Predictability Classes for Forecasting Clients Behavior by Transactional Data*

Elizaveta Stavinova, Klavdiya Bochenina, and Petr Chunaev

National Center for Cognitive Technologies, ITMO University
Saint Petersburg 199034 Russia
estavinova@icloud.com, k.bochenina@gmail.com, chunaev@itmo.ru

Abstract. Nowadays, the task of forecasting the client's behavior using his/her digital footprints is highly demanded. There are many approaches to predict the client's next purchase or the next location visited that focus on achieving the best possible prediction quality in terms of different quality metrics. Within such approaches, the quality is however usually evaluated on the entire set of clients, without dividing them into classes with a different predictability rate of client's behavior. In contrast to the approaches of this type, we propose a method for the identification of the client's behaviour predictability class by means of a foreign trip in the next month by using only client's historical transactional data. In a sense, this allows us to estimate the quality of forecasting the client's foreign trip before the actual prediction procedure. Our experiments show that the approach is rather efficient and that the predictability classes obtained quite agree with the prediction quality classes found within the actual forecasting.

Keywords: Event forecasting · Predictability · Transactional data.

1 Introduction

Forecasting the client's behavior is an extremely popular topic nowadays with different applications — from the prediction of the next purchase [19] to that of the future location [12]. There are many approaches that focus on achieving the best possible prediction quality in terms of different quality metrics. It is common that the predictions are made and evaluated due to training the forecasting models on a part of the data with further testing them on some test data. Within such approaches, the quality is however usually evaluated on the entire set of clients, without dividing them into classes with a different predictability rate of client's behavior. In contrast to the approaches of this type, we are interested in the identification of the client's behaviour predictability class by means of a foreign trip in the next month (a particular case of an *event*) by using only client's historical transactional data.

* This research was financially supported by the Russian Science Foundation, Agreement 17-71-30029.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

To be more precise, we propose a classification method that exploits the idea that all clients can be divided into classes based on the predictability rate of a foreign trip in the next month according to their transactional history. Once the clients are divided into the classes, a bidirectional Long Short-Term Memory (LSTM) network [16] is trained on some data to identify the predictability class before the actual forecasting. This trained network can be further used to identify the predictability class of a new client or to rearrange the client's class in the case of changing his/her travel behavior. The fact that we know that a client belongs to one of the predictability classes gives us the opportunity to estimate the trip prediction quality for a client before the actual forecasting.

Earlier works in the field of predictability were dedicated to the measurement of the time series predictability [9] and the predictability of the features [8]. It should be mentioned that the goal of the time series predictability analysis is to estimate how possible is it to capture the time series patterns while the feature predictability analysis is aimed to the selection of such features which are useful for the predictive model. The novelty of our research consists in the estimation of the event predictability and the usage of it for the assessment of the model performance before the prediction. By event predictability we mean the predictability of time series consisting of categorical values that are binary labels indicating the event.

Let us mention that we use the LSTM network for making the predictions because of its advantage in remembering time dependencies which are important for this kind of task [5]. This network was already used for the prediction of the client's location [17] and proved its applicability to this problem.

It is important to note that our classification method seems to be useful for financial organizations to develop beneficial rates or personalized advertising campaigns for those clients whose travel behavior is more predictable.

The paper is organized as follows. Section 2 contains the related work on the topic of predictability and different measures which were developed to estimate it. Section 3 describes the methodology of the research connected with the forecasting the foreign trips and defines the predictability phenomena. Finally, Section 4 presents the data and gives a look at how the raw transactional data are processed to be used further. Section 4 presents the results and their analysis.

2 Related work

The first definition of an unpredictable random variable was proposed in [2]. A random variable x_t is called unpredictable with respect to an information set Ω_{t-1} if the conditional distribution $F_{x_t}(x_t|\Omega_{t-1})$ and the unconditional distribution $F_{x_t}(x_t)$ of x_t coincide, i.e.

$$F_{x_t}(x_t|\Omega_{t-1}) = F_{x_t}(x_t). \tag{1}$$

In particular, if Ω_{t-1} consists of the past realizations of x_t, then (1) indicates that the knowledge about the past realizations does not improve the prediction

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

quality of x_t. Note that the unpredictability of x_t in this sense is an inherent property of x_t that is independent of a prediction algorithm.

