Debt Detection in Social Security by Adaptive Sequence Classification*

Shanshan Wu[1], Yanchang Zhao[1], Huaifeng Zhang[2],
Chengqi Zhang[1], Longbing Cao[1], and Hans Bohlscheid[2]

[1] Centre for Quantum Computation and Intelligent Systems (QCIS),
University of Technology, Sydney, Australia
{shanshan,yczhao,chengqi,lbcao}@it.uts.edu.au

[2] Data Mining Section, Payment Reviews Branch, Business Integrity Division,
Centrelink, Australia
{huaifeng.zhang,hans.bohlscheid}@centrelink.gov.au

Abstract. Debt detection is important for improving payment accuracy in social security. Since debt detection from customer transaction data can be generally modelled as a fraud detection problem, a straightforward solution is to extract features from transaction sequences and build a sequence classifier for debts. For long-running debt detections, the patterns in the transaction sequences may exhibit variation from time to time, which makes it imperative to adapt classification to the pattern variation. In this paper, we present a novel adaptive sequence classification framework for debt detection in a social security application. The central technique is to catch up with the pattern variation by boosting discriminative patterns and depressing less discriminative ones according to the latest sequence data.

Keywords: sequence classification, adaptive sequence classification, boosting discriminative patterns.

1 Introduction

In social security, each customer's transactional records form an activity sequence. From a business point of view, these sequences have a close relationship with debt occurrence. Here, a debt indicates an overpayment made by government to a customer who is not entitled to that payment. Debt prevention has emerged as a significant business goal in government departments and agencies. For instance, in Centrelink Australia (http://www.centrelink.gov.au), the Commonwealth Government agency responsible for delivery social security payments and benefits to the Australian community, approximately 63 billion Australian dollars (30% of government expenditure) is distributed each year to 6.4 million customers. Centrelink makes 9.98 million individual entitlement payments and

* This work was supported by the Australian Research Council (ARC) Linkage Project LP0775041 and Early Career Researcher Grant from University of Technology, Sydney, Australia.

D. Karagiannis and Z. Jin (Eds.): KSEM 2009, LNAI 5914, pp. 192–203, 2009.

5.2 billion electronic customer transactions annually [1]. In any year, the debt raised by Centrelink is significant, to say the least. In order to achieve high payment accuracy, it is imperative and important for Centrelink to detect and prevent debts based on the customer activity sequence data. In this paper, we present our research work of debt detection in Centrelink using adaptive sequence classification.

For each customer's activity sequence associated with a debt, it is labelled as *debt*. Contrarily, if no debt happens to a customer, the corresponding sequence is labelled as *normal*. Therefore, debt detection can be generally modelled as a sequence classification problem. Sequence classification has been a focused theme in the data mining research community. Since 1990's, along with the development of pattern recognition, data mining and bioinformatics, many sequence classification models have been proposed. Frequent pattern based classification is one of the most popular methodologies and its power was demonstrated by multiple studies [2,3,4].

Most of the conventional frequent pattern based classifications follow two steps. The first step is to mine a complete set of sequential patterns given a minimum support. The second step is to select a number of discriminative patterns to build a classifier. In most cases, mining for a complete set of sequential patterns on a large dataset is extremely time consuming. The discovered huge number of patterns make the pattern selection and classifier building very time consuming too. In fact, the most important consideration in sequence classification is not finding the complete rule set, but discovering the most discriminative patterns. Recently, more attention has been put on the discriminative frequent pattern discovery for effective classification [5,6]. In this paper, we proposed a novel measure, *contribution weight*, to boost the discriminative patterns. Contribution weight is induced by applying the frequent patterns to a set of evaluation data, and expresses the discriminative power of the patterns regarding to the evaluation data. The interestingness measure of frequent patterns are refined by contribution weight, so as to let the discriminative patterns pop up.

