

## The Method of Finding Potentially Churning Customers Based on Social Networks

Wanqiu Huang<sup>1</sup>, Xuguang Jia<sup>2</sup>, Fen Tian<sup>1</sup>, Yu Zhang<sup>1</sup> and Zhe Zhou<sup>1</sup>

<sup>1</sup>Beijing Technology and Business University Beijing, P.R. China

<sup>2</sup>TravelSky Technology Limited, P.R. China

[huangwq@th.btbu.edu.cn](mailto:huangwq@th.btbu.edu.cn)

### Abstract

*Customer churn analysis has become an important focus of corporate marketing. It will be a great help to profitability if there is a method can find losing customers in time. In the paper, a method based on RFM and Cross-correlation model is proposed. Firstly, the customer's value is calculated by RFM. Secondly, the typical losing curves of customer value are matched via cross-correlation. And finally, integrated with social network analysis (SNA) and community detection, the group of potential losing customers are revealed. The effectiveness of the presented method has been proven in a dataset of retail sales records.*

**Keywords:** SNA; Churn Analysis; Cross-correlation; Community Detection

### 1. Introduction

Customer resource is the lifeline of the enterprise now. In essence, the competition is exercised by providing products or services in order to compete for customers in the market. Therefore, the primary issue of enterprise-customer relationship management is how to attract new customers as well as to retain old customers. Moreover, prevention of the customer churn and timely analysis of the reasons for the loss is of more significance. According to the research of Reichheld F F [1], the cost of the service industry to attract a new customer is 4 to 6 times higher than to keep an old customer. If the retention rate of customers increase by 5%, the profits of enterprises will be increasing by from 25% to 85%. Discovering the losing or potential losing of individual and customer group, to take positive measures to retain them, to analyze the reason of the loss and to perfect the products and services play an important role in improving the efficiency of enterprises.

Over more than a decade of development, people usually take the customer churn prediction as a classification problem. The main research methods are: Decision Tree Classifiers, Logistic Regression, Bayesian Classifier, Cluster Analysis, Artificial Neural Network (ANN), Self Organizing Map (SOM), Support Vector Machine (SVM), etc.[2-8] Classification of the methods above is based on several characteristic attributes, although ANN, SOM, SVM method is more accurate, yet the classification results are difficult to explain.

What's more, these methods take customers as isolated individuals to study without considering the interaction and information exchange between customers. There is an increasing numbers of consumers who would prefer to take the rating&review from purchased customers or the friends they have on those newly emerging social medium such as Wechat or Weibo (Microblog) as reference during the process of their purchase decision. The enterprises should have a penetrating insight to Figure out if a highly influential customer no longer buy a brand, or even produce such effects on his surrounding customers to further bring about a customer churn phenomenon.

Based on the mentioned problems, this paper presents a simple effective method for analyzing the customer churn: Based on the value curve of customer and cross-correlation function, we can quickly identify the churn customers. The customer's consumption behavior is quantified by time period into a time-dependent value, therefore, the dynamic process of customer churn is taken into consideration in this method. Also, it applies such social network analysis methods as community detection and impact analysis *etc.* and is combined with the customer relationship network (generated according to a joint purchase records )to extract the potential losing customers.

The proposed method not only helps to find the churn of individual quickly, but also further gives the potential losing customer group .The effectiveness of the presented method has been proven in a dataset of retail sales records.

## 2. Method to Find Churn Customer Group Based on Cross-Correlation Model

The method presented to find losing customer group is consists of three steps: RFM customer value calculation, cross-correlation model matching, community detection.

### 2.1. RFM Customer Value Calculation

Define customer value. When a customer is given, there will be many indicators to indicate whether the customer is in the process of losing or potential loss. These features can be either static or dynamic. For instance, customer's attention, purchase quantity and purchase frequency of a certain brand declines in a period of time. Quantify this kind of standard and we can draw the curve of customer value change.

The RFM model [9] is a widely used customer value analysis model, in which, R (Recency) represents the most recent consumption time, F (Frequency) represents the consumption frequency, M (Monetary) represents the amount of consumption. These three indicators which are of great distinguishability to quantify and differentiate between high or low value customer.

