**Machine Learning Approach to counteraCt the bullwhip effect in supply chain**

**Shivahari Revathi Venkateswaran, M.S.\***

**Himlona Palikhe, Ph.D.**

**Manish Ranjit, Ph.D.**

**Northeastern University**

**\***[**venkateswaran.sh@northeastern.edu**](mailto:venkateswaran.sh@northeastern.edu)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Abstract**

The Bullwhip effect is a phenomenon that tends to increase the volatility in the demand distribution when moving upstream in the supply chain. Because of this, most of the industries are failing to accurately forecast the demand. This leads to increased manufacturing cost, higher inventory level, longer replenishment lead times, low product availability, higher transportation cost and overall, it reduces the supply chain profitability. Previous studies have proposed different methods to overcome the Bullwhip effect such as improving demand forecasting, reducing order batching, reducing the incentives of forward buying, better sharing of information, shorten lead and review period times and designing single-stage replenishment control. This study focuses on the customer-centric approach of categorizing the customers by incorporating Machine Learning techniques for better understanding the customers’ purchasing behavior. Specifically, this study analyzes the hidden patterns and similarities that exist between different customers and the products that they purchase. For the analysis, Principal Component Analysis (PCA) is used for dimensionality reduction and k-Means Clustering, an unsupervised learning technique is used for customer categorization. By this, the end customer demands are tracked downstream from upstream in the supply chain, which may help to reduce the demand variation.

**Keywords**

Bullwhip Effect, Supply Chain, Machine Learning, Customer Categorization, PCA, k-Means Clustering,

Unsupervised Learning.

# Introduction

Forecasting the demand is a big challenge for most of the Supply chain industries (Wieland & Wallenburg, 2011). The traditional demand forecasting methods include Cumulative forecasting, Naïve forecasting and Exponential smoothing technique (Holt, 2004). These forecasting techniques may also include the level, seasonality and trend in the demand distribution for better prediction of the results (Gamberini, Lolli, Rimini, & Sgarbossa, 2010). In previous studies, Regression technique has also been widely used for demand forecasting (Brazdil, Carrier, Soares, & Vilalta, 2008; Freedman, 2009; Levis & Papageorgiou, 2005). However, volatility in demand distribution makes it harder to predict the demand pattern. The volatility starts to increase when moving upstream in the supply chain i.e. from the end customers to the retailers, retailers to the distributers, distributers to the manufacturer and from the manufacturer to the suppliers. This effect is called the Bullwhip effect in the supply chain. This concept was initially discussed by Forrester (1997), thus it is also known as Forrester effect, which makes the demand pattern more volatile while moving upstream in supply chain. Although there are many techniques proposed to overcome the Bullwhip effect, it is still a big challenge to accurately predict the demand as the variability of the end customer demand starts to pile up when moving upstream in the supply chain. One way to overcome this problem would be reducing the variability in the demand distribution, which could ultimately produce a better forecasting result. In the supply chain, eliminating the multiple forecasts, that only use the immediate partner order data in implementing the collaborative forecasting (McCarthy & Golicic, 2002), will reduce the Bullwhip effect. A customer-centric approach of predicting, “what the end customers need” and “when they need it” would avoid multiple forecasting and thus reduces the variability in the demand distribution.

# Methodology

In demand forecasting, the variance in demand distribution plays a critical role. It is always better to reduce the variance for improving the forecasting results. Particularly, for retail industries, the variance in demand of different products will be more when the aggregated demand forecasting is used, as the end customer demand is not properly traced (Jin, Williams, Tokar, & Waller, 2015). Thus, to avoid the aggregated demand forecasting, clustering technique, an unsupervised learning technique can be used. This technique will group customers into different clusters based on their purchasing behavior. However, this technique has limitations of determining optimum number of clusters and finding the initial cluster centers within each cluster. In addition, there will always be a trade-off between the number of clusters and the change in variance within each cluster. When there is no significant change in the variance while increasing the number of clusters, the clustering process is stopped, and the algorithm is converged. This can be done by using k-Means clustering (Lloyd, 1982). Further, the problem with initializing the cluster centers can be overcome by using k-Means++ algorithm (Arthur & Vassilvitskii, 2006) which helps this study to group the customers into different clusters*.* The clustering technique will also help in the following (Burton, Shore, & Buck, 1983):

* Vector quantization that finds a finite set of representatives, which provides coverage of a complex, possibly infinite, high-dimensional space.
* Finding meaningful structures in data and finding salient grouping in data.