It is important to understand the differences between the concepts of the predictability and the prediction quality. These concepts are closely related but differ in the time moment of their determination as the predictability is determined before the prediction while the prediction quality is done after the model responses have been received. Thus, the predictability rate is an individual measure that is inherent to each object (time series, feature, event, etc.) and the prediction quality is an aggregated one, that is based on the model predictions. Further in this section a literature review on measures that are used to estimate the predictability of various objects is presented.

2.1 Predictability measures for time series

A significant part of all works on predictability measures is represented by studies dedicated to measuring the predictability of a time series before the actual forecasting. In most cases, the predictability measure should answer the question: is it worth to make predictions for a given time series, or in other words, does this time series contain some patterns that could be inferred and learned by a forecasting model? A first measure that tries to answer the above-mentioned question is now called the Kaboudan coefficient and is proposed in [9]. In addition, the author conduct experiments on the usage of this measure on time series from the financial sphere. To measure the predictability, the coefficient exploits the idea of comparing the original time series with a time series obtained by random shuffling the values of the original one.

The coefficient was further used as a base for other predictability measures. For example, a modified coefficient is proposed in [3] to overcome the following two problems of the initial one: (a) the dependence between the coefficient value and the time series length and (b) the narrow range of the coefficient distribution in the case of a long-term series. A few years later, the authors of [14] further modify the Kaboudan coefficient and apply it for measuring the predictability rate of financial time series.

In addition to the papers applying predictability measures to financial time series, there is one on the analysis of time series describing the streamflow observed in river basins [20]. Its authors propose to estimate the predictability rate of a univariate time series via the so-called coefficient of efficiency:

$$CE(n, Q) = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - \bar{Q})^2},$$

where n is the test period size, Q_i is the actual series value, $\hat{Q}_i$ is the predicted value, $\bar{Q}$ is the average value of the observation period series.

Moreover the concept of predictability is used to analyze the efficiency of the genetic programming models applied to time series prediction [1, 21]. The authors of [18] also work with it to determine the most appropriate predictors in the problem of genomic sequence identification. As for the multivariate time

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

series, there is a measure that allows to estimate the expected decrease in the prediction error of a multivariate model in relation to a univariate one [13].

2.2 Predictability measures for features

The purpose of the features predictability estimation consists in the selection of such a feature set that describes the object behavior in the best way and allows the model to make predictions of a desired quality. In fact, if a feature does not contain information about the future, its usage may add randomness to model responses and prevent it from making an accurate prediction. In the case of large input data dimensions, it becomes necessary to match the initial features with a certain feature set of a smaller dimension which can help the model to make more stable and efficient forecasts. However, standard dimensionality reduction algorithms are focused on the preserving the data properties that are not related to the predictability, and therefore there is a possibility of missing important information contained in the data.

One of the works devoted to the determination of the features predictability is [8], which identifies the most useful features for predicting the remaining time of the system performance. The authors define the predictability rate as a function that depends on the prediction horizon, the model class, the model parameters and the required accuracy threshold. The proposed predictability measure pulls the threshold and the accuracy achieved by the model into a single value between 0 and 1. Next, the set of features and the model providing the best predictability are chosen by brute force.

There are a few works in the field of the predictive features extraction developed for multivariate time series. The recently proposed method called Forecastable Component Analysis [6] is in fact one of the methods for reducing the dimension of time-dependent signals. Predictive Feature Analysis [15] is an unsupervised learning algorithm that aims to select only those input signals that behave as predictable as possible. Another approach to the dimension reduction of the input data is called Slow Feature Analysis [22]. It explicitly uses time dependencies in the data and distinguishes slowly changing features which can be regarded as predictable ones.

3 Method

3.1 Predictability measure

In this work we face the task of event prediction and therefore adapt (1) for providing the following definition of event predictability. If the distribution function of the label at the moment t does not depend on its previous values, i.e.

$$F(label_t|\Omega_{t-1}) = F(label_t), \tag{2}$$

where F is the distribution function, Ω_{t-1} is previous values of the label, then the label is called *unpredictable*. Here a label means a binary indicator of the event at time step t (1 is when the event takes place while 0 when it does not).