By multiply or calculating the weighted sum of the three RFM indicators ,customer value can be obtained. It's usually expressed in the form of a standardized value from 0 to 100. As shown in Figure 1, the value curve of a high value customers maintain at 96. Figure 2 shows the value curve of churn customer, during the 7 statistical time period, the customer value declines undulately by 70 to 0. This indicates that the purchase ability, profitability, loyalty of this customer in the given brand declines so that appropriate marketing projects need to be exercised to retain.

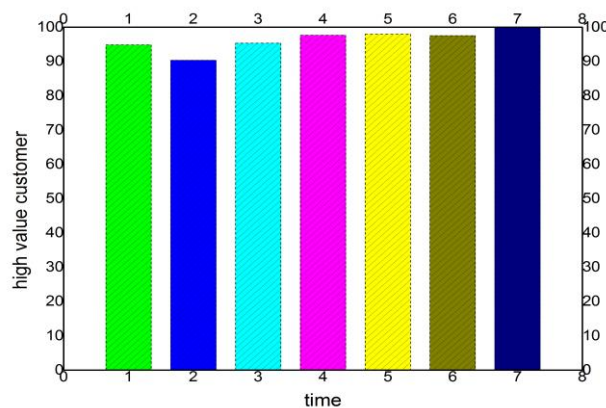
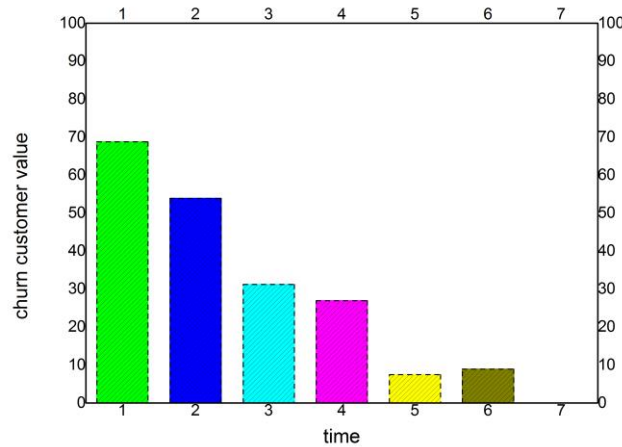


Figure 1. Value Curve of a High Value Customer



**Figure 2. Value Curve of a Churn Customer**

## 2.2. Correlation Model Matching

The churn discriminant algorithm based on cross-correlation function -- which is widely used in time series analysis, what follows below is its discrete form:

$$(f * g)[n] \triangleq \frac{1}{N} \sum_{m=-\infty}^{\infty} \frac{f^*[m]g[n+m]}{\sigma_f \sigma_g} \quad (1)$$

$f[n]$  and  $g[n]$  stand for two time series,  $\sigma_f$  and  $\sigma_g$  respectively stand for standard deviation of  $f[n]$  and  $g[n]$ .

If the two time series are completely identical or similar in shape (that is, they can overlap through translation), then we find the cross-correlation function for a maximum value of 1.

## 2.3. Community Detection

The customers are not isolated individuals, but the formation of relationship network. As shown in Figure 3, which is graphed by Gephi [10]. The points represent individual customer in the graph, the size of the dot represents customer value; the connecting edge of two nodes represent a joint consumption records of two clients, and note the weights of edges as joint consumption frequency which represents relationship strength.

Through the application of community detection method, closely connected individuals can be classified into the same group. Here gives community discovery algorithms based on graph: Image Segmentation Algorithm, Hierarchical Clustering Algorithm, Modular Function based-on Algorithm, the Dynamic Algorithm, etc. [11] This essay adopt modular function based-on algorithm proposed by the Blondel V D to exercise community detection fastly [12]. The computational time complexity of this method is  $O(m)$ . According to this method, Figure 3 can be classified into 4 communities, which has been marked in different color.

According to the cross-correlation model, after matching the potential churn customer, we substitute them into the community detection results to find out the other nodes which belong to the same community, then set the filter conditions such as relationship strength, to obtain potential losing customer.



