Combining Heterogeneous Features for Time Series Prediction

Charles Chu*, James Brownlow*, Qinxue Meng*, Bin Fu*, Ben Culbert*, Min Zhu* and Guandong Xu†

* Marketing, CFS, Sydney, Australia

† Advanced Analytics Institute, Faculty of Engineering and Information Technology,
University of Technology Sydney, Sydney, Australia

Email: {Charles.Chu, James.Brownlow, Qinxue.Meng, Bin.Fu, Ben.Culbert, Min.Zhu}@cba.com.au, Guandong.Xu@uts.edu.au

Abstract—Time series prediction is a challenging task in reality, and various methods have been proposed for it. However, only the historical series of values are exploited in most of existing methods. Therefore, the predictive models might be not effective in some cases, due to: (1) the historical series of values is not sufficient usually, and (2) features from heterogeneous sources such as the intrinsic features of data samples themselves, which could be very useful, are not take into consideration. To address these issues, we proposed a novel method in this paper which learns the predictive model based on the combination of dynamic features extracted from series of historical values and static features of data samples. To evaluate the performance of our proposed method, we compare it with linear regression and boosted trees, and the experimental results validate our method's superiority.

I. Introduction

Time series prediction is to build models based on previously observed values to predict future values. Currently, it has been a major research focus in the field of data mining and machine learning because of its wide application in industries including banking, finance, retailing, and manufacturing etc.

Until now, a variety of methods have been proposed or adapted for learning predictive models, i.e., linear methods [1], kernel based methods [2], Gaussian processes [3], recurrent neural networks (RNNs) [4], and tree-based optimization models [5] etc. However, these methods mainly use previously observed series of values to learn models and make predictions, while more features could be extracted from heterogeneous sources. Therefore, in order to incorporate various types of features into the learning process to improve predicting performance, we propose a novel method in this paper which is capable of integrating both previous series of value and features of data samples to learn predictive models (TS-F). In this way, more information are exploited and thus the potential issue of insufficient series of previous value is alleviated. To validate the practical effectiveness of our proposed method, we compare it with the other advanced models on a dataset collected from the Commonwealth Bank of Australia [1].

II. Methodology

This section formulates the research question to be resolved, followed by the description of the proposed predictive learning

[1] https://www.commbank.com.au/

model.

A. Problem Definition

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a dataset that consists of n instances or samples. Each instance usually consists of m features and is represented as a m-length vector. Furthermore, an instance could be associated with a time series $(x_1^i, x_1^i, \ldots, x_t^i)$ which is a sequence of successively observed values in which x_t^i is a real value that represents the value of sample x^i in time slot t. For instance, financial advisers have features like gender, age etc. and also have a sequence numbers of accounts in previous years. The research question in this paper is how to forecast the future values of samples based on information related to both of their previously observed series of values (dynamic features) and their intrinsic characteristics (static features).

B. TS-F Predictive Learning Model

The proposed learning method is based on the gradient boosting optimization [5] to minimize the loss between predictions (y') and actual values (y) with explicit consideration of both time series data and static features. The objective function is defined as
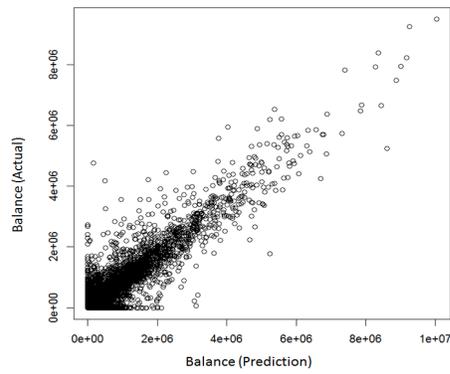
$$
\begin{aligned}
\Theta = \min_{\alpha, w_t, w_f} & \left((\alpha * f_{w_t}(X_t) + (1 - \alpha) * f_{w_f}(X_f)) - y \right)^2 \\
& + \lambda (\alpha * ||w_t||^2 + (1 - \alpha) * ||w_f||^2)
\end{aligned}
\tag{1}
$$

