commonly used approach for forecasting demand, if the historical transaction time series does not contain enough data points, the forecast it produces for the given item will be zero or very close to zero. If the few data points are actually significant, an additional method of forecasting must be applied to these exceptional cases.

Demand Time Series Generation

Generating four separate types of transaction time series patterns was the goal of the archetype demand simulation. These four types are identified as periodic, seasonal, level-shift, and sparse. Periodic demand refers to demand patterns that have a relatively regular pattern with a short demand interval. In the experimentation that follows, the transaction sets were assumed to be monthly transaction data, therefore a periodic demand pattern would be a set of demands occurring monthly with a high degree of regularity. A seasonal demand pattern is defined here as a periodic demand pattern with an interval between 3 and 12 months. This seasonality may exist concurrent with a periodic demand pattern or as a purely seasonal demand series. The level-shift demand pattern is defined as a significant positive or negative change in the regular demand pattern. Finally, the sparse demand pattern includes any demand time series that contains so few demand data points that reasonable estimates of future demand cannot reliably be forecasted by conventional methods.

The archetype demand simulation models were created using Visual SLAM and AweSim simulation software (Pritsker and O'Reilly, 1999). Each demand transaction stream generated represents 36 months or three years of simulated service part inventory demand. To simulate the *periodic* demand pattern, the mean and standard deviation of the demand were read from a set of sample demand series, including mean demand values ranging from 1 per month to 8500 per month with standard deviation values representing both low and high variance demand. These values were then used as the mean and standard deviation of a log normal distribution from which the mean demands of the transaction series samples were drawn. A separate sample from a log normal distribution provided the standard deviation of the demand for the transaction series. These samples of mean demand and standard deviation were then used as the mean and standard deviation in a separate log normal distribution to provide the simulated number of parts needed in a given repair operation. This number of parts needed per repair operation was then multiplied by a sample from a Poisson distribution, with a mean of 1, providing the estimate of the number of repair events in a given month. The defining aspect of the *seasonal* demand patterns were the same whether concurrent with an underlying *periodic* demand pattern or not. A seasonality or season periodicity was selected from a uniform distribution from 3 to 12. A seasonal demand delta was selected from a log normal distribution as a uniform increase or decrease of the mean demand. This positive or negative delta was then applied to the transaction stream at the season interval. The *level-shift* demand pattern was simulated by defining a single point in the transaction stream where a positive or negative demand shift occurs. This shift point was a sample from a uniform distribution from 4 to 30. The shift delta was determined by sampling from a uniform distribution from .3 to 5.5 and multiplying this factor by the mean demand. After the demand shift point had passed in the demand series, the *level-shift* was applied by adding (or subtracting) this value to (or from) a *periodic* demand transaction stream. The final demand pattern, the *sparse* demand pattern, was generated by creating the demand mean in the same method as the *periodic* demand. Then the arrival of the demand was controlled by taking the nearest integer sample from a normal distribution with a mean of 0 and standard deviation of 8 and testing whether this value was equal to 0. If this test was true, a demand was generated, otherwise a demand of 0 was generated in the demand series.

Data Mining Methods Examined

The archetypal training sets output from the simulation and the four test data sets were evaluated against the Naïve Bayes, Bayesian network, C4.5 decision tree classification algorithms, and the sequential minimal optimization (SMO) algorithm for training a support vector classifier as implemented within the Weka Explorer data mining workbench, version 3.4.4 (Witten and Frank 2005). The Naïve Bayes probabilistic learner and the C4.5 decision tree algorithm (J4.8) were selected because they represent two quite different approaches to machine learning and they are relatively fast, state-of-the-art algorithms that are often used in data mining applications (Hall and Holmes 2003). The Bayesian network method of data mining was selected for two reasons. First, it represents a method which combines the strengths of the decision tree learner and the probabilistic learner through the use of directed acyclic graphs. Bayesian networks are a special case of a wider class of statistical models called graphical models, which include networks (called Markov networks) with undirected edges (Witten and Frank 2005). Second, it is very likely that the demand points which comprise the transaction series contain a high degree of dependency among them. Therefore, the validity of using the Naïve Bayes probabilistic learner must be considered in light of the attribute independence assumption. Fortunately, Bayesian networks help answer this concern because they allow for modelling of arbitrarily complex dependencies between attributes (Wang and Webb 2002). The fourth data mining method applied uses John C. Platt's support vector classification approach called the sequential minimal optimization (SMO) algorithm.