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

As for the predictability rate, it can be estimated by the prediction error-based approach. Often the predictability rate is called the sample predictability rate because it is computed using the forecast errors. Here, we use the following simple sample predictability rate:

$$C(n, Q, m) = 1 - \frac{1}{n}\sum_{i=1}^{n}|Q_i - \hat{Q}_i| \in [0, 1],$$

where n is the test period size, Q_i is the actual event indicator, $\hat{Q}_i$ is the predicted probability of the event, m is a forecasting model.

The proposed coefficient is used in our method to divide all clients into classes according to the predictability rate of a foreign trip in the next month. This division is done after the model predictions are obtained for the train data. Basing on the values of C, we will determine two predictability classes, separated by the value of $\mathsf{median}(\mathbf{C})$, where $\mathbf{C}$ is the set of C-values for the entire dataset of clients. The motivation for this division in two classes is to distinguish objects with high and low predictability.

3.2 Predictability class identification

We identify the client's behavior predictability class by computing the value of C for a chosen dataset and a model. Thus, for the identification we should obtain model predictions and compare them with the real data to know the predictability rate. After that, we can estimate how predictable is the client's behavior. But in practice, it will be very useful to skip the step of using a model and having only the event indicators for the past few months claim if this client has predictable trip behavior or not. That is why we developed a method for the predictability class identification by the sequence of the event indicators.

In fact, this method is the way of solving the classification task. The goal is to identify which predictability class a client belongs to having only feature vectors of the train period. This task is as close to practice as possible, since at a certain point in time we have access only to the observation period data without possibility to access the data from the future (i.e. from the test period).

To solve this problem of the sequence classification we use a Bidirectional LSTM network [16]. The *input* of this network is a set of categorical sequences consisting of the event indicators (or the number of the events) with the step of one month. The length of the sequence is chosen to be six. The network *outputs* the predictability class for each sequence from the input. To train this model, we extracted the sequences describing the client's trip behavior throughout the last six months of the train period and estimated these clients predictability class using the prediction model, its answers and the coefficient C. After training the classification model on the data for the last six months, the quality is measured using the next six months of the transaction history. The main idea here is obtaining the trained model which can be further used in the case of new data arriving when we should recalculate the client predictability class or identify the predictability class for a new client.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

4 Results

4.1 Data description and processing

For our experiments we use three transactional datasets from Russian banks.

Firstly, we describe two datasets that contain transactions made by clients of a large regional Russian bank and that cannot be made publicly available due to the bank's policy. The first dataset ($D1$) contains the transaction history of almost 3,000 clients over two years (from January 2016 to January 2018). The second dataset ($D2$) consists of transactions made by 6,000 clients in 2018 (from January to December). Although $D2$ has the shorter observation period, it contains three times more transactions than in the first dataset. Thus, $D1$ is sparse and may have the incomplete transaction history for some clients. Also, one of the possible explanations of this fact is the difference in the clients samples: the first sample of clients can be more passive in their spending behavior.

Now we describe the third dataset used that is publicly available and serves for the purposes of reproducibility of our results. Namely, this dataset (called Raiffeisen below) is from a Kaggle competition initiated by the Raiffeisen bank. This dataset as well as the implementation of the approaches described in this paper can be accessed through Github[1]. The dataset contains the transaction history of 10,000 clients and describes their spendings during 2017 (from January to December).

The structure of the datasets is very similar so we can process them in a rather unified way. The first stage of the data processing is the categorization of transactions based on the Merchant Category Code (MCC). All MCCs are divided into 87 categories. Then all transactions are divided into the two groups according to the location: made in Russia and abroad. This is done using the ISO country code that is specified in the transaction details. We use feature vectors that characterize the client spending in each month to formalize his/her behavior. These vectors contain the information about the amount of money spent in roubles by the client in a particular month in each category. After that, a Yeo-Johnson transformation [23] is applied to these vectors to create a monotonic transformation of the data. The vectors obtained have 88 dimensions, where the first 87 coordinates correspond to 87 categories of MCC codes and the 88th coordinate is the label of the client's location in this month. The goal is to predict the client's location label in the next month. In this problem statement the month has the abroad location label if a client has at least one abroad transaction during this month. It should be mentioned that transactions made in Internet are marked as home transactions.