Moreover, sequence data represent the evolvement of data sources, and the sequential patterns generalize the trends of sequences. For long running sequence classification issues, even if the sequences come from the same source, the sequential patterns may vary from time to time. Therefore, the classifier built on a sequence dataset in the past may not work well on the current sequence dataset, not to mention future datasets. For example, based on our previous research work in Centrelink Australia, we found out that the classifier built on transaction data generated from Jul. 2007 to Feb. 2008 does not work quite well on the new data generated from Mar. 2008 to Sep. 2008, due to the changes of policies, economic situation and other social influences. Therefore, it is significant to improve the sequence classification to make it adapt to the sequential pattern variation. The most direct way is to rebuilt the classifier with the latest training dataset. However, the training is quite a time-consuming process, and if the pattern variation is not so much, an incremental updates would be much more efficient than rebuilding the classifier. In this paper, we propose an adaptive sequence classification framework to tackle the above problem. The adaptive model adapts the classifier in a timely fashion by adopting the proposed discriminative pattern boosting strategy, so as to catch up with the trends of sequential pattern variation and improve the classification accuracy.

There are three main contributions in this paper. Firstly, we propose a novel method to boost discriminative frequent patterns for sequence classification, which improves the accuracy of classifier. Secondly, we build up an adaptive sequence classification model which upgrades the sequence classification performance on time-varying sequences. Lastly, our strategies are applied to a real-world application, which shows the efficiency and effectiveness of the proposed methods.

The structure of this paper is as follows. Section 2 provides the notation and description to be used in this paper. Section 3 introduces how the discriminative frequent patterns are boosted. Our proposed adaptive sequence classification framework is given in Section 4. The case study is in Section 5, which is followed by the related work in Section 6. Then we conclude the paper in Section 7.

2 Problem Statement

Let $\mathcal{S}$ be a sequence database, in which each sequence is an ordered list of *elements*. These elements can be either *simple items* from a fixed set of items, or *itemsets*, that is, non-empty sets of items. The list of elements of a data sequence s is denoted by $< s_1, s_2, \cdots, s_n >$, where s_i is the i^{th} element of s.

Consider two sequences $s =< s_1, s_2, \cdots, s_n >$ and $t =< t_1, t_2, \cdots, t_m >$. We say that s is a subsequence of t if s is a "projection" of t, derived by deleting elements and/or items from t. More formally, s is a subsequence of t if there exist integers $j_1 < j_2 < \ldots < j_n$ such that $s_1 \subseteq t_{j1}$, $s_2 \subseteq t_{j2}$, ..., $s_n \subseteq t_{jn}$. Note that for sequences of simple items the above condition translates to $s_1 = t_{j1}$, $s_2 = t_{j2}$, ..., $s_n = t_{jn}$. A sequence t is said to *contain* another sequence s if s is a subsequence of t, in the form of $s \subseteq t$.

2.1 Frequent Sequential Patterns

The number of sequences in a sequence database $\mathcal{S}$ containing sequence s is called the support of s, denoted as $sup(s)$. Given a positive integer min_sup as the support threshold, a sequence s is a frequent sequential pattern in sequence database $\mathcal{S}$ if $sup(s) \geq min\_sup$. The sequential pattern mining is to find the complete set of sequential patterns with respect to a given sequence database $\mathcal{S}$ and a support threshold min_sup.

2.2 Classifiable Sequential Patterns

Let $\mathcal{T}$ be a finite set of *class labels*. A *sequential classifier* is a function

$$\mathcal{F} : \mathcal{S} \rightarrow \mathcal{T} \tag{1}$$

In sequence classification, the classifier $\mathcal{F}$ is built on the base of frequent *classifiable sequential patterns* $\mathcal{P}$.

Definition 2.1 (Classifiable Sequential Pattern). *Classifiable Sequential Patterns (CSP) are frequent sequential patterns for the sequential classifier in the form of $p_a \Rightarrow \tau$, where p_a is a frequent pattern in the sequence database $\mathcal{S}$.*

Based on the mined classifiable sequential patterns, a sequential classifier can be formulized as

$$\mathcal{F} : s \xrightarrow{\mathcal{P}} \tau. \tag{2}$$

That is, for each sequence $s \in \mathcal{S}$, $\mathcal{F}$ predicts the target class label of s based on the sequential classifier built with the classifiable sequential pattern set $\mathcal{P}$. Suppose we have a classifiable sequential pattern set $\mathcal{P}$. A sequence instance s is said to be *covered* by a classifiable sequential pattern $p \in \mathcal{P}$ if s contains the antecedent p_a of the classifiable sequential pattern p.

