Customer Shopping Pattern Prediction: A Recurrent Neural Network Approach

Miss. Pooja V. Utkhede[2], Prof. K. B. Bijwe[3]
[1, 2, 3] Dept of Computer Science & Engineering
[1, 2, 3] P.R. Pote (patil)collage of engineering and management,Amravati

Abstract- *Customer relationship management is a popular and strategic topic in marketing and quality of service. The availability of big transactions data as well as computing systems have provided a great opportunity to model and predict customer behaviour. However, there is a lack of modern modelling and analytical methods to perform analysis on such data. Deep learning techniques can assist marketing decision makers to provide more reliable and practical marketing strategic plans. In this paper, we propose a customer behaviour prediction model using recurrent neural networks (RNNs) based on the client loyalty number (CLN), recency, frequency, and monetary (RFM) variables. The experiment results show that RNNs can predict RFM values of customers efficiently. This model can be later used in recommender systems for exclusive promotional offers and loyalty programs management.*

Keywords- Customer Behaviour Prediction, Recurrent Neural Networks, Recency Frequency Monetary(RFM), Shopping Pattern.

I. INTRODUCTION

To manage the customers and their relationships are the most important topic in marketing and e-commerce. And to deal with this challenge One of the popular method is use called customer lifetime value(CLV) model. In marketing, customer lifetime value (CLV) is a prediction of the net profit attributed to the entire future relationship with a customer. Customer lifetime value is an important concept in that it encourages firms to shift their focus from quarterly profits to the long-term health of their customer relationships. CLV is also defined as the value of relationship with the current customers, by considering the future cash flows from the relationship with the customers [1]. Since acquiring new customers are mostly more expensive than keeping current customers, the CLV model attempts to manage the long-term relationships, rather than short-term ones. The decisions are generally made upon classification of customers to a number of loyalty classes. Then, marketing decision makers can manage marketing communication programs through methods such as personalized advertisements and exclusive promotional offers[2].

The CLV models use different methodologies for customer behavior modelling. One of the most reliable ones is using the recency (R), frequency (F), and monetary value (M) variables, called RFM. RFM is a method used for analyzing customer value. It is commonly used in database marketing and direct marketing and has received particular attention in retail and professional services industries. RFM stands for the three dimensions: Recency – How recently did the customer purchase? Frequency – How often do they purchase? Monetary Value – How much do they spend?. These variables represent to understanding the customer behavior and try to answer the above questions. Customer segments containing weighted RFM scores and demographic data in the same clusters conclude stronger and more accurate association rules to understand the customer behavior and customer value[3].

Predicting future consumer behavior is fundamental to many use-cases in e-commerce. Such predictions are based on indicators found in previous consumer behavior. For example, the time since a consumer last visited the webshop, the products that were looked at or the amount of products added to the cart. Behavior is captured in consumer histories, which are, in their raw form, sequences of interactions with the webshop. RNNs operate on sequences of varying length and therefore provide an ap-propriate match to consumer histories. We apply RNNs directly to series of captured consumer actions. RNNs maintain a latent state that is updated with each action. RNNs are trained to detect and preserve the predictive signals in the consumer histories[4].

II. LITERATURE REVIEW

Internet shopping is still in evolutionary stage in India and very few studies have undertaken research exploring customer acceptance and diffusion of internet shopping in India. Although there has been a dearth of internet shopping related studies in Indian context, theoretical exploration can be based on various international studies carried out in other countries. As an initiative to explore the internet shopping acceptance and diffusion in India, this section discusses theories relevant to predicting and explaining actual behavior and behavioral intention and innovation diffusion within the context of internet shopping[lr]. Prior to choice decision or repurchase intention, consumers place a number of attributes in his or her choice sets, in order of importance and relevance. Among these attributes are price and quality, and consumers tend to use price as a proxy to quality (Dodds, Monroe, and Grewal, 1991; Ofir, 2004). However, studies also reveal that, besides price and quality, other cues that are also considered as more important to assess the product‟s worth, are attributes such as brand, store name, past experience, attitude and product information (Curry and Riesz, 1988; Zeithaml, 1988; Tellis and Geath, 1990; Dodds, Monroe, and Grewal, 1991). Brand name, for example, often signals as a cue or as a surrogate of product quality use by consumers in their evaluation of goods or services before they decide to purchase[5].

III. RECURRENT NEURAL NETWORKS

A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed graph along a sequence. RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. Recurrent networks can work with sequences of arbitrary lengths rather than sequences of fixed lengths. This, for example, allows for the analysis of sentences of variable lengths, making them effective for tasks such as translation and sentiment analyses. In order to work with recurrent neural networks, we must choose the number of timesteps. For example, if a network is to predict the next word of a sentence, the network will need to know about the previous words of the sentence. The number of words, in this case, would be the number of timesteps back in time. For each timestep forward in time, the network will take into account the previous output computed upon previous words, consequently making the network remember. In short, the output of previous timesteps affects future timesteps; we refer to this as the memory of a network. Outputs of recurrent nodes and layers depend on previous timesteps, hence recurrent networks are built on memory [6,7] .

