A Meta Forecasting Methodology for Large Scale Inventory Systems with Intermittent Demand

Vijith Varghese, Manuel Rossetti
Department of Industrial Engineering
Bell Engineering Center
University of Arkansas, Fayetteville, AR 72701, USA

Abstract

This paper presents a meta-forecasting approach for recommending the most appropriate forecasting technique for an intermittent demand series based on a multinomial logistic regression classifier. The meta-forecaster is based on a mapping between a demand attribute space and the best forecasting technique. The demand attribute space is based on the estimates from the demand series of the following attributes: probability of non-zero demand after zero demand, probability of non-zero demand after non-zero demand, mean demand, demand variance, lag 1 correlation coefficient of the interval between non-zero demand and lag 1 correlation coefficient. Based on the mapping, the best forecasting technique for an unknown demand vector can be predicted. Given the demand series, the demand attributes are estimated and then the classifier is used to predict the best forecasting technique. After training, the classifier was tested. The results indicate an accuracy rate of 70.87% for the recommended best forecasting technique; and an 87.94%accuracy rate for the recommended top two forecasting techniques.

Keywords
Intermittent demand forecasting, classification, inventory control

1. Introduction and Motivation
The focus of this research is on inventory system with intermittent demands. Intermittent demand is characterized by demand data that has many time periods with zero demands. However, other definitions can be found in the literature [14] [3] [9], for intermittent demand. Intermittent demand is hard to model using conventional distributions and is hard to forecast. It is commonly found in vital military supply networks. The Naval Aviation Maintenance Program (NAMP) of the US Navy is a multi-echelon supply network with 3 levels. At the lower level the demand for a reparable spare part arrives when the part fails. The demand at the higher levels occurs when the lower levels are unable to repair the failed spare part. The repair cycle may include shipping time, processing time, repair time, waiting time, and delivery time. Because of the repair cycle as well as the failure cycle, the demand for repair and spare parts is often intermittent in nature. This research is motivated by Varghese [20] [21], in which a naïve categorization (classification is the term used in statistics papers) scheme for intermittent demand was created. Each demand category was mapped to an intermittent demand forecasting technique most appropriate for that demand category. Each demand category is characterized by its demand attributes π_Z, CV_{NZ}^2 and $\phi_{1,NZ}$: probability of zero demand, coefficient of variation of non-zero demand, lag 1 correlation coefficient of non-zero demand respectively. Varghese [20] [21] showed that this naïve mapping scheme can be used to choose an appropriate forecasting technique for demand series that are similar to those used to develop the classification scheme. First, the estimates of the three demand attributes (demand attribute vector) of the SKU (stock keeping unit) are computed. Second, the corresponding demand category is identified and then the forecasting technique to which the demand category is mapped is chosen. This meta-forecasting technique was applied on a real data and was found to reduce the forecast error, when compared with the existing technique.

A meta-forecaster is relevant in large scale inventory systems where inventory managers encounter decision making for the selection of a forecasting technique. It is a tool used to select the best forecasting technique for a given demand series. The selection of the best forecasting technique for a given demand series can be approached in several ways. The most common approach is to select a forecasting technique that minimizes the forecast error using the available demand history. Another approach is to forecast based on several forecasting techniques and subsequently combine the forecasted values into a single forecast. Armstrong [4] recommends this approach instead of selecting a single forecasting technique. This paper develops an approach, which we term meta-forecasting. In a meta-forecasting approach, one creates a classification scheme for a demand attribute space (e.g. a 6-dimensional vector space of demand attributes). Our classifier handles categorical responses (best forecasting technique) that are nominal. There are various multi-variate classification approaches that can be considered. Hastie et al. [7] lists Nearest neighbor classification, Multinomial logistic regression, Artificial neural network, Discriminant analysis, Separating hyper plane, Support vector machine, Flexible Discriminant analysis etc. to train data for classification. Our meta-forecaster is based on classification using multinomial logistic regression. The classification scheme requires constructing a demand attribute space and mapping to its best forecasting technique. This is followed by training this data and thus developing a classification scheme. The classification scheme holds the demand attributes vector and the best forecasting technique associated with it. This research considered the forecasting techniques: moving average, simple exponential smoothing, Syntetos' approximation method [18], and a cumulative average (CA) forecasting method. We evaluated the efficacy of the classification scheme using the standard accuracy test for statistical learning techniques, which is discussed in section 2.2. An evaluation of the meta-forecasting approach with the combining-forecast approach in Armstrong [4] is being considered for future work.

