Predicting Customer Profitability Dynamically over Time: An Experimental Comparative Study

Daqing Chen[1](✉), Kun Guo[1], and Bo Li[2]

[1] School of Engineering, London South Bank University, London SE1 0AA, UK
{chend, guok}@lsbu.ac.uk
[2] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
libo803@nwpu.edu.cn

Abstract. In this paper a comparative study is presented on dynamic prediction of customer profitability over time. Customer profitability is measured by Recency, Frequency, and Monetary (RFM) model. A real transactional data set collected from a UK-based retail is examined in the analysis, and a monthly RFM time series for each customer of the business has been generated accordingly. At each time point, the customers can be segmented by using the *k*-means clustering into high, medium, or low groups based on their RFM values. Twelve different models of three types have been utilized to predict how a customer's membership in terms of profitability group would evolve over time, including regression, multilayer perceptron, and Naïve Bayesian models in open-loop and closed-loop modes. The experimental results have demonstrated a good, consistent and interpretable predictability of the RFM time series of interest.

Keywords: Time series analysis · RFM model · CRM · Predictive modelling

1 Introduction

Over the last decades marketing has made a significant shift from traditional product/brand based to customer-centric and data-driven by intensively using analytical models and tools. One of the important aspects of applying analytics in marketing is to predict customer profitability over time based on customer purchasing history and a certain profitability measure, such as customer life-time value (CLV), and recency, frequency, and monetary (RFM) values. With regard to modelling techniques, there are mainly two categories of models [1]: probabilistic models and machine learning models. A fundamental question to be asked here is that if a customer's profitability is predictable, and what models can be best suitable a given prediction problem [2]. The primary aim of this research is to provide a case study for such a prediction.

In this paper, a UK-based online retail is examined for customer profitability prediction. A real transactional data set collected from the retail is used for the analysis. The RFM values of each customer are employed as a profitability measure, and an associated monthly RFM time series for every customer can be created accordingly

© Springer Nature Switzerland AG 2019
I. Nyström et al. (Eds.): CIARP 2019, LNCS 11896, pp. 174–183, 2019.
https://doi.org/10.1007/978-3-030-33904-3_16

based on their historical purchasing records. By using the *k*-means clustering, at any given time point, all the customers are segmented into three groups based on their RFM values: high, medium, or low profitability groups. The prediction problem concerned here is how a customer's membership in terms of profitability group would evolve over time. For comparison purpose, twelve different models of three types are utilized for prediction including both probabilistic and machine learning models in open-loop and closed-loop modes: regression, multilayer perceptron (MLP), and Naïve Bayesian models. Choosing RFM-based measure is because of its simplicity and easy interpretability in practice, and the models selected are classic, simple and widely used in business for marketing purpose.

A comparative analysis with the given data set and the models has demonstrated a good predictability of the chosen measure for the business under consideration in terms of customer profitability. It also shows how to use the certain context of the business to help to interpret the modelling outcomes.

The remainder of this paper is organized as follows. Section 2 gives a brief discussion on the relevant work. Sections 3 describes in detail the methodology adopted in this work including the creation of RFM-based time series, customers grouping and model selection. Detailed experimental settings and the experimental results are provided in Sect. 4, and a discussion on the outcomes is given in Sect. 5. Finally concluding remarks are given in Sect. 6 along with suggested further work.

2 Related Works

In recent years, predicting customer's profitability over time has been an active, yet very challenging research topic. In general, such a prediction mainly involve three interrelated factors:

- The nature of the business under consideration;
- Which measure(s) to be used to indicate a customer's profitability; and
- Which models to be employed to best fit the modelling requirements.

The nature of the business under consideration will be directly linked to what measures could be adopted, for example, an on-line business in the retail industry and a marketing consultancy company in the fashion industry may use a completely different set of measures. On the other hand, depending on the measures to be adopted, a static or a dynamic model could be applied for modelling purpose.

In [3] an RFM score-based time series was created using the *k*-means clustering analysis and used to measure and describe a customer's profitability for an on-line retail. Furthermore, multilayer feedforward neural network models were trained to identify the dynamics in terms of how customer profitability evolved over time.

Interestingly in [4], RFM was employed to calculate customer loyalty and Apriori algorithm was used to determine the association rules of product bundles. In addition, the work in [5] suggested convolutional neural network structures for predicting the CLV of individual players of video games, and in [6], recurrent neural networks were proposed for customer behavior prediction based on the client loyalty number and RFM values.

Other measures have been also considered, such as Pareto/NBD (negative binomial distribution) [7].