3 Discriminative Frequent Patterns Boosting

Given a dataset, the more samples a pattern can correctly classifies, the more discriminative the pattern is on the dataset. In other words, the more samples a pattern incorrectly classifies, the less discriminative the pattern is on the dataset. To make it more statistically significant, the definitions of *positive contribution ability* and *negative contribution ability* are given as follows.

Definition 3.1 (Positive Contribution Ability). *Given a dataset S, the Positive Contribution Ability (PCA) of pattern P is the proportion of samples that can be correctly classified by P out of all the samples in dataset S.*

Definition 3.2 (Negative Contribution Ability). *Given a dataset S, the Negative Contribution Ability (NCA) of pattern P is the proportion of samples that are incorrectly classified by P out of all the samples in the dataset S.*

For a classifiable sequential pattern P in the form of $p_a \Rightarrow \tau$, PCA of P on S can be denoted as

$$PCA_S(P) = \frac{\|\{s | p_a \subseteq s \wedge s \in S_\tau\}\|}{\|S\|}, \tag{3}$$

and NCA of pattern P on S can be denoted as

$$NCA_S(P) = \frac{\|\{s | p_a \subseteq s \wedge s \in S_{\neg\tau}\}\|}{\|S\|}, \tag{4}$$

where S_τ and $S_{\neg\tau}$ represent the subsets of S in which samples are of class τ and are not of class τ, respectively.

Above all, PCA and NCA describe the classification ability of patterns on a given dataset. In order to enhance classification performance, it is intuitive to boost the patterns with higher PCA and lower NCA, while depress those with lower PCA and higher NCA. Thereafter, a measure of *Contribution Weight* is proposed to measure the discriminative power that a pattern contributes to the classification on a dataset.

Definition 3.3 (Contribution Weight). *Given a dataset S, Contribution Weight of a classifiable sequential pattern P is the ratio of Positive Contribution*

Ability $PCA_S(P)$ on S and Negative Contribution Ability $NCA_S(P)$ on S. It can be denoted as

$$CW_S(P) = \frac{PCA_S(P)}{NCA_S(P)} = \frac{\|\{s|p \subseteq s \wedge s \in S_\tau\}\|}{\|\{s|p \subseteq s \wedge s \in S_{\neg\tau}\}\|}. \tag{5}$$

The proposed measure of contribution weight tells the relative discriminative power of a classifiable sequential pattern on a given dataset, which is based on the classification performance of the pattern on the dataset. According to the definition, contribution weight has following characters.

- The greater the value of contribution weight is, the more discriminative a pattern is on a given dataset, and vice versa.
- Contribution weight is a measure with regard to a dataset on which classification performance is evaluated.
- Contribution weight is independent of the algorithm that is used for classifiable sequential pattern mining, and it does not matter which interestingness measure is used for classification.

Therefore, we introduce contribution weight as a factor to boost the discriminative frequent patterns on a certain dataset. The term of *Boosted Interestingness Measure* is defined as follows.

Definition 3.4 (Boosted Interestingness Measure). *For a classifiable sequential pattern P with an interestingness measure R, the corresponding Boosted Interestingness Measure on dataset S is denoted as*

$$R_S^* = R \times CW_S(P). \tag{6}$$

In other words, boosted interestingness measure of a pattern can be regarded as a weighted interestingness measure, and the weight tells how much contribution the corresponding pattern can make to the classification on the given dataset. Patterns that are more discriminative on a given dataset are strongly boosted by higher contribution weights, and vice versa. From this point of view, boosted interestingness measure adjusts the original interestingness measure so as to make it indicating the discriminative ability of classifiable sequential patterns on the given dataset more vividly.

4 Adaptive Sequence Classification Framework

In order to catch up with the pattern variation over time, an adaptive sequence classification framework is introduced in this section. The main idea of the adaptive framework is to include the latest pattern into the classifier with the proposed boosted interestingness measure, so as to improve the classification performance on dataset of near future.