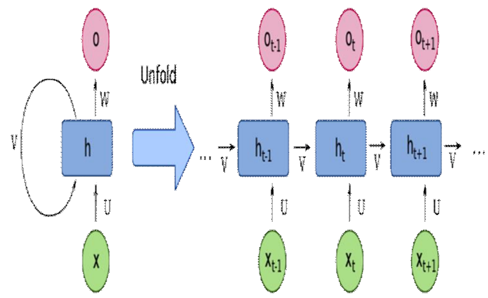

Fig3.1: General RNN model. Left: folded version of the RNN. Right: unfolded version of the RNN.

Figure 3.1 shows the basic RNN model, where the rectangular box contains the nodes of the hidden layer. On the right side, we can see the unfolded version of the RNN, which can be seen as a deep feed-forward neural network with shared parameters between layers. U, V, and W are the weight matrices that are learned. The input at time step Ot+1 is Xt+1, which is connected to the hidden state ht+1 of the network through the weights U. The hidden state Ht+1 is connected to itself via V and to the output Ot+1 via W. The network can be described with the next two equations:

ht+1 = f(Uxt+1 +Vht+1)
Ot+1 = g(W ht+1)
Where Xt+1 2 RD, ht+1 2 RH and Ot+1 2 RK. The parameters learned are the matrices

U 2 RH*D, V 2 RH*H, and W 2 RK*H. f is usually a non-linear and differentiable function applied to the hidden nodes, such as tanh, and g is a function which may depend on the task, for example for a classification task with only one valid class the softmax function is normally used. Given an input sequence x = (x1, x2, …… xT-1, xT) of length T, we can feed the network with each element xt one by one. The function of the network is to store all relevant information through the different time steps in order to capture temporal patterns in sequences. The RNN maps the information contained in the sequence until time step t into a latent space, which is represented by the hidden state vector ht. Therefore, RNNs can solve the problem of modeling the temporal aspect of the data[8].

3.1 LSTM(Long Short Term Memory)

Long Short Term Memory networks "LSTMs" are a special kind of RNN, capable of learning long-term dependencies. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn[9]. The LSTM is composed by a memory cell (a vector) and three gate units. The gates are vectors where sigmoid functions are applied, making their values to be between 0 and 1. The mechanism of LSTM block helps to combat the vanishing and exploding gradient problems and produce better results than the standard RNN. This LSTM block is not the unique model using gates. In an LSTM variant was introduced using peephole connections, so the previous internal cell state of the LSTM is connected to the gates [9,10].

3.2 GRU (Gated Recurrent Units)

Gated Recurrent Units (GRU) are a simplified version of LSTM. They combine the forget and input gates into a single update gate. The model is simpler than the standard LSTM and has fewer parameters, which has lead in some cases to a better performance. There are other alternatives, but most of them contain only small variances. Instead of the input, forget, and output gates in the LSTM cell, the GRU cell has two gates, an update gate, and a reset gate. The update gate defines how much previous memory to keep around and the reset gate defines how to combine the new input with the previous memory. There is no persistent cell state distinct from the hidden state as in LSTM. In LSTM process the last output in a very different way but for the GRU it is easy to repeat the last output, by just saturating the input gate to 0. Thus the LSTM cannot decide, at some point to start reproducing the last output. On the other hand this is easily possible for the GRU within the limits of the approximation of the input gate saturating to 0[11].

IV. PROPOSED MODEL

4.1 Customer Shopping Pattern Model

We study the customer behaviour though time with equal time steps (intervals) as demonstrated in Figure 4.1.1. The time interval can be weekly, bi-weekly, monthly, or etc. Since the first purchase time among customers is different, we define a lower limit and upper limit during the time. The lower limit refers to the start point of our study and the upper limit refers to the end point of our study through time. In this case, we can define some equal time intervals between the lower and upper limits, as shown in Figure 4.1.1. The shopper's purchase is then identified in each time interval and the R, F, and M variables are computed with respect to any point of interest. For example, if the point of interest is at t4, the recency is the time difference between the last purchase before the time of interest and the purchase itself. The time difference can be represented in scale of hour, day, week or etc., depending on the application. The frequency is the number of conducted purchases between the lower limit and the time of interest, which is F = 3 in this example. The monetary is the value customer has spent on the purchases between the lower limit and the time of interest. The R, F, and M values are computed for each customer with a CLN and for all times of interest[2].

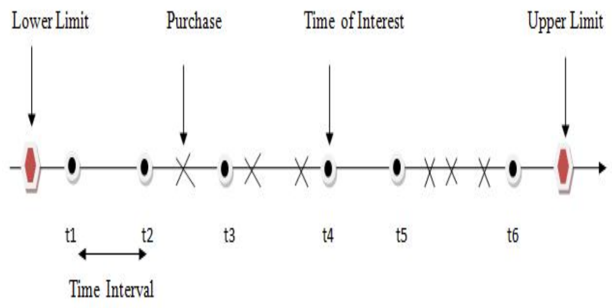

Fig. 4.1.1: A sample of shopper's behaviour during different time intervals.