In summary, the main work in this area appears to be subject/domain-specific and has no unified approaches. CLV and RFM are the most popular measures adopted to reflect customer profitability/loyalty. The most diverse aspect of the relevant research is on modelling approaches, and a range of models have been proposed, from very classic regression models to deep learning paradigms.

This research presents a case study for customer profitability prediction in which multiple models are used with a simple yet practically easy-to-implement profitability measure.

3 Methodology

This Section gives in detail the main approaches, models, and procedures adopted in this research.

3.1 Recency, Frequency, and Monetary Model

The RFM model [8] has received much attention and has been widely used in customer relationship management (CRM) and direct marketing due to its simplicity and effectiveness for evaluating a customer's profitability.

Given a set of transactional records of a business over a certain period of time, Recency indicates how recently a customer made a purchase with the business; Frequency shows how often a customer has purchased; and Monetary indicates the total (or average) a customer has spent. Therefore, each customer of a business can characterized by a set of RFM values, and furthermore all the customers can be grouped into meaningful segments based on their RFM values so that various marketing strategies can be adopted to different customer groups accordingly.

Note that a time series of RFM values can be generated for each customer if they are calculated at consecutive time points, such as at the end of each calendar month over a period of time.

3.2 *k*-Means Clustering

k-means clustering is one of the most popular algorithms in data mining for grouping samples into a certain number of groups (clusters) based on Euclidean distance measure. Assume $V_1, V_2, \ldots, V_n$ are a set of vectors, for instance, a vector represents a customer's RFM values in the form a vector, and these vectors are to be assigned to k clusters $S_1, S_2, \cdots, S_k$. Then the objective function of the k-means clustering is expressed as

$$f(\mu_1, \mu_2, \cdots, \mu_k) = \sum_{i=1}^{k} \sum_{V_j \in S_i} \left\| V_j - \mu_i \right\|^2 \tag{1}$$

where μ_i represents the centroid of cluster S_i. The k-means clustering algorithm in the form of pseudocode is shown in Table 1.

Table 1. The k-means clustering algorithm.

Step 0: Initialise the centroids of k clusters $\mu_1, \mu_2, \cdots, \mu_k$
Step 1: Assign V_j $(1 \le j \le n)$ to cluster S_i, if $\left\|V_j - \mu_i\right\|^2 \le \left\|V_j - \mu_n\right\|^2$ $(1 \le n \le k)$
Step 2: Update centroids $\mu_1, \mu_2, \cdots, \mu_k$ using $\mu_i = \frac{1}{m}\sum_{V_j \epsilon S_i} V_j$, where m represents the number of vectors in a cluster
Step 3: Stop if the centroids $\mu_1, \mu_2, \cdots, \mu_k$ remain unchanged; Otherwise, go back to Step 1.

In this paper, a group of customers are segmented into three segments using the k-means clustering based on their RFM values: low, medium, or high profitability groups.

3.3 Open-Loop Model and Closed-Loop Model for Time Series Prediction

Time series prediction can be in general formalized by open-loop and closed-loop models. Given a time series $\{\theta(t)|t = 1, 2, \cdots n)\}$, a prediction based on an open-loop model is expressed as

$$\hat{\theta}(t) = f(\theta(t-1), \theta(t-2), \cdots, \theta(t-n)) \tag{2}$$

where $f(\cdot)$ donates a mapping, and $\hat{\theta}(t)$ represents the predicted value of variable $\theta(t)$ at time t using the prior n observed values of the variable at time points $t-1$, $t-2$, $\cdots$, $t-n$.

A closed-loop model can be expressed as

$$\hat{\theta}(t) = f\left(\hat{\theta}(t-1), \hat{\theta}(t-2), \cdots, \hat{\theta}(t-n)\right) \tag{3}$$

which uses the previous n predicted values $\hat{\theta}(t-1), \hat{\theta}(t-2), \cdots, \hat{\theta}(t-n)$ to predict to the value of variable $\theta(t)$ at time t.

3.4 Model Selection

The mapping $f(\cdot)$ in an open-loop or a closed model (Eqs. (2) and (3)) can be in different forms. In this paper, three models are considered for comparison purpose: Linear Regression, Multilayer Perceptron (MLP), and Naïve Bayesian.

Linear regression is perhaps the simplest model to be considered. Using this model for prediction, Eqs. (2) and (3) can be re-written, respectively, as

$$\hat{\theta}(t) = w_0 + \sum_{i=1}^{n} w_i \theta(t-i) \tag{4}$$

and

$$\hat{\theta}(t) = w_0 + \sum_{i=1}^{n} w_i \hat{\theta}(t-i) \tag{5}$$

where $\{w_i | i = 0, 1, \ldots n)\}$ are regression coefficients.

