論文の要旨

題目：**Studies on Time Series Forecasting by using Hybrid Deep Learning Architectures**

（ハイブリッドディープラーニングによる時系列予測に関する研究）

氏名 **YEPENG CHENG**

Abstract

This dissertation discusses logistic regression analysis-based retailer competition analysis, and hybrid deep learning architectures-based multi-conditional time series forecasting, respectively.

In today's supermarket business, the ID-POS database enables supermarkets to analyze customer behavior and adopt more targeted and personalized marketing strategies such as customer relationship management (CRM), to improve the competitiveness of supermarkets. The ID-POS database digitally records customer ID, customer information, sales records., etc. Therefore, customer behavior is measurable by counting their daily shopping records as customer values. Generally speaking, customer value analysis, which is also known as RFM analysis, mainly depends on three parametric indicators, customer shopping recency, frequency, and monetary. They can reflect the customer loyalty of a store. The models consist of RFM indicators with other statistical parameters that are trainable by clustering analysis and other machine learning methods to investigate the customer shopping preference.

In general, there are many supermarkets in a city, and other nearby competitor supermarkets significantly affect the customer value of customers of a supermarket. However, it is impossible to get detailed ID-POS databases of competitor supermarkets. This thesis firstly focused on the customer value and distance between a customer's home and supermarkets in a city, and then constructed the models based on logistic regression analysis to analyze correlations between distance and purchasing behaviors only from a POS database of a supermarket chain. Logistic regression analysis is widely used in parametric impact analysis. The coefficients of logistic regression mathematically considered as the parameters in the Odds ratio. The Odds ratio can reflect the influence of variable parameters on a particular parameter. During the modeling process, there are four primary problems existed, including the incomparable problem of customer values, the multicollinearity problem among customer value and distance data, the number of valid partial regression coefficients and the target values of logistic regression analysis for training (loyal customer mark). The RFM score, Huff's gravity model, inverse

attractiveness frequency and Decyl analysis are considered to solve these problems, respectively. We present three types of models based on these methods for loyal customer classification and competitors' influence analysis from the viewpoint of valid partial regression coefficients and accuracy.

We also discuss the multi-conditional time series forecasting via hybrid deep learning architectures in this dissertation. In big data analysis, time series forecasting is an essential branch developed in recent years. Traditional methods have some limitations for time series forecasting since the time series possess characteristics such as non-linearity, non-stationarity and unknown dependencies. Deep learning is an advanced approach to overcome these problems. It depends on non-linear modules to learn the fully features from the input data. In recent research of deep learning, it existed a state-of-art deep learning structure named SeriesNet, which combined the dilated causal convolutional neural networks (DC-CNN) and the long-short term memory (LSTM). This model has higher forecasting accuracy and greater stableness. LSTM and DC-CNN are widely applied to time series forecasting with excellent performance. However, DC-CNN and LSTM include a large number of parameters, resulting in tremendous computation cost. Gated recurrent unit network (GRU) and LSTM have a comparable performance on time series forecasting, but parameter quantity significantly reduced. So does the dilated depthwise separable temporal convolutional networks (DDSTCNs) compared with DC-CNN. The SeriesNet can directly input raw time series sequences by conditioning the target time series on the additional time series. But the specific conditioning method is not clarified in their work. In addition, they did not consider the attention mechanisms in SeriesNet. Recently, most researches focus on the recurrent neural network (RNN) based attention to improve the deep learning structure. However, the heavyweight attention mechanism within massive training parameters will influence the computation efficiency. The convolutional block attention module (CBAM) [9] is a lightweight attention structure, but has only been successfully applied to image recognition so far. Therefore, we introduce the conditioning methods for CNN and RNN and propose a lightweight hidden state attention module (HSAM) on RNN layers. We have utilized the attention mechanisms in SeriesNet and present an attention-based SeriesNet (A-SeriesNet) combined CBAM on convolutional layers and HSAM on RNN layers for time series forecasting. We adopt GRU and DDSTCNs instead of LSTM and DC-CNN of SeriesNet to reduce the parameters in neural network layers.