1.1. Intermittent Demand and Categorization Scheme

The literature refers to the "hard to forecast" demand scenarios as intermittent demand, lumpy demand, erratic demand, sporadic demand, slow-moving demand etc. and often these words are used interchangeably which amounts to much confusion. As previously discussed, the demands are generally characterized by the attributes: intermittence (or sporadicity) and lumpiness. Usually, intermittent demand (or sporadic demand) is defined as demand occurring randomly with many time periods with zero demands. However, this limits the definition to the attribute of intermittence or sporadicity. Silver [14] proposed a definition for intermittent demand as "*infrequent in the sense that the average time between consecutive transactions is considerably larger than the unit time period, the latter being the interval of forecast updating*." Smart [3] defined intermittent demand as a demand series with at least 30% zero demand. Representative US Navy [13] inventory managers consider those demand series with less than or equal to 60 - 70% non-zero demands as intermittent. Johnston et al. [9] proposed that if the mean interval between non-zero demands is 1.25 times greater than the inventory review period, the demand series can be considered as intermittent. Most of the definitions of intermittent demand (or sporadic demand) do not include the demand attribute: lumpiness. Slow demands are usually defined as those with infrequent demands, which occur in very few units [10] [18] [25]. Slow demands are usually intermittent demands. Meanwhile erratic (or irregular) demand is described as in [18] as patterns with high variability in non-zero demands. Syntetos [18] based his definition on the demand size and excluded demand incidence and so did Silver [14]. Syntetos [18] defined lumpy demand as those demand patterns with some zero-demands and with non-zero demand having high variability. He considered all lumpy demands as intermittent demands; however not all intermittent demand is lumpy demand. Ward [23] also used intermittent demand and lumpy demand interchangeably. These types of demand scenarios overlap with similar characterizations of intermittent demand. In this paper we view these demand scenarios by the difficulty to forecast. Some of the previous literature on demand classification on demand attributes is William [24] and Syntetos [18] [19]. William's categorization scheme [24] is one of the earliest ones of its kind and is based on a concept called variance partitioning, in which the variance of the lead time demand is split, to classify the demand. His classification represented intermittence, by how often the demand occurs during the lead time. He also considered the variance of non-zero demand, commonly called as lumpiness. Syntetos et al. [18] [19] in their research on intermittent demand forecasting techniques, proposed a demand categorization scheme with recommendations for an appropriate cut-off value for squared coefficient of variation and mean interval between non-zero demands.

The research literature on demand characteristics usually considers only intermittence and lumpiness. Varghese [20] [21] considered dependence (through $\phi_{1,NZ}$) in addition to intermittence and lumpiness. In this research we expand the demand attribute space to dimensions: $p_{01}, p_{11}, \mu, \sigma^2$, ϕ_1 and $\phi_{1,ibt}$, probability of non-zero demand after zero demand, probability of non-zero demand after non-zero demand, mean demand, demand variance, lag 1 correlation coefficient and lag 1 correlation coefficient of the interval between non-zero demand. These demand attributes are expected to measure not just the intermittence and lumpiness but also the burstiness of demand. Bursty demand is seen usually in telecommunications network; it refers to a demand scenario in which non-zero demands arrive consecutively. Bursty demand can be intermittent demand; there can be time periods when there is no demand and when a demand occurs, it is followed by consecutive non-zero demands. It is most likely that the highly bursty demands are those with highly positively correlated intervals between transactions. An expanded demand attribute space, including burstiness may be able to improve the accuracy of the classifier. The application of the classification scheme will be in the mapping of the demand attribute vector to its best forecasting technique. Using this mapping, we can predict the best forecasting technique for a given demand scenario.