When these individual clusters are analyzed and their distributions are studied, the variance associated within each cluster will show a possible improvement in comparison with the overall aggregated data. When these clusters with reduced variances are considered for forecasting, the loss function associated with the model can be reduced (Chu, Keerthi, & Ong, 2004).

In summary, this study follows the steps below:

1. Clustering: The customers were categorized into different clusters using k-Means clustering.
2. Evaluation: The purchasing behavior of each cluster of customers is evaluated by calculating the average number of orders and average number of days taken by the customers in each cluster to reorder the products. These averages can be used to predict “what the end customers in each cluster need” and “when they need it”.

**Analysis and Results**

This study used the data from “Instacart - market basket” dataset (Kaggle, 2017) for analysis. This dataset contains data on customer orders over time for previously purchased orders from Instacart, which is a grocery ordering and delivery application. The data variables are shown in Exhibit 1 and a sample dataset is shown in Exhibit 2.

**Exhibit 1. Data Description**.

|  |  |
| --- | --- |
| **Columns** | **Definitions** |
| Order\_id (*OI*) | Unique identity number for each order |
| Product\_id (*PI*) | Unique identity number for each product |
| Add\_to\_cart\_order (*ATCO*) | Order in which each product is added to the cart |
| Reordered (*RO*) | Boolean for reordered items |
| Product\_name (*PN*) | Name of the Product |
| Aisle\_id (*AI*) | Unique identity number for each aisle |
| User\_id (*UI*) | Unique identity number for each user |
| Order\_number (*ON*) | Orders from each customer |
| Order\_dow (*DOW*) | Determines the days of the week |
| Order\_hour\_of\_the\_day (*HOD*) | The time when the order is placed |
| Days-since\_prior\_order (*DSPO*) | Orders from each customer  (ranges between 4 and 100) |
| Aisle (*aisle*) | Product Categories |

## Exhibit 2. Sample Dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *OI* | *PI* | *ATCO* | *RO* | *PN* | *AI* | *UI* | *ON* | *DOW* | *HOD* | *DSPO* | *aisle* |
| 2 | 33120 | 1 | 1 | Organic Egg Whites | 86 | 202279 | 3 | 5 | 9 | 8 | eggs |
| 26 | 33120 | 5 | 0 | Organic Egg Whites | 86 | 153404 | 2 | 0 | 16 | 7 | eggs |
| 120 | 33120 | 13 | 0 | Organic Egg Whites | 86 | 23750 | 11 | 6 | 8 | 10 | eggs |
| 327 | 33120 | 5 | 1 | Organic Egg Whites | 86 | 58707 | 21 | 6 | 9 | 8 | eggs |
| 390 | 33120 | 28 | 1 | Organic Egg Whites | 86 | 166654 | 48 | 0 | 12 | 9 | eggs |

To categorize the customers into different clusters, firstly cross-table between the users and the aisles are formulated. A sample of the cross-table data is shown in Exhibit 3. Then, Principal Component Analysis (PCA) is used to reduce the dimension of the dataset. In addition, PCA also reduces the storage and computational time, reduces the redundancy from the high-dimensional data and removes noise or irrelevant features (Bingham & Mannila, 2001). In the PCA, the scree plot is used to determine the number of principal components to retain (Zhu & Ghodsi, 2006). In this study, six principal components are retained for further analysis as the scree plot in Exhibit 4 and Exhibit 5 shows that the six principal components explain around 70% of the variance in the data.

## Exhibit 3 Sample Data in Cross-Table.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *UI* | *aisle* | | | | | | |
| tortillas  flat bread | trail mix snack mix | trash bags  liners | vitamins  supplements | water seltzer  sparkling water | white  wines | yogurt |
| 206205 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 206206 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 206207 | 2 | 1 | 0 | 0 | 11 | 0 | 15 |
| 206208 | 7 | 0 | 0 | 0 | 0 | 0 | 33 |
| 206209 | 0 | 0 | 1 | 0 | 0 | 0 | 3 |

## Exhibit 4. Scree Plot.



## Exhibit 5. Variance Explained by Principal Components.

|  |  |  |
| --- | --- | --- |
| **Principal**  **Components** | **Variance** | **Cumulative**  **Variance** |
| **PC1** | 48.24 | 48.24 |
| **PC2** | 9.59 | 57.82 |
| **PC3** | 5.19 | 63.01 |
| **PC4** | 3.59 | 66.60 |
| **PC5** | 2.93 | 69.53 |
| **PC6** | 2.39 | 71.93 |

After determining the number of principal components, Elbow method is used to determine the optimum number of clusters (k) to be used for k-Means clustering (Bholowalia & Kumar, 2014). The Elbow plot in Exhibit 6 shows that k is four, which is the elbow point in the Elbow plot, for further analysis. Then, k-Means clustering is used to group customers with similar purchasing behavior. Exhibit 7 shows the plot of the clusters and their centers considering only the two principal components PC1 and PC2.