A multi-layer perceptron can be thought of as a regression model on a set of derived inputs via layered and successive non-linear transformations. In this paper, an MLP is used with a single hidden layer and a linear transformation for output nodes, which can be expressed as

$$h_j(t) = \frac{1}{1 + e^{-(w_{0j} + \sum_{i=1}^{n} w_{ij}\theta(t-i))}}, j = 1, 2 \cdots, m \tag{6}$$

$$\hat{\theta}_l(t) = w_{0l} + \sum_{j=1}^{m} w_{jl} h_j(t), l = 1, 2 \ldots, k \tag{7}$$

where w_{ij} and w_{jl} are connection weights between the ith input node to the jth hidden node, and the jth hidden node to the lth output node, respectively; w_{0j} and w_{0l} donate the bias to the jth hidden node and the bias to the lth output node, respectively; and $h_j(t)$ and $\widehat{\theta}_l(t)$ donate the output of the jth hidden node and the lth output node, respectively. For the closed-loop model the inputs $\{\theta(t-i)\}$ are substituted by $\left\{\hat{\theta}(t-i)\right\}$.

3.5 Naïve Bayesian Model

A Naïve Bayesian model is based on Bayes' theorem as shown below

$$p(A|B) = \frac{p(A, B)}{p(B)} = \frac{p(B|A)p(A)}{P(B)} \tag{8}$$

where $p(\cdot)$ and $p(\cdot|\cdot)$ represent a probability and a conditional probability, respectively. Applying Naïve Bayesian model, Eqs. (2) and (3) can be re-written and simplified as

$$\begin{aligned} p\left(\hat{\theta}(t) | \theta(t-1), \ldots, \theta(t-n)\right) &= \frac{p\left(\hat{\theta}(t), \theta(t-1), \ldots, \theta(t-n)\right)}{p(\theta(t-1), \ldots, \theta(t-n))} \\ \propto p\left(\hat{\theta}(t), \theta(t-1), \ldots, \theta(t-n)\right) &\approx p\left(\hat{\theta}(t)\right) \prod_{i=1}^{n} p(\theta(t-i)) \end{aligned} \tag{9}$$

$$p\left(\hat{\theta}(t)|\,\hat{\theta}(t-1),\ldots,\hat{\theta}(t-n)\right) = \frac{p\left(\hat{\theta}(t),\hat{\theta}(t-1),\ldots,\hat{\theta}(t-n)\right)}{p(\theta(t-1),\ldots,\theta(t-n))}$$
$$\propto p\left(\hat{\theta}(t),\hat{\theta}(t-1),\ldots,\hat{\theta}(t-n)\right) \approx p\left(\hat{\theta}(t)\right)\prod_{i=1}^{n} P\left(\hat{\theta}(t-i)\right) \tag{10}$$

4 Case Study

4.1 Data Set and Data Pre-processing

A UK-based online retail is considered in this study [3, 9]. A data set was collected from the retail which contains all the transactions occurring from December 2010 to November 2011. The data set has 11 variables as described in Table 2. Note that the data set can be found at: https://archive.ics.uci.edu/ml/datasets/online+retail.

It is worth mentioning that, over the years, the business has been functioning as both wholesale and retail, and has maintained a stable and healthy number of customers.

Table 2. Variables in the dataset.

Variable	Data type	Description; Typical values and meanings
Invoice	Nominal	Invoice number
StockCode	Nominal	Product (item) code
Description	Nominal	Product (item) name; CARD I LOVE HAVANA
Quantity	Numeric	Quantity of each product (item) per transaction
Price	Numeric	Product price per unit in sterling; £45.23
InvoiceDate	Numeric	Day and time when a transaction occurred; 23/08/2011 15:59
Address Line 1	Nominal	Delivery address line 1; 103 Borough Road
Address Line 2	Nominal	Delivery address line 2; Elephant and Castle
Address Line 3	Nominal	Delivery address line 3; London
PostCode	Nominal	Delivery address postcode; SE1 0AA
Country	Nominal	Delivery address country; England

Appropriate pre-processing was carried out to address quality issues of the data set. Outliers and extreme values have been removed. The resultant target data set contains 751 valid customers from the UK only.

4.2 Settings for Modelling

To start the analysis, a time series of RFM values for each customer was first calculated at the end of each calendar month successively from December 2010 to November 2011, and therefore each RFM time series consists of 12 data points.