1.3 Intermittent Demand Forecasting

The simpler traditional forecasting methods like simple exponential smoothing and moving average are often unsuitable in intermittent demand scenarios. There are several forecasting techniques relevant for intermittent demand. These techniques are discussed in detail by Varghese [20] [21]. Croston's [6] approach and its variant Syntetos [18] are two of the primary techniques. Syntetos was recommended by [Syntetos 2006, Boylan 2007, Eaves 2004] to be very competitive because it removes the bias associated to Croston's technique. In addition to these approaches, Willemain [24] developed a non-parametric bootstrapping approach forecasting especially intermittent demand. Meanwhile, Snyder [16] proposed a parametric bootstrapping to forecasting slow demand. Interested readers are referred to [22] for more information on parametric and non-parametric bootstrapping approaches to forecasting intermittent demand. The performance of forecasting techniques can be measured by the forecast error's mean absolute deviation (MAD), mean square error (MSE) and mean absolute percentage error (MAPE). Though MAD, MSE and MAPE are sufficient to compare between errors associated with each of the demand scenarios, when it comes to identifying the best forecasting technique the winners may be different across each of these error metrics. For illustrating the applicability of this meta-forecaster approach we will select the best technique based on MAD. We plan to extend the training of the meta-forecaster based on other forecast metrics in future research. The following section discusses the methodology that we adopted in building the meta-forecaster. We describe how the demand attribute vector space is created and mapped to its best forecast technique. Section 2.1 shows how multinomial logistic regression is applied on the data and in section 2.2 we discuss how the meta-forecaster is tested and we summarize the results.

2. Experiment and Results

The classification scheme requires constructing a demand attribute space and mapping to its best forecasting technique. This is followed by training this data and thus developing a classification scheme. The demand attribute space is of 6-dimension $\vec{X} = (p_{01}, p_{11}, \mu, \sigma^2, \phi_1, \phi_{1,ibt})$. The 6-dimensional demand attribute space is based on real data provided by the US Navy. This data set consists of 3725 demand series for items which are intermittent in nature. The best forecasting technique corresponding to each of the 3725 demand attribute vectors were selected through a simulation approach. We simulate the demand at each of the vectors. The demand generation is based on the demand attributes $\vec{Y} = (p_{01}, p_{11}, \mu_{NZ}, \sigma^2_{NZ}, \phi_{1,NZ})$ estimated for each of the 3725 items. The demand attributes are probability of non-zero demand after zero demand, probability of non-zero demand after non-zero demand, mean non-zero demand, non-zero demand variance, and lag 1 correlation coefficient of non-zero demand respectively. During the simulation run, forecasts are made based on relevant forecast techniques. At the end of the run, the MAD associated with each forecast techniques is observed. The demand generation and corresponding forecasts are replicated for each of the demand attribute vectors. After the replication, for each demand attribute vector the best forecasting technique is selected. The comparison across the forecasting techniques is based on a multiple comparison approach called multiple comparison with the best (MCB) developed by Hsu [8]. At each of these vectors, the forecast techniques with low MAD were identified and ranked. Hsu's approach compares each forecasting technique with the best of the remaining forecasting techniques [8] [11]. When compared with other multiple comparison techniques, in Hsu's MCB approach, the comparison procedure is implemented in a single-stage. In addition, the relative performance of each forecasting technique can be estimated. Thus we obtain the demand attribute space with each of the vectors mapped to its best forecasting technique. The training set within the 3725 mappings are trained and following that the test data set is tested to estimate the accuracy of the classifier. This approach is illustrated as in Figure 1.

The experimental set up is implemented in a Java platform and is integrated with a Java simulation library (JSL). The JSL is an open source simulation library for Java. The JSL has packages that support random number generation, statistical collection, basic reporting, and discrete-event simulation modeling. The interested reader may also refer to Rossetti [12] for further information on the design and use of the JSL. For our experiment we decided to replicate 100 times. The demand is generated by the artificial demand generator discussed in Varghese [20]. This is a demand generator that models a demand occurrence process and demand amount process. The demand occurrence process is a 2 stage Markov Chain and the demand amount process is based on an algorithm named ARTA (Auto regressive to anything). The ARTA algorithm induces correlation within the demand amount process [5]. The user controls the intermittence, lumpiness and dependence by assigning the demand generator with the input vector, $\vec{Y} = (p_{01},\ p_{11}, \mu_{NZ}, \sigma^2_{NZ}, \phi_{1,NZ})$, .. The input demand attribute vector corresponds to the 6-dimensional demand attribute vector with which the demand attribute space is then created from the generated demand series. The forecasting methods that we selected are Simple Exponential Smoothing (SES), Moving Average (MA), Syntetos' Approximation method and Cumulative Average (CA). The levels are: Simple Exponential Smoothing with alpha value 0.1 and 0.2: SES(0.1) and SES(0.2), Moving Average with N value 19 and 9: MA(19) and MA(9), Syntetos' method with alpha value 0.1and 0.2: Syntetos(0.1) and Syntetos(0.2) and Cumulative average. The best technique is selected from these techniques. Since, Syntetos is an improvement of Croston, we decided to select Syntetos over Croston. The range of values of alpha are recommended by previous research [6] [9].