## Exhibit 6. Elbow Method for Optimal k.



## Exhibit 7. Cluster Assignments and Centroids.

## 

To analyze the purchasing behavior of the customers in each cluster, the average order count (*OC*) i.e. the average number of orders in each aisle is calculated. In addition, average reorder frequency (*RF*) i.e. the average number of days taken by the customer to reorder the products in each aisle, also known as lead time in supply chain analysis (Kumar, 1989), is calculated for each cluster. Exhibit 8 shows the ten aisles with the most *OC* for each cluster. In addition, Exhibit 8 also shows the *RF* for the corresponding aisle in each cluster.

## Exhibit 8. Average Order Count (OC) and Reorder Frequency (RF) in Different Clusters.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cluster 1** | | | **Cluster 2** | | | **Cluster 3** | | | **Cluster 4** | | |
| *Aisle* | *OC* | *RF* | *Aisle* | *OC* | *RF* | *Aisle* | *OC* | *RF* | *Aisle* | *OC* | *RF* |
| fresh fruits | 5.85 | 15.63 | fresh fruits | 155.74 | 5.82 | fresh fruits | 29.63 | 11.59 | fresh fruits | 77.06 | 8.01 |
| fresh  vegetables | 5.28 | 16.45 | fresh  vegetables | 154.94 | 6.03 | fresh  vegetables | 29.20 | 12.36 | fresh  vegetables | 68.05 | 8.52 |
| packaged  vegetables  fruits | 3.15 | 15.83 | packaged  vegetables  fruits | 66.94 | 5.92 | packaged  vegetables  fruits | 14.84 | 11.78 | packaged  vegetables  fruits | 34.66 | 8.17 |
| yogurt | 2.61 | 15.71 | yogurt | 50.95 | 5.95 | yogurt | 11.92 | 11.64 | yogurt | 30.29 | 8.02 |
| water seltzer  sparkling  water | 2.27 | 14.77 | Packaged  cheese | 30.89 | 6.02 | packaged  cheese | 8.07 | 11.71 | packaged  cheese | 18.02 | 8.21 |
| packaged  cheese | 2.05 | 15.84 | milk | 29.63 | 5.71 | milk | 7.13 | 10.76 | milk | 17.19 | 7.57 |
| milk | 1.81 | 14.71 | s | 19.64 | 5.80 | water  seltzer  sparkling  water | 6.80 | 10.48 | water seltzer  sparkling  water | 13.01 | 7.42 |
| chips pretzels | 1.78 | 15.14 | bread | 18.12 | 5.75 | chips pretzels | 5.94 | 10.98 | chips pretzels | 11.83 | 7.79 |
| ice cream ice | 1.39 | 14.60 | baby food  formula | 18.09 | 5.98 | soy  lactosefree | 5.42 | 11.42 | soy  lactosefree | 11.76 | 7.84 |
| soft drinks | 1.32 | 14.24 | chips pretzels | 16.97 | 5.81 | refrigerated | 4.78 | 10.78 | bread | 10.61 | 7.86 |

Exhibit 8 shows that the fresh fruits aisle has the highest OC in each cluster. The customers in Cluster 1 ordered the products from fresh fruits aisle 5.85 times on average and the average reordering frequency was 15.63 days. However, the customers in Cluster 2 ordered the products from fresh fruits aisle 155.74 times on average and reordered from the same aisle every 5.82 days on average. This shows that PCA and k-Means clustering techniques can be used to evaluate the information regarding purchasing behavior of the customers in different clusters. Such information can be useful for any company to set up their review policies like periodic review policy or continuous review policy (Jain, 1999) and to set up their safety stock (Monk & Wagner, 2012) according to the level of desired service quality.