Further at each time point of the monthly-based RFM time series, the customers were grouped using the *k*-means clustering into three profitability groups as shown in Fig. 1, where Recency is in month and Monetary is in Sterling, and symbols '*', '+', and 'o' indicate high, medium, and low profitability groups, respectively. The sub-graphs in Fig. 1 are arranged sequentially by month in ascending order. As such, each customer belongs to a certain profitability group at a given time point of the time series. Before conducting the clustering, the RFM values were normalised by using range normalisation.

Next, the three types of predictive models discussed in the previous section were applied to predict each customer's profitability group using open- and closed- loop models. The three profitability groups were encoded into three orthogonal unit vectors $[1, 0, 0]$, $[0, 1, 0]$ and $[0, 0, 1]$, and these vectors were used as the desired outputs of all

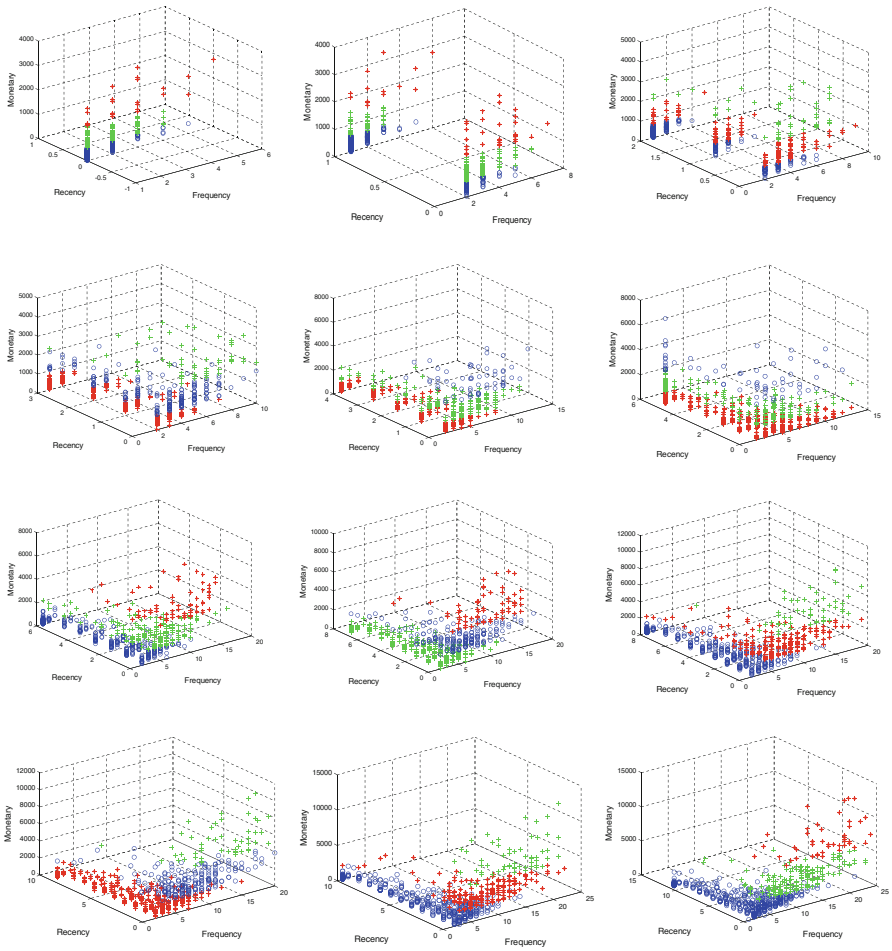

Fig. 1. Customers segmented into three profitability groups: high (*), medium (+), or low (o). Calculations were made at the end of each calendar month from Dec. 2010 to Nov. 2011.

the models for training to represent mutually exclusive three classes. Both the open- and closed- loop linear regression models had two or three terms. The topology of the MLP models were set to: three input nodes, ten hidden nodes and three output nodes. The initial connection weights and biases were generated randomly.

All the models were trained and tested 10 times, and each time 70% of the samples in the data set were randomly selected for training and the remaining 30% for testing. The data in December 2010 and January 2011 was used as the initial inputs for the closed-loop models. Note that, regardless what predictive models to be used, the training procedures for both the open-loop and closed-loop models are the same; However, when applying a trained closed-loop model, the first n observations will be used as the initial inputs to the model, and then the predicted values will be fed back sequentially to the model as inputs to generate further predications in an autonomous manner.