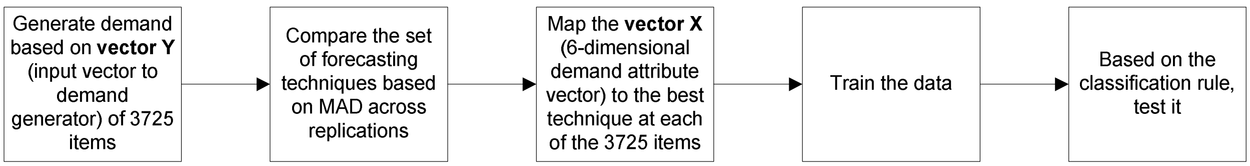

Figure 1 Methodology

2.1 Multinomial logistic regression

One of the widely used classification schemes to train data, Multinomial logistic regression or Polytomous logistic regression or Multinomial logit model is considered in our approach. It is very similar to logistic regression modeling. In logistic regression modeling, the categorical response variable has only two levels and the probability of falling into each category is modeled as a binomial distribution. Meanwhile in multinomial regression, the response variable is categorical and nominal and has more than two levels and the classification is based on multinomial distribution. In multinomial logistic regression [1] [15], we develop a model relating the predictors to the probability of falling in each of the levels, $p_J = 1, \cdots m$ of the response variables. In our case the predictors of the classification scheme are the demand attributes: $p_{01}, p_{11}, \mu, \sigma^2, \phi_1$ and $\phi_{1,ibt}$; and the response variable is the best forecasting technique. The p_J is assumed to be multinomial and there is no inherent ordering within the *m* response levels. Let $(p_{01},\ p_{11}, \mu, \sigma^2, \phi_1, \phi_{1,ibt}) = (\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6)$. The logit odds or probability of falling in a category relative to the baseline category (*base*) is given by:

$$log\left(\frac{p_j}{p_{base}}\right) = \beta_{0j} + \beta_{1j}\,x_1 + \cdots + \beta_{1j}\,x_6 \qquad \begin{matrix} j = 1, 2, \ldots 7 \\ j \neq base \end{matrix} \tag{1}$$

SES (0.1) is selected arbitrarily as the base line category arbitrarily. From this we can compute $p_J = \frac{exp(\beta_{0j}+\beta_{1j}\,x_1+\cdots+\beta_{1j}\,x_6)}{\sum_{l=1}^{m} exp(\beta_{0l}+\beta_{1l}\,x_1+\cdots+\beta_{1l}\,x_6)}$, where β estimate of the baseline category is equal to zero. The β estimate can be obtained by the maximum likelihood estimation (MLE); the log-likelihood is given by,

$$\mathcal{L}(\vec{\beta}) = \sum_{j=1}^{6} \sum_{y_i=j} p_{j(i)} \tag{2}$$

The second summation is over all observations *i* with response level *j*, and $p_{j(i)}$ is the probability, substituting in the predictor value the *i*th observation. The demand attribute vector space mapped to the best forecasting technique will be created after the simulation run. The coefficients, β estimate from Equation 2 can be computed using the CATMOD procedure in SAS software [2] [17]. Thus the probability p_j can be estimated for an unknown demand attribute vector. The forecasting technique with highest probability will be selected. The use of multinomial regression to train the demand attribute vector data and classify into the appropriate forecasting technique and subsequently using the classification rule to predict the best forecasting technique; is the first of its kind.