Although our proposed A-SeriesNet is superior to state-of-art deep learning models for time series forecasting, it has some disadvantages for high feature dimensional time series. The A-SeriesNet is a two subnetworks hybrid neural network architecture, which contains augmented attention residual learning module-based convolutional neural network (augmented ARLM-CNN) subnetwork and hidden state attention module-based recurrent neural network (HSAM-RNN) subnetwork. Both of its subnetworks are not encoder-decoder structures. In recent research, the dual-stage attention recurrent neural network (DA-RNN) proved that the attention-based encoder-decoder framework is an effective model for dealing with the high feature dimensional time series. Therefore, the A-SeriesNet is not suitable for time series datasets that feature dimension higher than 15. The CNN subnetwork conditions the multi-condition series on a target time series by simultaneously feeding them to an augmented residual learning module. With the multi-condition series's feature dimension increasing, the relation between multi-condition series becomes more complicated. Without prior feature extraction of raw multi-condition series may pollute the target time series to some extent. The conditioning method of the RNN subnetwork reshapes the multi-condition series into the first GRU layer's hidden state size in advance. The reshaped multi-condition series are fed to the first GRU layer as its initial hidden state. When the multi-condition series own large feature dimensions, this method may lose some information during the reshaping process. The RNN subnetwork only has a global HSAM attention mechanism which generates an attention weights vector for all hidden states after the first GRU layer performed all its update steps. The GRU layers and the HSAM are independent of each other. The mere HSAM can not detect each multi-condition series's importance for prediction results due to the reshaping preprocess. The concatenation method of A-SeriesNet limits the number of its subnetworks. The overall prediction is liable to be impacted by either of its subnetworks.

Aim to above disadvantages of A-SeriesNet, we consider and present the TA-SeriesNet and its subnetworks for high feature dimensional time series forecasting. Intuitively, we apply the DA-RNN to the HSAM-RNN subnetwork of A-SeriesNet and present the triple-stage attention-based recurrent neural network (TA-RNN) subnetworks, TA-LSTM and TA-GRU. Furthermore, we consider a CNN-based encoder-decoder structure named dual attention residual learning module-based convolutional neural network (DARLM-CNN) subnetwork to improve the augmented ARLM-CNN subnetwork of A-SeriesNet. The DARLM-CNN subnetwork used ARLM-CNN as the encoder and augmented ARLM-CNN as the decoder. All subnetworks of TA-SeriesNet are encoder-decoder structures more effective for high dimensional time series. The DARLM-CNN structure extracts the feature context vector from high feature dimensional multi-condition series by its ARLM-CNN encoder. The augmented ARLM-CNN is fed by the generated feature context vector and the target time series to reduce the raw multi-condition series's pollution for the target time series. In the TA-RNN subnetwork, the multi-condition series are fed to an RNN encoder as the input (not a hidden state) directly without shape variation, which ensures the information integrity of the multi-condition series. The TA-RNN subnetworks have one global HSAM attention and two local attention mechanisms, input attention and temporal attention. They extract the importance of each multi-condition series for feature context vector generation from the feature dimension axis and the time step axis. The input attention weights are updated by each time unit states of the RNN encoder. The learned each time input attention weights are element-wise multiplied by each time multi-condition series and fed to the RNN encoder again to generate its next unit states. This process will not terminate until the RNN encoder performed all its time steps. Similarly, this update process also happens between the temporal attention and the RNN decoder. The input attention, the temporal attention and the RNN encoder-decoder are not independent. We adopt a new concatenation method instead of the element-wise multiplication of A-SeriesNet to free the parallel connection number of subnetworks. Each subnetwork's learnable output weight promotes to reduce the output dependence of TA-SeriesNet on a certain subnetwork.

The outline of this dissertation is organized as follows.

Chapter 1 discusses the motivations and gives the organization of this dissertation.

Chapter 2 introduces the logistic regression analysis-based retailer competition analysis in detail. In this chapter, we first introduce the related works about RFM analysis, Huff’s gravity model, and Decyl analysis. Secondly, we give the definition of them. And then, we propose the inverse attractiveness frequency. Finally, we integrate these methods to generate three models and train them by logistic regression. The loyal customer classification accuracy and the valid partial regression coefficients of logistic regression are used for model evaluation.

Chapter 3 explains the detail of the hybrid deep learning architectures-based multi-conditional time series forecasting. In this chapter, we first show the related works about CNN, DC-CNN, DDSTCNs, residual learning module, RNN, LSTM, GRU and DA-RNN. Then we give the definition of them. After that, we introduce our proposed A-SeriesNet and TA-SeriesNet. Finally, we comprehensively evaluate the performance of different state-of-art deep learning models with our proposed models for different feature dimensional datasets.

Chapter 4 summarizes the conclusions and contributions of this dissertation.