4.2 THE PREDICTION MODEL

The recency, frequency, monetary and time (RFMT) method is an approach used to measure costumer's loyalty and segment customers into various group for future personalization services. This study identifies customer behavior using (RFMT) model**.** An ANN have been shown to be very promising systems in many forecasting applications and business classification applications due to their ability to "learn" from the data, their non parametric nature and their ability to generalize. A feed-forward back propagation neural network with tansigmoid transfer functions in both the hidden layer and the output layer is proposed to predict the customer behavior. It consists of three layers: the input layer, a hidden layer, and the output layer. A one hidden with 20 hidden layer neurons is created and trained. The proposed network uses the scaled conjugate gradient algorithm for training. The input and target vectors are automatically divided into three sets: 60% are used for training, 20% are used to validate that the network is generalizing and to stop training before over fitting, and the last 20% are used as a completely independent test of network generalization. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy. The information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes. There are no cycles or loops in the network. The proposed neural networks are shown in Fig4.1.2[12].

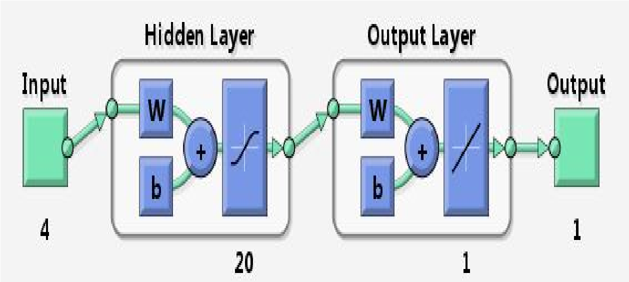

Fig. 4.1.2 The proposed network

4.3 Learning Machine Architecture

The proposed RNN model is consisted of one input layer, one hidden (recurrent) layer, and one output layer. The input layer is an auto-encoder which extracts features from inputs. The CLN, R, F, and M values for each customer at each timestep t are the input sequence to the RNN model. The R, F, and M value of the next time-step t+1 is shown to the model as target through predefined time intervals. The time-step t can be set depending on the application. For example, for grocery stores it can be weekly or bi-weekly and for sports wear every season. In general, the CLNs are provided as large integer digits in transaction data. We use the one-hot encoding method to break the dependencies between integers.

The one-hot encoded CLN and binary representation of Rt, Ft, and Mt are fed to an auto-encoder, which represents each input vector with a feature representation vector of fixed length. Each input vector is fully connected to the representation layer, where WV I is the weight matrix to be optimized while training the model. The feature representation is:

$$v = WV\ Iu;$$

where v is the feature representation of each input parameter CLN, Rt, Ft, and Mt at time t. Then the features are concatenated such as xt = [v, rt, ft, mt] and fed to the recurrent layer[2].

V. CONCLUSION

In this paper we introduce the customer shopping pattern behavior and proposed an approach to apply RNNs to pre-dict future consumer behavior in e-commerce, and also proposes a new model for RFM prediction of customers based on recurrent neural networks. We studied that the recurrent neural networks have powerful pattern classification and prediction capabilities. RNNs can be very effective to model customer interaction data to predict items that the user will purchase in the future. We explored RNNs in a scenario where we have a high number of different items and where using embeddings called LSTM and GRU.

REFERENCES

[1] S. Gupta, D. R. Lehmann, and J. A. Stuart, “Valuing customers,” Journal of marketing research, vol. 41, no. 1, pp. 7–18, 2004.

[2] H. Salehinejad,S. Rahnamayan, “Senior Member, IEEE Department of Electrical, Computer, and Software Engineering University of Ontario Institute of Technology Oshawa, Ontario, Canada” “Customer Shopping Pattern Prediction:A Recurrent Neural Network Approach”.

[3] V. Kumar, Customer lifetime value: the path to profitability. Now Publishers Inc, 2008.

[4] T.Lang, M. Rettenmeier

[5] 'International Review of Business Research Papers Volume 6. Number 4. September 2010. Pp. 279 – 294International Review of Business Research Papers Volume 6. Number 4. September 2010. Pp. 279 – 294'

[6] https://en.wikipedia.org/wiki/Recurrent neural_network

[7] F. Gers and J. Schmidhuber. “Recurrent nets that time and count”. Proc. IEEEINNS- ENNS International Joint Conf. on Neural Networks 3 (2000), pp. 189–194.

[8] Daniel Sánchez Santolaya July 7, 2017 36 EC February 2017 - July 2017

[9] S. Hochreiter and J. Schmidhuber. “Long short-term memory”. Neural Computation 9 (1997), pp. 1735–1780.

[10] colah.github.io/posts/2015-08-Understanding-LSTMs.

[11] J. A. Freeman, and D. M. Skapura, Neural networks: algorithms, applications and programming techniques, Addison Wesley Longman, 1991.

[12] Qeethara Kadhim Al-Shayea, *Member, IAENG*, and Tamara Kadhim Al-Shayea Proceedings of the World Congress on Engineering 2014 Vol I, WCE 2014, July 2 - 4, 2014, London, U.K.