2.2 Training

The first question that has to be addressed is how to train and test the data. One of the naïve approaches is to train and test by re-substitution in which we train and test the same data. However, the accuracy rate in re-substitution may be over estimated. The other approach is the cross validation approach and it is of 3 types: hold-out cross validation, n-fold cross validation and "leave one out" cross validation. Each case has its advantages and disadvantages. Hold-out cross validation trains on fewer observations when compared to others. However, hold-out cross validation is considered computationally less intensive. Hence, in this study, we opt for hold-out cross validation. Now the next question is how much of the data has to be trained or how much is to be held out. Lacking specific guidance from the literature, we arbitrarily decided to use 70% of data for training and to test on the 30% of the remaining data. The data for training and testing is randomly chosen, in such a way that the percentage of each of the 7 categories in the training data and the testing data will be the same as that in the whole data. Thus, we classify the test data and training data. The multinomial regression was analyzed by the SAS procedure named CATMOD, which models the data based on multinomial logistic regression where βestimate is based on MLE. Table 1 summarizes the βestimate each category with respect to the arbitrarily selected baseline category SES (0.1) and the corresponding 6 predictors.

Table 1 $\boldsymbol{\beta}$ estimate each category

	SES(0.1)	MA(19)	Syntetos (0.1)	MA(9)	Syntetos (0.2)	CA
intercept	-4.81	-3.41	-13.31	*	-5.51	-6.53
p_{11}	-12.34	*	-14.09	21.97	-55.20	-9.94
p_{01}	34.09	65.60	77.86	102.90	87.76	56.13
ϕ_1	-60.27	-56.25	-64.76	-23.31	-30.33	-77.14
μ	9.49	-15.32	*	-85.42	9.49	5.58
σ^2	-0.99	1.27	*	6.92	-0.79	-0.48
$\phi_{1,ibt}$	-45.97	-29.05	-33.7102	*	-24.82	-62.39

The SAS output showed that the MLE estimation is a good fit. It also shows the statistical significance of each predictor on the response. Those predictors that are not statistically relevant are excluded from the regression equation, (* value of βestimate in Table 1 denotes that the predictor is not statistically significant). For example the βestimate of the predictor "mean" has no statistically significant relevance with the response Syntetos (0.1). Once the significant coefficients are estimated, we compute the probability that an observation from the test data will fall into each category. Our classification rule then chooses the category with the maximum probability as the most appropriate predicted best-forecast-technique to this observation. We compare the recommended technique with the actual best technique which the test data has already observed from the simulation. The accuracy for how many times the best forecasting technique is predicted as the recommended best was estimated as 70.87%. The accuracy of how many times the best forecasting technique is predicted as the one of the two recommended best was also estimated and is 87.94%. This is a reasonable estimate for the accuracy. We see the usability of the meta-forecasting approach. This approach is relevant in large scale inventory system where selecting the best technique is not a trivial problem. The meta-forecaster eases this problem.

3. Conclusions and Future Research

With a classification rule, we can predict the most appropriate forecasting technique for a new data set. We estimated an accuracy rate of 70.87% for the best forecasting technique to be predicted as the best and 87.94% for the best forecasting technique to be predicted as one in the two best. This is a reasonable level of accuracy. This is an initial work towards the concept of meta-forecasting; the accuracy of the meta-forecaster based on the 6 demand attributes justifies extending the research to considering means to improve the meta-forecasting approach. In future research, an organization's existing forecasting technique will be compared to the technique selected via the meta-forecaster. This will give us a better metric to the benefit of using the classification rule. We can extend the demand attribute space by considering other attributes like the mean and the variance of interval between non-zero demand and thus possibly improve the accuracy of the classifier. In addition to that, we may consider generating demand attribute space using multi-variate random number generation with correlation induced by approaches like NORTA [5]. In this paper, the demand attribute space is based on the existing 3725 demand attribute vectors; NORTA and similar approaches generates demand attribute vectors according to distribution of each demand attributes without losing the correlation information. A larger population of demand attribute vectors gives a better defined attribute space. In future research, we may also consider mapping the demand vector with the most appropriate inventory practice (e.g. base-stock policy, (s, S) policy, (r, Q) policy etc.).