Experimentation

Ten training files were generated using the archetype demand simulation. Each simulation run used different input parameters for the mean demand and the amount of demand variation represented in the transaction streams created for the training datasets. These training files were then used for input into each of the data mining methods mentioned in the previous section. The trained classification models built with each of the data mining algorithms using 10-fold cross validation were then used to classify four test files. The four test files represent a set of hand-classified aircraft repair item inventory transactions (*H*), a set of aircraft rivet transactions "classified" using the Box-Jenkins time series analysis (*R*), a set of military tracked vehicle repair part transactions analyzed using Box-Jenkins (*V*), and a set of oil and chemical transactions analyzed with the Box-Jenkins method (*P*). Three of these test data files come from service repair part inventories. The oil and chemical demand transaction series were included to provide the initial validating support that the demand sources modelled in the simulation were service repair part items and not commodity type inventory items like oil. Therefore, the expectation was that the classifiers would perform poorly on the oil and chemical test data. The results of the testing are presented in table 1. The Bayesian network classifier was clearly the most effective of the four tested, followed by the C4.5 decision tree classifier.

Success Rate (%)

	Training	**Test *H***	**Test *R***	**Test *V***	**Test *P***
Bayesian Network	70.31	69.71	75.37	55.38	41.30
C4.5	68.11	66.73	73.11	54.26	37.77
SMO Linear	46.49	63.64	68.07	43.59	26.68
Naïve Bayes	41.71	52.70	57.72	44.31	26.11
SMO 2-Poly	37.69	59.46	74.51	41.61	20.11

Kappa Statistic

	Training	**Test *H***	**Test *R***	**Test *V***	**Test *P***
Bayesian Network	0.59	0.44	0.43	0.35	0.18
C4.5	0.57	0.38	0.38	0.34	0.17
SMO Linear	0.24	0.22	0.14	0.06	0.03
Naïve Bayes	0.20	0.12	0.01	0.09	0.06
SMO 2-Poly	0.03	0.02	0.11	0.00	0.00

Table 1: Data mining results

Conclusions

Using the archetype demand simulation to generate transaction series for use as training input to data mining algorithms appears to provide an effective means of building classification models for service part inventory items. Forecasting demand for this type of inventory presents numerous challenges and prediction accuracy is consistently low. By classifying the transaction streams into the demand archetypes, an inventory item manager could focus automatic methods of forecasting, such as the Box-Jenkins method, toward the analysis of more "well-behaved" inventory items. This would raise both the accuracy of the forecasts and the confidence of the inventory managers. Also, alternative methods could then be used for the forecast and control of the inventory items with sporadic or no demand history. In addition, knowing the type of demand an inventory item is expressing through its transaction history could be exploited to alter control policies in order to maintain more appropriate inventory levels without decreasing service rates.

References

1. Ghiani, G., Laporte, G., and Musmanno, R.,2004, *Introduction to Logistics Systems Planning and Control*, (Wiley, West Sussex).
2. Simchi-Levi, D., Chen, X., and Bramel, J., 2005, *The Logic of Logistics: Theory, Algorithms, and Applications for Logistics and Supply Chain Management, 2nd ed.*, (Springer, New York).
3. Muckstadt, J.A., 2005, *Analysis and Algorithms for Service Parts Supply Chains*, (Springer, New York).
4. Paschalidis, I. C., Liu, Y., Cassandras, C. G., Panayiotou, C., 2004, “Inventory control for supply chains with service level constraints: a synergy between large deviations and perturbation analysis”, *Annals of Operations Research* 126, pp. 231-258.
5. Ramirez Cerda, C. B., and Espinosa de los Monteros F., A. J., 1997, “Evaluation of a (R,s,Q,c) multi-item inventory replenishment policy through simulation”, *Proceedings of the 1997 Winter Simulation Conference*, pp. 825-831.
6. Grange, F., 1998, “Challenges in modeling demand for inventory optimization of slow-moving items”, *Proceedings of the 1998 Winter Simulation Conference*, pp. 1211-1217.
7. Sugita, K., and Fujimoto, Y., 2005, “An optimal inventory management for supply chain considering demand distortion”, *Industrial Informatics, 3rd IEEE International Conference on (INDIN)*, pp. 425-430.
8. Pritsker, A. A. B., and O’Reilly, J. J., 1999, *Simulation with Visual SLAM and AweSim*, (John Wiley & Sons, New York).
9. Hall, M.A., and Holmes, G., Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering*, 2003, **15**(6), 1437-1447.
10. Wang, Z., and Webb, G.I., Comparison of lazy Bayesian rule, and tree-augmented Bayesian learning, *Proc. of the 2002 IEEE Int. Conf. on Data Mining (ICDM 2002)*, 9-12 December 2002, pp.490 – 497.
11. Witten, I.H. and Frank, E., *Data Mining: Practical Machine Learning Tools and Techniques -2nd ed.*, (Morgan Kaufmann, San Francisco), 2005.