**Figure 3. Customers Social Relation Network**

The customer churn discriminant algorithm can be summarized as the following steps:

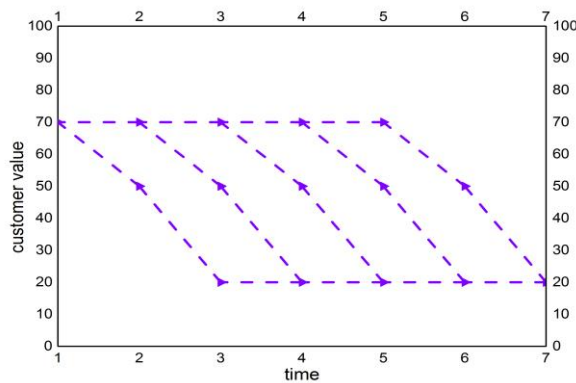
- (1) According to the RFM model, calculate the customer value curve  $f$  over a period of time;
- (2) According to the clustering method, define the typical erosion sample curve family  $g_1, g_2, \dots, g_n$ ;
- (3) Calculate  $(f * g_i) \in [0, 1], i = 1, 2, \dots, n$ , find the maximum value, denoted by  $C_{\max}$ ;
- (4) If  $C_{\max} \geq$  threshold 0.9, then the customer is defined as a match of the typical churn model and need to be marked as potential churn object.
- (5) Substitute the community division results, find out the other customers belong to the same community, set filter condition according to relationship strength and extract the relevant customer group.

The typical erosion sample curve family  $g_i$  can be figured out according to the sample set K-means clustering or manually set by industry expert. Due to the multiple forms of value curve of churn samples may have, it should be matched to several curves in order to reduce the missing analysis.

### 3. Simulation Results and Analysis

Take a retail enterprise for example, here shows 3.5 years of its customer consumption records, including 102299 individual customers, each one of them is identified by a sole number. The record shows specific information of date and the amount of purchase .

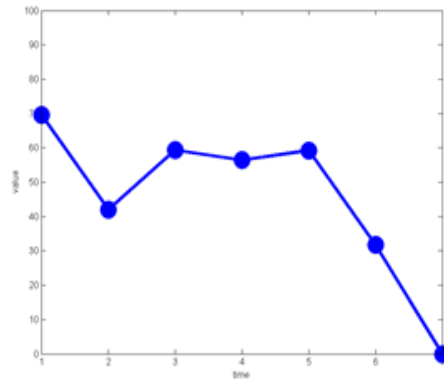
As is shown in Figure 4, we choose 0.5 years as the time window length, then calculate customer value in each segment according to the RFM model, finally define the curves of typical churn customer.



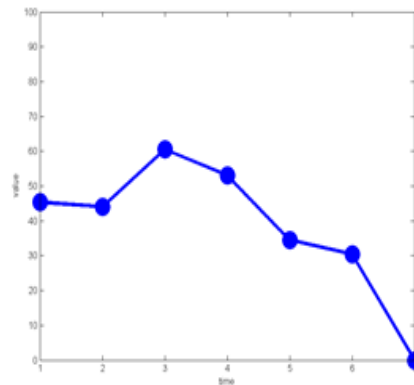
**Figure 4. Curves of Typical Churn Customers**

Substitute what we had into the Churn Analysis Algorithm, we can draw a conclusion that the on-going customer churn number is 4716, customer churn rate is about 4.61%.

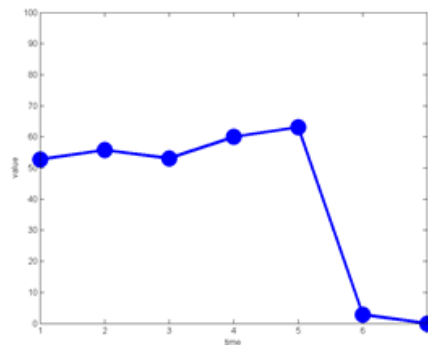
Take some of them, customer A~C as an example, their value curve has shown below in Figure 5 to 7. The value of A customers is relatively stable in the first 5 time periods, while it gradually decline to zero in the sixth, seventh time period; the value of customer B continues to decline in third to seventh time period; similar to customer A, customer value of C during first to fifth time period stably remained at about 50, while suddenly drop to 0 in the sixth time period.