As for the Raiffeisen dataset, the prediction of a foreign trip is difficult there since only approximately 0.08% of clients took at least one foreign trip during 2017. It is likely that training the model on such a small amount of data may result in a poor forecasting quality. That is why it was decided to predict not the foreign trip but the fact of the transaction in a particular category. The

[1] https://github.com/stavinova/predictability-classes.git

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

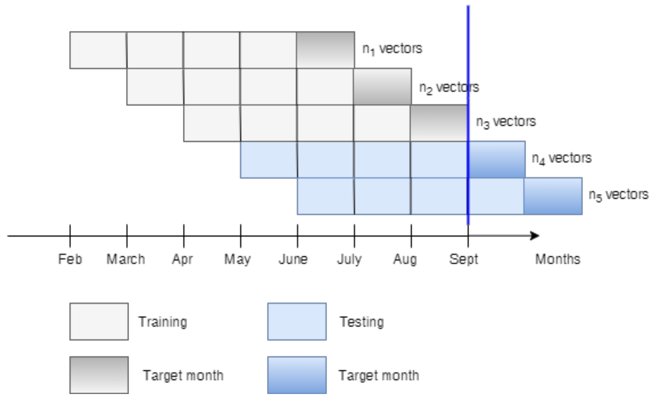


Fig. 1. The time-based cross validation procedure (the months correspond to $D1$).

restaurants category is chosen to be predicted so that the task of forecasting it is rather challenging and, at the same time, there is enough information to successfully complete the training process (the frequency of restaurants-related transactions is 0.41). This goal reformulation let us use the data related to all clients and to train the model of a sufficient quality. Moreover, this task provides some changes in the feature vectors. They contain the information about the number of transactions made by the client in a particular month in each category. No transformation is performed with these vectors and, of course, there is no need in the client's location label, thus, the vectors have 87 dimensions.

Note that in what follows, an event means a foreign trip for the datasets $D1$ and $D2$ and a visit to a restaurant for the Raiffeisen dataset.

After the stage of data preprocessing and feature vectors extraction, all vectors are divided into subsequences of vectors as follows. Each subsequence contains seven vectors (i.e. months) for $D1$ and five vectors in case of $D2$ and the Raiffeisen dataset, all elements of the subsequence describe the spending of one client and this subsequence is continuous during these months. Next, the train period ($D1$: from November 2016 to August 2017, $D2$: from January 2018 to June 2018, Raiffeisen: from January 2017 to June 2017) and the test period ($D1$: from September 2017 to January 2018, $D2$: from July 2018 to December 2018, Raiffeisen: from July 2017 to December 2017) are defined. All subsequences which end in months of the test period are set aside for testing the quality of the model that is trained using the feature vectors from the training period. Model training process is based on the time-based cross validation procedure which is shown in Fig. 1.

The LSTM network is used to predict an event for a client. The number of hidden layers and the number of neurons are determined via several experiments with different settings. Six experiments are conducted with different network configurations: 1, 2, 3 hidden layers with 32, 64 neurons in each hidden layer. A combination of two hidden layers with 64 neurons is selected basing on the

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

performance in the experiments. The learning rate value is set 0.001 basing on another series of experiments.

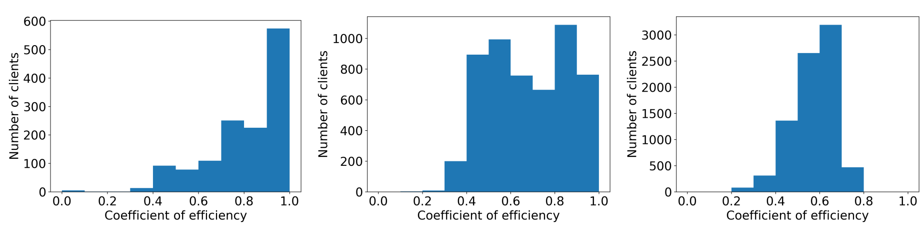
In the current problem statement, the considered event is a foreign trip (or, in the case of the Raifeissen dataset, a visit to a restaurant) which will probably be made by a client in the next month. It is necessary to answer the question: how predictable is this event? Or more precisely, how predictable is the fact of a foreign trip (or a visit to a restaurant) in the next month for a client in terms of the described LSTM model with feature vectors as inputs?