Acknowledgements

This material is based upon work supported by the U.S. Air Force Office of Sponsored Research, US Navy (NAVSUP) and the Air Force Research Laboratory. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors.

References

1. Agresti, A. 2002. Categorical Data Analysis, 2nd ed. Wiley-Interscience, New York
2. Allison, P.D., SAS Institute. 1999. Logistic regression using the SAS System: Theory and Application, SAS Institute, Cary, N.C.
3. Anonymous 2003, "New Technology Enables Accurate Forecasting of Intermittent Inventory," Inventory Management Report, 3(2), 2.
4. Armstrong, J.S., 2001, Principles of Forecasting: a Handbook for Researchers and Practitioners, Kluwer Academic, Boston, MA.
5. Banks, J., Carson, J. S., Nelson, B. L. and Nicol, D. M. 2005. Discrete-event System Simulation, 4th ed. Pearson Prentice Hall, Upper Saddle River, NJ.
6. Croston, J. D., 1972, "Forecasting and Stock Control for Intermittent Demands," OR Quarterly, 23(3), 289-.
7. Hastie, T., R. Tibshirani, J.H. Friedman. 2001. The Elements of Statistical Learning: Data mining, Inference, and Prediction, Springer, New York.
8. Hsu, J.C., 1996, Multiple Comparisons: Theory and Methods, 1st Edition, Chapman & Hall, London.
9. Johnston, F. R., and Boylan, J. E., 1996, "Forecasting for Items with Intermittent Demand," Journal of the Operational Research Society, 47(1), 113-121.
10. Kwan, H.W., 1991, "On the Demand Distribution of Slow Moving Items," Ph.D. dissert., Lancaster Univ.
11. Nelson, B.L., 1992, Statistical Analysis of Simulation Results," appears in Handbook of Industrial Engineering, Salvendy, G. (eds.), IIE; Wiley, Norcross, Ga.; New York, 2586-2588.
12. Rossetti, M.D. 2007. JSL: An Open-source Object-oriented Framework for Discrete-event Simulation in Java. International Journal for Simulation and Process Modeling
13. Schad, T., LCDR, 2004, "Questions Related to 9 COG Material Category Matrix."
14. Silver, E.A., Pyke, D.F., and Peterson, R., 1998, Inventory management and production planning and scheduling, 3rd Edition, Wiley, New York.
15. Simonoff, J.S. 2003. Analyzing categorical data, Springer, New York
16. Snyder, R., 2002, "Forecasting Sales of Slow and Fast Moving Inventories," European Journal of Operational Research, 140(3), 684-699.
17. Stokes, M.E., C.S. Davis, G.G. Koch. 2001. Categorical data analysis using the SAS system, 2nd ed. SAS Institute; Wiley, Cary, NC; New York
18. Syntetos, A., 2001, "Forecasting of Intermittent Demand," Ph.D. dissertation, Brunel University.
19. Syntetos, A. A., Boylan, J. E., and Croston, J. D., 2005, "On the Categorization of Demand Patterns," Journal of the Operational Research Society, 56(5), 495-503.
20. Varghese, V.M., 2005, "Categorizing and forecasting intermittent demand," Master's thesis, University of Arkansas, Fayetteville.
21. Varghese, V. M., and Rossetti, M. D., 2008, "A Classification Approach for Forecasting Intermittent Demand," Proceedings of the 2008 Industrial Engineering Research Conference, Vancouver Canada
22. Varghese, V. M., and Rossetti, M. D., 2008, " A Parametric Bootstrapping Approach to Forecast Intermittent Demand," Proceedings of the 2008 Industrial Engineering Research Conference, Vancouver
23. Ward, J. B., 1978, "Determining Reorder Points when Demand is Lumpy," Management Science, 24(6), 623-633.
24. Willemain, T. R., Smart, C. N., and Schwarz, H. F., 2004, "A New Approach to Forecasting Intermittent Demand for Service Parts Inventories," International Journal of Forecasting, 20(3), 375-387.
25. Williams, T. M., 1984, "Stock Control with Sporadic and Slow-Moving Demand," Journal of the Operational Research Society, 3(5), 939-948.