As illustrated in Figure 1, the initial classifiable sequential pattern set CSP_0 is extracted from the dataset DS_0, and then is used to perform prediction/ classification on coming dataset DS_1 and get the predicted labels $L_1^{'}$. Once L_1,

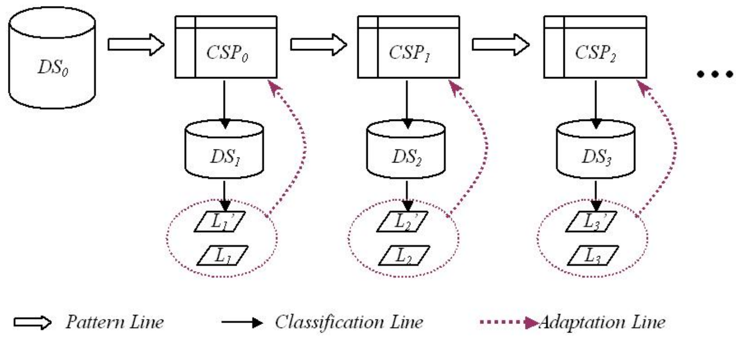

Fig. 1. Architecture of Adaptive Sequence Classification

Algorithm 1. Adaptive classification model.

Data: Dataset DS_i and corresponding real labels L_i that are available after classification/prediction, $i = 0, 1, ...$
Basic classification algorithm$F$(F_1:Classifier construction;F_2:Classifying)

Result: Predicted labels $L_i^{'}$, $i = 1, 2, ...$
Classifiers CSP_i, $i = 0, 1, 2, ...$

begin
  $CSP_0 = F_1(DS_0, L_0)$
  $i = 1$
  while i **do**
    $L_i^{'} = F_2(DS_i, CSP_{i-1})$
    $Wait\ till\ L_i\ is\ available$
    $Modify\ CSP_{i-1}\ with\ R^*_{(DS_i,L_i)}\ to\ get\ CSP_i$
    $i = i + 1$
end

---

the real class labels of dataset DS_1 , is available, interestingness measure of the classifier CSP_0 could be refined and CSP_0 evolves into CSP_1 with boosted interestingness measure, which brings the timely trends of patterns in dataset DS_1 into the classification model. The boosted classifier will be applied to continuously coming dataset for prediction/classification. The procedure goes on as dataset updates all along, which is generalized in Algorithm 1. The boosted classifier CSP_i, $i = 1, 2, ...$ not only takes the latest pattern variation into the classification model, but also tracks the evlovement of the patterns ever since the initial classifier is built. Therefore, the performance of classification is expected to outperform that of the initial classifier.

Since the adaptive model is based on boosted interestingness measure, it inherits the properties of boosted interestingness measure congenitally. To be more precise, it is independent of interestingness measure and classifiable sequence mining method.

5 Case Study

The proposed algorithm has been applied in a real world business application in Centrelink, Australia. The purpose of the case study is to predict and further prevent debt occurrence based on customer transactional activity data. In this section, the dataset used for debt prediction in Centrelink is described firstly. Then a pre-experiment is given to evaluate the effectiveness of discriminative pattern boosting strategy, followed by the experimental results of adaptive sequence classification framework.

5.1 Data Description

The dataset used for sequence classification is composed of customer activity data and debt data. In Centrelink, every single contact (e.g., a circumstance change) of a customer may trigger a sequence of activities running. As a result, large volumes of activity based transactions are encoded into 3-character "Activity Code" and recorded in activity transactional files. In the original activity transactional table, each activity has 35 attributes, in which 4 attributes are used in the case study. These attributes are "CRN" (Customer Reference Number) of a customer, "Activity Code", "Activity Date" and "Activity Time", as shown in Table 1. We sort the activity data according to "Activity Date" and "Activity Time" to construct the activity sequence. The debt data consist of the "CRN" of the debtor and "Debt Transaction Date". In our case study, only the activities of a customer before the occurrence of his/her debt are kept for the sequence classification.

