Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 3, July 2021: 2933 - 2937

Research Article

Hybrid Attention based GRU BiLSTM (GRBiLSTM) for Banking Customer Churn Prediction

[1]Mr. M. John Britto, [2]Dr. R. Gobinath,
[1]Research Scholar, Department of Computer Science, VISTAS, Chennai.
[2]Associate Professor, Department of Computer Science, VISTAS, Chennai.
[1]britto.chennai@gmail.com, [2]drgobinathramar.scs@velsuniv.ac.in

Abstract
Early Customer churn detection is a vital aspect of Customer Relationship Management. The behavioral RFM features of the customer can be used for Banking Customer Churn prediction. The transactions made by the Customer are time series dataset. The recent activity features from the transactional data has influence in predicting the churned customers. The proposed Hybrid Attention based GRU BiLSTM (GRBiLSTM) model uses an attention layer based on recent activity behavioral pattern with the BiLSTM. This model outperforms other existing models like LSTM, CNN, and BiLSTM.

Key words:Churn, Deep Learning, Behavioral features, Time series Data, Gated recurrent unit, BiLSTM.

1. Introduction

The customer retention is a major challenge for an organization in the present economic environment. It is faced in many domains telecom, banking, and other non contractual organizations. The Customer retention in a retail bank needs efficient early churn prediction mechanism. It in turns reflects on the growth and goodwill of an organization. The Behavioral pattern attributes of customer such as Recency, Frequency, and Monetary are used along with the months. These values can be extracted from the transactions made by the customer.

Customer churning is not seasonal or periodic it does not need any decomposition rather it need some attention. The overall transactions of a customer and the behavioral pattern over the months are separated into equal bins of four months. The trend feature for the recent months has influence in the churning customer prediction.(Leung and Chung, 2020)

The deep Learning models performs better in prediction mechanism. Some techniques like LSTM which does not suffer from vanishing gradient problem are used for time series data prediction(Venkatesh and Jeyakarthic, 2020). The BiLSTM Technique allows the network to handle the sequence time step information in both forward and backward direction.GRU is used for its fast computational abilities and thus avoids computational complexities(Munawar et al., 2021). In spite of those mechanisms it needs attention for performing better churn prediction. It is achieved by attention layers based on the recent activity trends from RFM features.

2. Related Works

Customer churn prediction can be achieved through Deep Learning Concepts. For effective capture of features from customer data through multilayer feed-forward architecture were implemented. It used the benefits of multi layers in feed forward neural network. It does not process the time series data effectively (Wangperawong et al., 2016).

A framework using Portuguese retail bank with Neural Network implementation was done. Here the author had used behavioral aspects from the transactional data for predicting credit card customer churn using neural network approach and CRISP methodology (Predescu, 2019). (hegde and Mundada, 2019) The EDFFN- Enhanced Deep Feed Forward Neural Network based model was used to predict the attrition rate in the banking sector. Using optimized data preprocessing an enhancement was made to the traditional Deep Neural Network model with Adam optimizer technique.

(Li et al., 2019)The fully-connected neural network, the recurrent neural network formed with stacked Long-short term memory (LSTM) and the ensemble model which combined the advantages of the two types of neural networks architectures were implemented for customer churn prediction. Here the data reduction was not better focused. (Li and Xie, 2020) The GAN network is used for predicting churn. It has focused on generating minority class samples to improve imbalanced data. F1 and precision were used as evaluation metrics. It used the demographic details of the customer.

(Leung and Chung, 2020)A dynamic classification for optimize customer churn prediction. This approach uses trend factor to extract behavioral features. It also used under sampling and SMOTE to handle imbalanced data set. Even though it used transactional data set, time series prediction was not implemented.

(Zhang et al., 2017)Deep and shallow model for Insurance churn prediction used generalized linear model for shallow part and feed forward neural network for deep model. It uses demographic data alone. Time series data set was ignored. It focused on insurance domain. (Simion-Constantinescu et al., 2018) Deep neural pipeline is a customer churn prediction model for telecom industry based on deep neural network model uses CNN and LSTM. The recall achieved by them can be improved via BiLSTM attention models.

3. Data Set Description

The Czech Financial dataset has 4500 Customer data over 14 topics such as account balance, credit card information, demographic data, Full date and time of transaction and so on. It is dated from January 2015 to December 2018 around 1 million transactions.

Unlike other Datasets, the target variable is not present and so it is computed based on time-stamp differences. Customers with last three month of inactivity or with no balance amount are labeled as churned. The transactional data such as ATM transactions, withdrawal, deposit, internet transactions, Time are present in the dataset. The behavioral features are extracted from the dataset for last one year of transactions from that dataset for calculating recent month trend from each bin.