**Figure 5. Value Curve of Churn Customer A**

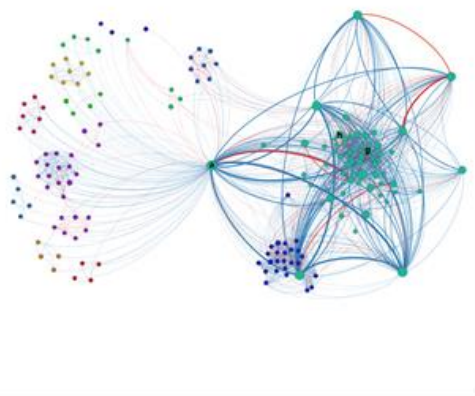


**Figure 6. Value Curve of Churn Customer B**

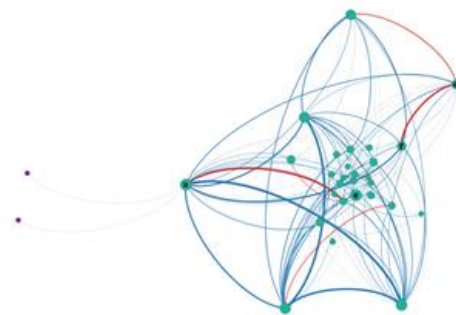


**Figure 7. Value Curve of Churn Customer C**

Take Customer A's one-hop neighbor network as an example, as shown in Figure 8, there's 160 nodes (including A) and 1640 edges included in the network. They respectively belong to 13 communities, marked with different color. Set the edge weight threshold to 3.0, after the filtration, the remaining 29 nodes (including A) and 162 edges are respectively belong to 2 community as shown in Figure 9. It also contains 5 customers which are regarded as churn customers according to the cross-correlation function. Moreover, based on the algorithm definition of this essay,  $28-5=23$  these neighbors of customer A, is the potential churn customers.

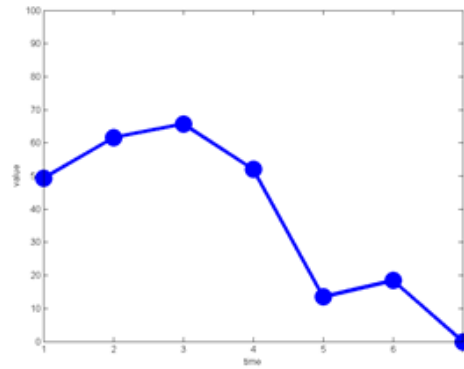


**Figure 8. 1-Hop Neighbors Network of Customer A**

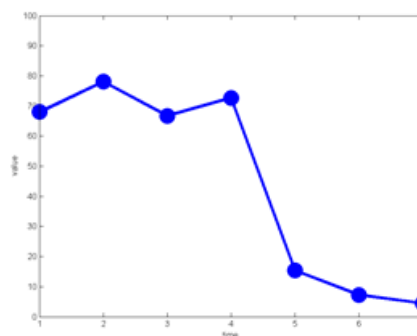


**Figure 9. 1-Hop Neighbors Network of Customer A Filtered by Edge Weight**

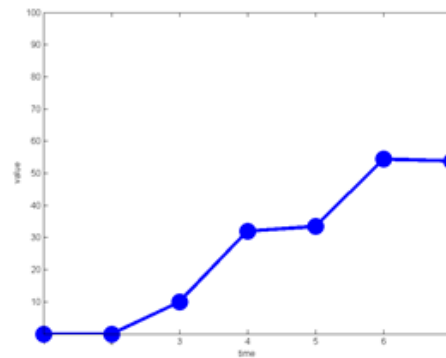
Now we select several high-value neighbors from customer A's potential churn customers neighbors, verify whether the churn existed from the value curve. Customer A's neighbors are classified into three categories: 1. found by the cross-correlation curve such as customer D (Figure 10) and other 5 customers; obtained by filtered edge weight such as customer E, F (Figure 11); removed by filtered edge weight such as customer G, H (Figure 12).