4.2 Predictability measurement

Earlier in this paper, we introduced a measure for estimating the predictability rate which we call the coefficient C. This coefficient is based on the values which are predicted by the model for the event forecasting. Our goal is to calculate the values of this coefficient for each client in all datasets. The coefficient values are computed by the following procedure. Firstly, the model based on the LSTM network is trained on feature vectors from the training period. After receiving the model responses, the values of the coefficient C for each client are calculated using the actual event indicators from the test period, predicted probabilities of the event and the length of the test period for a client. Fig. 2 shows the distribution of coefficient values for the three datasets.

The next step is the creation of predictability classes based on the values of the coefficient C. We decide to divide the clients into classes with the value of C separated by $\mathsf{median}(\mathbf{C})$, where $\mathbf{C}$ is the set of C-values for the entire dataset of clients. This division provides the classes of almost equal size, moreover, this parameter should not be readjusted for different samples. Note that $\mathsf{median}(\mathbf{C}) = 0.87$ for $D1$, $\mathsf{median}(\mathbf{C}) = 0.68$ for $D2$ and $\mathsf{median}(\mathbf{C}) = 0.59$ the Reiffeisen dataset. Two classes, of low and high predictability, are further formed according to the value of the coefficient C. The distribution of the clients by predictability classes is shown in Table 1. The clients from the high predictability class have more predictable behavior in terms of LSTM model since the model answers are closer to the actual situation comparing to those for the low predictability class.

After defining the predictability classes, the study of the forecast quality for each class is conducted. The precision and recall metrics are used for the quality



(a) (b) (c)

Fig. 2. Distribution of the coefficient C values: (a) $D1$, (b) $D2$, (c) Raiffeisen.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

Table 1. The predictability classes obtained

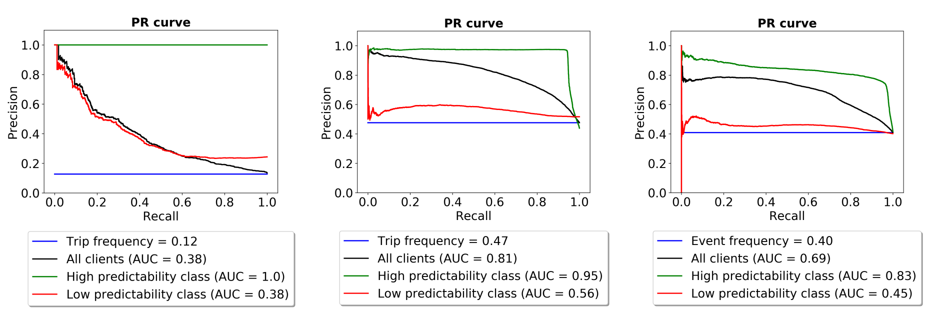
Class name	Coefficient C values	$D1$	$D2$	Raiffeisen
High predictability	$[\mathsf{median}(\mathbf{C}), 1]$	674	2687	4030
Low predictability	$[0, \mathsf{median}(\mathbf{C}))$	675	2687	4031

assessment. Fig. 3 shows the precision-recall curves for the predictability classes described in Table 1. Moreover, the frequency of events in the test set is shown to compare the classifiers quality with the random guessing. For all datasets the proposed classifiers are better than the random guessing. Fig. 3 shows that the clients from $D1$ have far less foreign trips comparing to $D2$ where the average trip frequency is 0.47.

The high predictability class in the dataset $D1$ has perfect prediction quality but the foreign trip frequency for them are almost zero (0.001). That means that almost the half of the $D1$ clients did not take a foreign trip in the test period. Thus, the high predictability class in $D1$ consists of the clients who stayed at home for the test period and are correctly classified by the model, and that is why the curve corresponding to the low predictability class almost coincides with the curve for all the clients.