where X_t is time series data, f_{w_t} is the predictive learning model of time series data with parameter w_t while X_f is the static features of samples, f_{w_f} is the predictive learning model of static features with parameter w_f; α is the trade-off parameter and λ is a hyper-parameter to control the strength of the regularization which is introduced to decrease model complexity and avoid over-fitting. In the experiment, we applied Gaussian Regression to deal with time series data and regression tree model to deal with static features as most features are categorical values. It should be noted that heterogeneous features from different sources are combined and exploited in this experiment, including previously observed number of accounts and balance amount as well as advisers' demographics, historical behaviors, transactions, and so on.
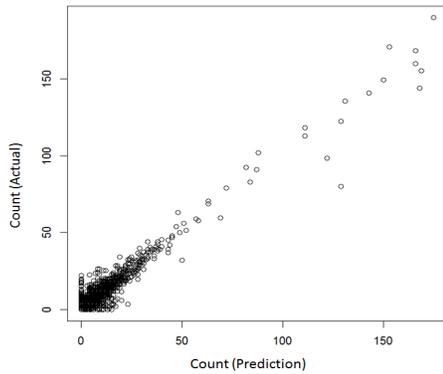
(a) Result of predicting number of customers   (b) Result of predicting amount of balance

Fig. 1. Experimental results of using our proposed method to predict (a) the number of customers and (b) the amount of balance for each financial adviser.

TABLE I
PREDICTIVE LEARNING RESULTS

Method	Prediction	MAE	RAE	RMSE	R^2
Linear regression	Count	2.437	0.516	3.27	0.521
	Balance	65,647	0.727	109,273	0.422
Boosted trees	Count	1.746	0.329	2.363	0.871
	Balance	14,782	0.397	75,219	0.811
TS-F	Count	0.039	0.155	0.351	0.929
	Balance	4,841	0.169	5,253	0.879

C. Evaluation

The performance of the proposed forecasting model is evaluated by a set of benchmark regression evaluation metrics including Mean Absolute Error (MAE), Relative Squared Error (RAE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2).

III. EXPERIMENT

This section describes the experimental dataset followed by the configuration of experiment and the comparative results of linear regression, boosted trees and the proposed TS-F predictive learning model.

A. Experimental Datasets

The experimental dataset is collected by Commonwealth Bank of Australia that contains 19,535 financial advisers with time series data of their annual customer balance ranging from 2012 to 2016 and 53 adviser profiling features. It can be found that the time series data is short-term with only six time slots.

B. Experimental Results

In this experiment, we forecast the number of customers and their balances in terms of ages for each financial adviser. For evaluating and comparing the performance of predictive learning models, 10-fold cross validation is used and the results of the average performance of each model is demonstrated in Table I. It can be found that the performance of our proposed TS-F predictive learning model outperforms linear

regression and boosted trees. For Count, the proposed TS-F model achieves highest value (0.929) in R^2 followed by boosted tree (0.871) and linear regression (0.521). For balance, TS-F achieves 0.879 in R^2 followed by boosted trees and linear regression as 0.811 and 0.422 respectively. The reversed pattern can be found in MAE, RAE and RMSE. Figure 1 further illustrates that the predicated results by TS-F are highly correlated to the actual results where dots are condensed into an area along with the diagonal of the coordinates.

IV. CONCLUSION

The proposed TS-F method considers both time series and sample features so as to decrease the information loss and then increase the accuracy of prediction. Specifically, TS-F achieve highest values in R^2 and lowest values in MAE, RAE and RMSE. In the future, we will further evaluate the effectiveness and efficiency of TS-F model on public benchmark datasets.

REFERENCES

[1] C. Liu, S. C. Hoi, P. Zhao, and J. Sun, "Online arima algorithms for time series prediction." in *AAAI*, 2016, pp. 1867–1873.

[2] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.

[3] K. Yamanishi and J.-i. Takeuchi, "A unifying framework for detecting outliers and change points from nonstationary time series data," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2002, pp. 676–681.

[4] T. Guo, Z. Xu, X. Yao, H. Chen, K. Aberer, and K. Funaya, "Robust online time series prediction with recurrent neural networks," in *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 816–825.

[5] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794.