**Figure 10. Value Curve of Churn Customer D**



**Figure 11. Value Curve of Churn Customer E**



**Figure 12. Value Curve of Churn Customer G**

Customer D's value curve is in accordance with the typical churn model. Customer E meets the churn characteristic, but failed to match with cross-correlation function. The customer F does not meet the definition of churn model presented in this essay, but this curve begin to show a declined trend from the fifth time period , which is worth paying certain attention to. According to the algorithm presented above, customer churn does not occur in normal customer G,H, from which the validity of the algorithm is confirmed from another perspective.

## 5. Conclusion

In this paper, we propose a group loss analysis algorithm based on cross-correlation model and social network analysis. Through this method, we will be able to find the losing customers, and further to locate the potential customers churn. The effectiveness of the presented algorithm has been proven in the datasets of a retail sales records, the result is relatively ideal. At the same time, the process of method is comparatively simple and the results have strong interpretability.

## Acknowledgments

This work was supported by The General Program of Beijing Municipal Education Committee, P. R. China 2015 (Research on growth mechanism of Zhongguancun industrial cluster based on social networks) (NO. SM201510011006), Beijing philosophy and social science planning projects, P.R.China (NO. 11JGB028) research result of stage.

## References

- [1] Rechinheld F. and Sasser W., "Zero defections: Quality comes to service", Harvard Business Review, vol. 68, no. 5, (1990), pp. 105-111.
- [2] G. Xia, "Research status and development of customer churn prediction", Application research of computers, vol. 27, no. 2, (2010), pp. 413-416.
- [3] G. Xia and Y. Chen, "Telecom Customer Churn Prediction Model", Statistics and decision making, vol. 20, (2006), pp. 163-165.
- [4] G. Jiang and X. Si, "Study of telecom customer churn prediction based on cost sensitive SVM", Application research of computers, vol. 26, no. 2, (2009), pp. 521-523.
- [5] Y. Zhao, B. Li, X. Li and X. Si, "Research of customer churn analysis based on support vector machine improvement", Computer integrated manufacturing system, vol. 13, no. 1, (2007), pp. 202-207.
- [6] M. Guo, H. L. Zheng and Y. W. Lu, "The customer churn analysis based on bayesian network", Journal of Nanjing University of posts and telecommunications natural science, vol. 25, no. 5, (2006), pp. 79-83.
- [7] J. Ye and S. Lin, "The application of inference in mobile customer churn analysis based on bayesian network", Journal of computer application, vol. 25, no. 3, (2005), pp. 673-675.
- [8] L. Tian, H. Qiu and L. Zheng, "Telecom Customer Churn prediction modeling and implementation of based on Neural Network", Journal of computer application, vol. 27, no. 9, (2007), pp. 2294-2297.
- [9] Miglautsch J. R., "Thoughts on RFM scoring", The Journal of Database Marketing, vol. 8, no. 1, pp. 67-72. [5]
- [10] Gephi. <http://gephi.org/>
- [11] Fortunato S., "Community detection in graphs", Physics Reports, vol. 486, no. 3, (2010), pp. 75-174.
- [12] Blondel V. D., Guillaume J. L. and Lambiotte R., "Fast unfolding of communities in large networks", Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, (2008), pp. 10008.

## Authors



**Wanqiu Huang**, received her M.S. degree in computer science from Beijing Technology and Business University. She is currently a lecturer in Beijing Technology and Business University. Research area include Computer Application, network, economic management and social network analysis.





**Xuguang Jia**, received his M.S. degree in computer science from Beijing Jiaotong University. He is currently a engineer in TravelSky Technology Limited. Research area include Data Ware and Data Mining , Computer Application Technology, network, and social network analysis.



**Zhe Zhou**, received his B.S. degree in management from Beijing Technology and Business University. He is currently a experimentalist in Beijing Technology and Business University. Research area include economic management.