The Bank transaction dataset has millions of transactions over the period. Thus data cleaning and reduction is necessary. The data with credit transaction of last one year transaction is taken into account for prediction of bank customer churn. Here, down sampling through reduction through exclusive homogeneous clustering (REHC) technique is used.

4. Proposed Model

The Hybrid Attention based GRU BiLSTM is proposed for early banking churn prediction. Recent activity trend for each bin is used to attain new weights in attention layer. The behavioral features extracted from the transaction dataset have been given as input. Influencing recent activity among RFM activity is calculated from each bin separately. The proposed model is depicted in figure 1.

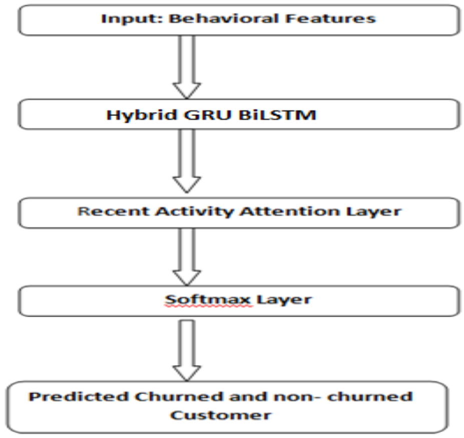

Figure 1 Hybrid Attention based GRU BiLSTM Model

4.1. BiLSTM Model

The BiLSTM is a sequence processing model with two LSTMs. IT uses both forward and backward direction processing. LSTM solves the issue of vanishing gradient in the traditional RNN. Figure 2 shows BiLSTM architecture. It introduces memory cell and gate mechanism(Venkatesh and Jeyakarthic, 2020). It uses following gates

Input gate: $i_t = \sigma (w_f . [c_{t-1} , h_{t-1}, x_t] + b_f)$ (1)

Forget gate: $f_t = \sigma (w_{i.} [c_{t-1} , h_{t-1}, x_t] + b_i)$ (2)

Output gate: $o_t = \sigma (w_o . [c_t , h_{t-1}, x_t] + b_o)$ (3)

Hidden state: $h_t = o_t * \tanh [c_t]$ (4)

Final output of ht will be from computed from forward and backward ht

$$h_t = [fh_t, bh_t] \quad (5)$$

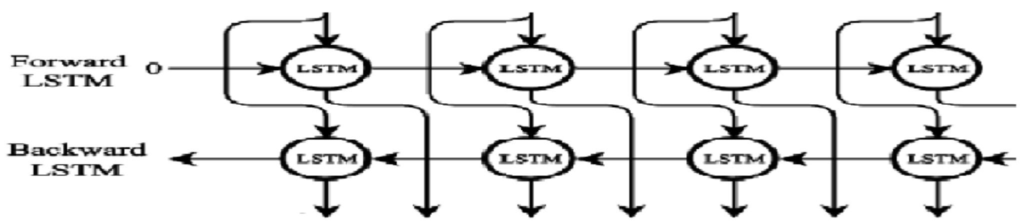

Figure 2 BiLSTM Architecture

4.2 Attention Layer

The attention weights are transferred into the network at right part and produce a corresponding loss value based on the predicted error(Li et al., 2019). The churning customer of a bank is not seasonal or periodic. It needs special attention based on recent activity trends of a customer. The recent month activity trend from each bin is computed and is given as attention weight. The figure 3 shows the attention based BiLSTM. The attention weights can be calculated and computed as follows

If (R> M and R>F)
Then activity ← R
Else if (M>F)
Then activity ← M
Else
activity ← F
tf = (Σ (weight * activity) * (No._of_Months/ Σ weight)) / Σ activity.

Based on the attention weights attention vector is calculated as follows.