4.3 Experimental Results

With the given settings, the relevant experiments were conducted accordingly to examine how well a customer's membership in terms of profitability groups can be predicted over time. The average prediction accuracies generated by different models are given in Tables 3 and 4.

Table 3. Average prediction accuracy using observations at one previous time point.

Model	Open-loop		Closed-loop	
	Training	Testing	Training	Testing
Linear Regression	84.82	84.34	84.82	74.90
MLP	84.82	84.34	84.82	74.90
Naïve Bayesian	84.82	84.34	84.82	74.90

Table 4. Average prediction accuracy using observations at two previous time points.

Model	Open-loop		Closed-loop	
	Training	Testing	Training	Testing
Linear Regression	83.79	83.75	83.79	73.89
MLP	84.85	84.90	84.85	74.91
Naïve Bayesian	84.85	84.90	84.85	74.91

5 Discussion

From the experiment results obtained, it is evident that the RFM time series under consideration was well predictable, and a customer's profitability group was stable.

Under all the experimental conditions, the prediction models using observations at one previous time point performed well and had a similar performance to those using observations at two previous time points. This can be further interpreted as the transit

probability of a customer from one profitability group to another at any two consecutive time points was low.

An examination on the transit probability of the customers from one profitability group to another over time has revealed that, on average, the transit probability was not more than 6%. A summary of the average transit probability is given in Table 5, where the element TP_{ij}, $i, j = 1, 2, 3$, in the 3×3 matrix indicates the average transit probability from the ith group to the jth group if $i \neq j$, and the average percentage of customers remained the ith group if $i \neq j$.

Since the business has been running as wholesale as well, the prediction results are quite interpretable and understandable. As such, the profitability of a customer in month t only depended on the profitability of the customer in month $t - 1$. Therefore, it's not necessary to use more past time points in the prediction.

Table 5. Average customer transit probability over time.

Group	1	2	3
1	0.05	0.01	0.01
2	0.01	0.19	0.05
3	0.01	0.06	0.61

In addition, the MLP and the Naïve Bayesian models were slightly more stable than the regression models.

The open-loop prediction models could achieve 84% accuracy and those models were useful for a short-term prediction. The closed-loop prediction models have achieved an accuracy of 79% and they could be applied for a long-term prediction.

6 Conclusions and Future Work

In this study, a comparative study has been conducted on predicting customer profitability dynamically based on monthly RFM time series using multiple models. The study shows a good predictability of the time series under consideration. The context of the business of interest has helped to interpret the prediction results.

Further work includes:

- Using real transactional data collected over a longer period of time, such as two or three years, to examine the predictability of the RFM time series;
- To investigate how prediction accuracy might be affected by the frequency at which the RFM values are calculated with a given transactional data; and
- Using other possible profitability measures to conduct comparative research.

References

1. Google Cloud: Predicting customer lifetime value with AI platform: introduction. https://cloud.google.com/solutions/machine-learning/clv-prediction-with-offline-training-intro. Accessed 10 May 2019
2. Malthouse, E.C., Blattberg, R.C.: Can we predict customer lifetime value? J. Interact. Mark. **1** (19), 2–16 (2005)
3. Chen, D., Guo, K., Ubakanma, G.: Predicting customer profitability over time based on RFM time series. Int. J. Bus. Forecasting Mark. Intell. **2**(1), 1–18 (2015)
4. Beheshtian-Ardakani, A., Fathian, M., Gholamian, M.: A novel model for product bundling and direct marketing in e-commerce based on market segmentation. Decis. Sci. Lett. **7**(1), 39–54 (2018)
5. Chen, P.P., Anna Guitart, A., Fernandez, A., Perianz, A.: Customer lifetime value in video games using deep learning and parametric models. In: IEEE International Conference on Big Data, pp. 2134–2140. IEEE, Seattle (2018)
6. Hojjat Salehinejad, H., Rahnamayan, H.: Customer shopping pattern prediction: a recurrent neural network approach. In: 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, Athene (2016)
7. Gauthier, J.R.: An introduction to predictive customer lifetime value modeling. https://www.datascience.com/blog/intro-to-predictive-modeling-for-customer-lifetime-value. Accessed 02 May 2019
8. Blattberg, R.C., Kim, B.-D., Neslin, S.A.: Database Marketing. ISQM, vol. 18. Springer, New York (2008). https://doi.org/10.1007/978-0-387-72579-6
9. Chen, D., Sain, S.L., Sain, S.L., Guo, K.: Data mining for the online retail industry: a case study of RFM model-based customer segmentation using data mining. J. Database Mark. Customer Strategy Manag. **19**(3), 197–208 (2012)