As for the dataset $D2$, the division into the predictability classes gives us a class with almost perfect precision-recall curve and a class with the curve located a little higher than the random guessing. It should be mentioned that the trip frequencies for these two classes are close to each other (0.43 and 0.51, respectively). This means that the proposed division of $D2$ into these predictability classes is useful. Moreover, it shows that in the case of $D2$ the model is able to predict trips not only by their frequency.

The prediction quality for the Raiffeisen dataset classes is quite similar to the quality of $D2$ classes. The trip frequencies for both predictability classes are close to each other, too.



(a) (b) (c)

Fig. 3. Precision-recall curves for different predictability classes obtained after the model prediction and the forecast for six months ahead: (a) $D1$, (b) $D2$, (c) Raiffeisen.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

Table 2. Confusion matrices for predictability class identification on: (a) $D1$, (b) $D2$, (c) Raiffeisen

(a)

		Forecast	
		High	Low
Actual	High	548	1
	Low	336	267

(b)

		Forecast	
		High	Low
Actual	High	2050	544
	Low	170	2458

(c)

		Forecast	
		High	Low
Actual	High	1048	305
	Low	366	877

4.3 Predictability class identification

Earlier in this paper, we introduced the approach of division the clients into two predictability classes using the coefficient C. That approach assumes that the model predictions are already obtained and the coefficient values are computed after the forecast moment. But now we would like to focus on the task of the client predictability class identification before the forecast moment. For that purpose, we solve the problem of the client's predictability class identification using the data related exclusively to the train period. This problem is solved using the method proposed in the previous section. The confusion matrices that allow us to estimate the Accuracy of the predictability class identification for three different datasets are presented in Table 2. The Accuracy metric of the predictability class identification achieved on the dataset $D1$ is 70.74%, on the second dataset $D2$ is 86.32%, while that on the Raiffeisen dataset is 74.15%.

It is possible to see that the results for $D1$ are too optimistic as the classifier marks 336 clients as high predictable while their actual behavior is poorly predictable. As for $D2$, the results are more balanced and accurate. In the case of the Raiffeisen dataset, the results are quite balanced but the Accuracy is not as high as on $D2$.

4.4 Analysis of the results

To analyze the effectiveness of the division into predictability classes before the prediction moment, we evaluate the prediction quality for classes obtained by the identification algorithm. Fig. 4 shows the prediction quality for the estimated predictability classes compared with the overall prediction quality. To understand the predictability class identification quality, we can compare Fig. 3 and Fig. 4 because they should be as close to each other as possible.

In case of $D1$ the estimated high predictability class became the class with low predictability and drawn in Fig. 4 by the red color. This situation appeared because the high predictability class in the original division consists of the clients who stayed at home during the test period and a few clients who took a trip which is guessed by the model. But now according to Table 2 (a), the estimated high predictability class contains not only high predictable clients but also 336 low predictable clients. In this situation, the trip frequency for this class is increased (0.04) and the model does not guess the correct answer for the new clients. That is why the precision-recall curve is located on the value of 0.04 in this case. The

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

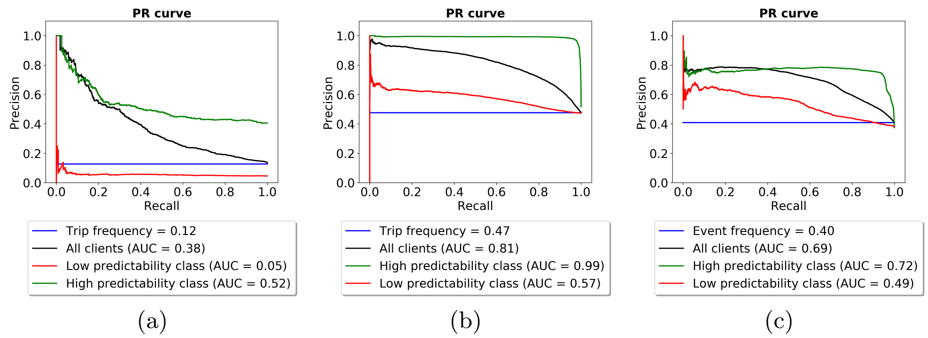


Fig. 4. Precision-recall curve for the estimated classes: (a) $D1$, (b) $D2$, (c) Raiffeisen.