**Conclusion**

This study used PCA and k-Means clustering techniques to group customers into different clusters based on their purchasing behavior. After the grouping, different purchasing behavior metrics such as average order count (*OC*) and average reorder frequency (*RF*) are evaluated for each cluster for different products. Such evaluation helps to track the end customer demands for any products, downstream from upstream in the supply chain. This may eventually help to reduce the demand variation for each product. In addition, this evaluation may also help to establish the replenishment policies and to determine the economic order quantity, which may further help to reduce the Bullwhip effect that occurs due to order batching. Overall, this approach helps to counteract the Bullwhip effect that may occur during demand forecasting. However, the analytical approach used in this study is not applicable to forecast the demand of products for which previous sales data is unavailable.

**References**

Arthur, D., & Vassilvitskii, S. (2006). *k-means++: The advantages of careful seeding*. Stanford.

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications, 105*(9).

Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: applications to image and text data*.* *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*.

Brazdil, P., Carrier, C. G., Soares, C., & Vilalta, R. (2008). *Metalearning: Applications to data mining*: Springer Science & Business Media.

Burton, D., Shore, J., & Buck, J. (1983). A generalization of isolated word recognition using vector quantization*.* *IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Chu, W., Keerthi, S. S., & Ong, C. J. (2004). Bayesian support vector regression using a unified loss function. *IEEE transactions on neural networks, 15*(1), 29-44.

Forrester, J. W. (1997). Industrial dynamics. *Journal of the Operational Research Society, 48*(10), 1037-1041.

Freedman, D. A. (2009). *Statistical models: theory and practice*: cambridge university press.

Gamberini, R., Lolli, F., Rimini, B., & Sgarbossa, F. (2010). Forecasting of sporadic demand patterns with seasonality and trend components: an empirical comparison between Holt-Winters and (S) ARIMA methods. *Mathematical Problems in Engineering, 2010*.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting, 20*(1), 5-10.

Jain, P. (1999). *Theory and problems in financial management*: Tata McGraw-Hill Education.

Jin, Y. H., Williams, B. D., Tokar, T., & Waller, M. A. (2015). Forecasting with temporally aggregated demand signals in a retail supply chain. *Journal of Business Logistics, 36*(2), 199-211.

Kaggle. (2017). Instacart Market Basket Analysis. Retrieved from *https://www.kaggle.com/c/instacart-market-basket-analysis*.

Kumar, A. (1989). Component inventory costs in an assembly problem with uncertain supplier lead-times. *IIE transactions, 21*(2), 112-121.

Levis, A., & Papageorgiou, L. (2005). Customer demand forecasting via support vector regression analysis. *Chemical Engineering Research and Design, 83*(8), 1009-1018.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory, 28*(2), 129-137.

McCarthy, T. M., & Golicic, S. L. (2002). Implementing collaborative forecasting to improve supply chain performance. *International Journal of Physical Distribution & Logistics Management*.

Monk, E., & Wagner, B. (2012). *Concepts in enterprise resource planning*: Cengage Learning.

Wieland, A., & Wallenburg, C. (2011). Supply-chain-management in stürmischen Zeiten. Univ. *Verlag der TU*.

Zhu, M., & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis, 51*(2), 918-930.

**About the Authors**

**Shivahari Revathi Venkateswaran** is a Master's student in Engineering Management at Northeastern University, Boston, MA. He received his Bachelors of Engineering from Anna University, Chennai in 2016. He has completed a Micro Masters in Supply Chain Management from Massachusetts Institute of Technology and a Micro Masters in Data Science from the University of California San Diego. Currently, he is working as a Business Operations Strategy Co-op/Intern at Vertex Pharmaceuticals, Boston, MA. His research interests include Applied Machine Learning for driving useful and meaningful insights in Supply Chain Management and using Operational Research for Optimizing the results produced.

**Himlona Palikhe** is an Associate Teaching Professor in the Mechanical and Industrial Engineering Department at Northeastern University. She also serves as the Co-Program Director for the MS in Engineering Management Program at Northeastern University. She received her Ph.D. in Systems and Engineering Management from Texas Tech University in 2013. Her research interests include engineering management, economic modeling, quality management, cost of quality and electric utilities.

**Manish Ranjit** is an Assistant Teaching Professor in the Mechanical and Industrial Engineering Department at Northeastern University, Boston, MA. He received his Ph.D. degree in Industrial Engineering (2017) and the M.S. degree in Electrical Engineering (2010) from Texas Tech University, Lubbock, TX. His research interests are in the areas of operations research, biomechanics, human factors and ergonomics.