$\alpha_t = f(c_t, h_t) = \tanh(w_c [c_t . h_t])$ (6)

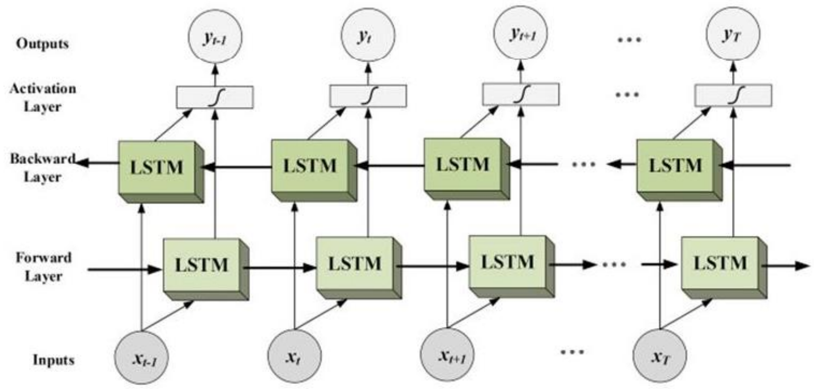

Figure 3 BiLSTM with Attention Layer

5. Results and Discussions

Customer churning in a Bank is imbalanced problem. Numbers of non-churned customers are comparatively high against churned customers. This churned customer is merely about 2 to 3 percent. Thus accuracy cannot be considered as evaluation metric.

Here, the recall value is considered as the evaluation metrics. It determines the churned customers are predicted well by the model. The table 1 depicts the precision and recall values of a few models.

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

Table 1: Evaluation Metrics of different models

Models	**Precision**	**Recall**
RNN (LSTM)	87.09	90.97
BiLSTM	84.30	95.56
Hybrid Attention based- GRU BiLSTM	83.74	**97.88**

The recall value of proposed Hybrid Attention based GRU BiLSTM model is 97.88. It implies that the churned customers are predicted well by this model.

6. Conclusion

The Customer churn prediction in banking sector using time series data set is classified using Hybrid attention Based GRU BiLSTM. For resolving local features relationships during temporal classification, an attention layer for influence based recent activity trend is proposed. This Hybrid Attention based GRU BiLSTM captures long term dependence in time series prediction with local relationship attention. Since, it is imbalanced data set the recall value is used for evaluation metrics. It predicts the churning customers (True Negative) and thus recall value is used and it is 97.88 for the proposed Hybrid Attention based GRU BiLSTM (GRBiLSTM). Experiments shown that Hybrid Attention based GRU BiLSTM can provide competitive churn prediction performance compared to other existing models. For future work this model can be hybrid with other deep learning models which utilizes demographic data.

References:

1. hegde, S. and Mundada, M. R. (2019) 'Enhanced Deep Feed Forward Neural Network Model for the Customer Attrition Analysis in Banking Sector', International Journal of Intelligent Systems and Applications, 11(7), pp. 10–19. doi: 10.5815/ijisa.2019.07.02.
2. Leung, H. C. and Chung, W. (2020) 'A dynamic classification approach to churn prediction in banking industry', 26th Americas Conference on Information Systems, AMCIS 2020, pp. 0–5.
3. Li, B. and Xie, J. (2020) 'Study on the Prediction of Imbalanced Bank Customer Churn Based on Generative Adversarial Network', Journal of Physics: Conference Series, 1624(3). doi: 10.1088/1742-6596/1624/3/032054.
4. Li, Y. et al. (2019) 'EA-LSTM: Evolutionary attention-based LSTM for time series prediction', Knowledge-Based Systems. Elsevier B.V., 181, p. 104785. doi: 10.1016/j.knosys.2019.05.028.
5. Munawar, S. et al. (2021) 'Electricity Theft Detection in Smart Meters using a Hybrid Bi-directional GRU Bi-directional LSTM Model', (April).
6. Predescu, A. (2019) A Multiple-Layer Clustering Method for Real-Time Decision Support. Springer International Publishing. doi: 10.1007/978-3-030-04849-5.
7. Simion-Constantinescu, A. et al. (2018) 'Deep Neural Pipeline for Churn Prediction', Proceedings - 17th RoEduNet IEEE International Conference: Networking in Education and Research, RoEduNet 2018. doi: 10.1109/ROEDUNET.2018.8514153.
8. Venkatesh, S. and Jeyakarthic, M. (2020) 'Adagrad Optimizer with Elephant Herding Optimization based Hyper Parameter Tuned Bidirectional LSTM for Customer Churn Prediction in IoT Enabled Cloud Environment', Webology, 17(2), pp. 631–651. doi: 10.14704/WEB/V17I2/WEB17057.
9. Wangperawong, A. et al. (2016) 'Churn analysis using deep convolutional neural networks and autoencoders', pp. 1–6. Available at: http://arxiv.org/abs/1604.05377.
10. Zhang, R. et al. (2017) 'Deep and Shallow Model for Insurance Churn Prediction Service', Proceedings - 2017 IEEE 14th International Conference on Services Computing, SCC 2017, pp. 346–353. doi: 10.1109/SCC.2017.51.