precision-recall curve for another predictability class is situated higher than on Fig. 3 because it becomes smaller than the original one according to Table 2. As for the dataset $D2$, the curves represented in Fig. 3 and Fig. 4 are very close to each other, thus, the class identification algorithm works at the high quality level. Identification results for the Raiffeisen dataset are worse than on $D2$ but still precise enough to distinguish two classes of high and low predictability which can be useful in practise.

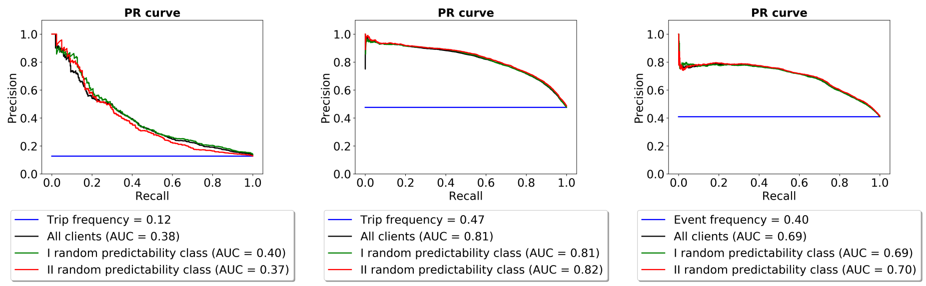
It is worth saying that the task of the event prediction using only the transactional data is highly depends on the quality of the data. This fact is demonstrated by the three different datasets. In the case of $D2$ and Raiffeisen, our approach of estimating the predictability class before the prediction moment can be used to infer a high predictability class with the prediction quality higher than the overall one. In the case of $D1$, we showed that the high predictability class consists of the clients almost without trips whose behavior is guessed by the model.

To demonstrate the effectiveness of the proposed division into predictability classes, we performed the random division of all clients into two classes and estimated the prediction quality for them. The results are shown in Fig. 5. It can be seen that the precision-recall curves in this case are very close to the original one which is obtained for the data without division. It emphasizes the usability of the proposed approach as it suggests a reasonable division of clients while the random division fails.

5 Conclusion

We have proposed an approach for measurement the predictability of an event, for example, a foreign trip or a visit to a restaurant in the next month by a client after obtaining the model predictions. For a predictability estimation we proposed a special measure comparing the actual event indicator and the probability of this event estimated by the model. Using the coefficient values we divided the clients into two groups according to the event predictability for them after the prediction moment. After that, we proposed a classification method for

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16



(a) (b) (c)

Fig. 5. Precision-recall curves for random classes: (a) $D1$, (b) $D2$, (c) Raiffeisen.

the identification of the predictability class for a client before the prediction by the usage of the historical data. The method helps to infer the class of clients whose behavior (events) can be predicted with high quality (on the condition that these clients have a complete transaction history). If we know that a client belongs to one of the predictability classes, we can further estimate the prediction quality for this or another client, similar in behavior to the former, before the actual forecasting.

Our future work will be related to testing our approach in the situations where data concept drift or another non-stationarity is present and where specific forecasting methods should be applied, see e.g. [4, 7, 10, 11]. It seems that in such cases one should be careful about a possible predictability rate drift of the client's behavior, too.

References

1. Chen, S.H., Navet, N.: Failure of genetic-programming induced trading strategies: Distinguishing between efficient markets and inefficient algorithms. In: Computational Intelligence in Economics and Finance, pp. 169–182. Springer (2007)
2. Clements, M., Hendry, D.: Forecasting economic time series. Cambridge University Press (1998)
3. Duan, M.: Time series predictability. Ph.D. thesis, Marquette University (2002)
4. Gama, J.a., Žliobaitundefined, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Comput. Surv. **46**(4) (Mar 2014). https://doi.org/10.1145/2523813, https://doi.org/10.1145/2523813
5. Gers, F.A., Schraudolph, N.N., Schmidhuber, J.: Learning precise timing with lstm recurrent networks. Journal of machine learning research **3**(Aug), 115–143 (2002)
6. Goerg, G.: Forecastable component analysis. In: International Conference on Machine Learning. pp. 64–72 (2013)
7. Janardan, Mehta, S.: Concept drift in streaming data classification: Algorithms, platforms and issues. Procedia Computer Science **122**, 804–811 (2017). https://doi.org/https://doi.org/10.1016/j.procs.2017.11.440, https://www.sciencedirect.com/science/article/pii/S1877050917326881, 5th International Conference on Information Technology and Quantitative Management, ITQM 2017

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16

8. Javed, K., Gouriveau, R., Zemouri, R., Zerhouni, N.: Features selection procedure for prognostics: An approach based on predictability. IFAC Proceedings Volumes **45**(20), 25–30 (2012)
9. Kaboudan, M.: A measure of time series' predictability using genetic programming applied to stock returns. Journal of Forecasting **18**(5), 345–357 (1999)
10. Krawczyk, B., Minku, L.L., Gama, J., Stefanowski, J., Woźniak, M.: Ensemble learning for data stream analysis: A survey. Information Fusion **37**, 132–156 (2017). https://doi.org/https://doi.org/10.1016/j.inffus.2017.02.004, https://www.sciencedirect.com/science/article/pii/S1566253516302329
11. Mehmood, H., Kostakos, P., Cortes, M., Anagnostopoulos, T., Pirttikangas, S., Gilman, E.: Concept drift adaptation techniques in distributed environment for real-world data streams. Smart Cities **4**(1), 349–371 (2021). https://doi.org/10.3390/smartcities4010021, https://www.mdpi.com/2624-6511/4/1/21
12. Moon, G., Hamm, J.: A large-scale study in predictability of daily activities and places. In: MobiCASE. pp. 86–97 (2016)
13. Peña, D., Sánchez, I.: Measuring the advantages of multivariate vs. univariate forecasts. Journal of Time Series Analysis **28**(6), 886–909 (2007)
14. Prelipcean, G., Popoviciu, N., Boscoianu, M.: The role of predictability of financial series in emerging market applications. In: Proceedings of the 9th WSEAS International Conference on Mathematics & Computers in Business and Economics (MCBE'80). pp. 203–208 (2008)
15. Richthofer, S., Wiskott, L.: Predictable feature analysis. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). pp. 190–196. IEEE (2015)
16. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE transactions on Signal Processing **45**(11), 2673–2681 (1997)
17. Stavinova, E., Bochenina, K.: Forecasting of foreign trips by transactional data: a comparative study. Procedia Computer Science **156**, 225–234 (2019)
18. Teodorescu, H.N., Fira, L.I.: Analysis of the predictability of time series obtained from genomic sequences by using several predictors. Journal of Intelligent & Fuzzy Systems **19**(1), 51–63 (2008)
19. Vaganov, D., Funkner, A., Kovalchuk, S., Guleva, V., Bochenina, K.: Forecasting purchase categories with transition graphs using financial and social data. In: International Conference on Social Informatics. pp. 439–454. Springer (2018)
20. Wang, W., Van Gelder, P.H., Vrijling, J.: Measuring predictability of daily streamflow processes based on univariate time series model. In: Proc. iEMSs 4th Biennial Meeting - Int. Congress on Environmental Modelling and Software: Integrating Sciences and Information Technology for Environmental Assessment and Decision Making, iEMSs 2008. pp. 1378–1385 (2008)
21. Wilson, G., Banzhaf, W.: Fast and effective predictability filters for stock price series using linear genetic programming. In: IEEE Congress on Evolutionary Computation. pp. 1–8. IEEE (2010)
22. Wiskott, L., Sejnowski, T.J.: Slow feature analysis: Unsupervised learning of invariances. Neural computation **14**(4), 715–770 (2002)
23. Yeo, I.K., Johnson, R.A.: A new family of power transformations to improve normality or symmetry. Biometrika **87**(4), 954–959 (2000)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_16