Table 1. Centrelink Data Sample

CRN	Act_Code	Act_Date	Act_Time
******002	DOC	20/08/07	14:24:13
******002	RPT	20/08/07	14:33:55
******002	DOC	05/09/07	10:13:47
******002	ADD	06/09/07	13:57:44
******002	RPR	12/09/07	13:08:27
******002	ADV	17/09/07	10:10:28
******002	REA	09/10/07	7:38:48
******002	DOC	11/10/07	8:34:36
******002	RCV	11/10/07	9:44:39
******002	FRV	11/10/07	10:18:46
******002	AAI	07/02/08	15:11:54

Table 2. Data Windows

Window	Start Date	End Date
W0	02/07/07	31/10/07
W1	01/08/07	30/11/07
W2	01/09/07	31/12/07
W3	01/10/07	31/01/08
W4	01/11/07	29/02/08
W5	01/12/07	31/03/08
W6	01/01/08	30/04/08
W7	01/02/08	31/05/08
W8	01/03/08	30/06/08
W9	01/04/08	31/07/08
W10	01/05/08	31/08/08

5.2 Effectiveness of Boosting Discriminative Patterns

In order to evaluate the effectiveness of discriminative patterns boosting, two groups of experiments are presented in this section. In both groups, we compare the performance of classification which uses discriminative pattern boosting strategy with that does not boost discriminative patterns. In group (a), the activity sequence data generated from Jul. 2007 to Oct. 2007 are used. After data cleaning, there are 6, 920 activity sequences including 210, 457 activity records used. The dataset is randomly divided into the following 3 portions.

– Training data(60%): To generate the initial classifier.
– Evaluation data(20%): To refine classifier.
– Test data(20%): To test the performance of classification.

While in group (b), some data generated in Nov. 2007 is added to the evaluation data and test data, expecting to include some pattern variation.

According to the property of contribution weight, the boosted interestingness measure is independent of basic classification. Therefore, we use the classification algorithm proposed in our previous work [13] to generate the initial classifier on the training dataset. And we use confidence as the base interestingness measure. For classification which uses boosting strategy, the evaluation dataset is used to refine the initial classifier, and the refined classifier is evaluated on the test dataset. While for the classification that does not boost discriminative patterns, we combine training data and evaluation data to generate the initial classifier, and then apply the initial classifier to the test dataset for debt prediction.

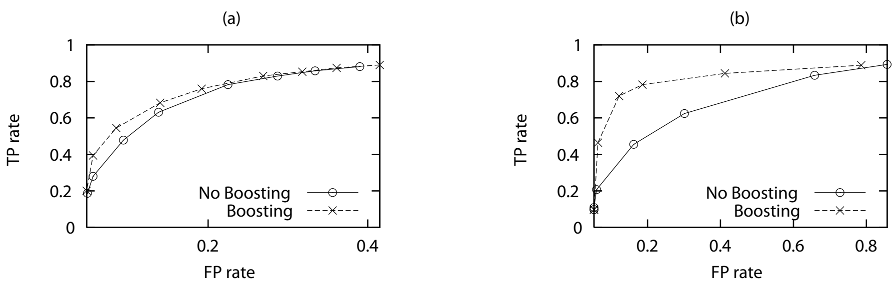

Fig. 2. Effectiveness of Discriminative Patterns Boosting

ROC curve (Receiver Operating Characteristic) is used to plot the fraction of true positives *vs.* the fraction of false positives of each classifier. The best possible prediction method would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity(no false positives). Therefore, the more close to the upper left corner the curve is, the better the classification method is. As illustrated in Fig. 2, the boosted classifier outperforms the classifier without boosting in both experiments. In group (a), training data, evaluation data and test data all come from the dataset generated in the same time period. By boosting discriminative patterns with evaluation data, classification power of initial classifier is refined by boosting discriminative patterns and depressing less discriminative patterns, so it outperforms the classification without boosting. As for group (b), since some new data generated in different time period is added to the evaluation data and test data, some pattern variation might be included in the corresponding dataset. In this circumstance, the proposed boosting strategy notices the pattern variation in the updated dataset, refines the interestingness measure of the classifiers with evaluation data, and performs much better in the test data than the classifier without boosting.

In all, the discriminative pattern boosting strategy improves the classification performance, especially when the sequence data evolves with pattern variation.

5.3 Performance of Adaptive Sequence Classification Framework

In this subsection, we will evaluate the adaptive sequence classification framework on the sequence datasets obtained with a sliding window applied on the activity sequence data. After applying sliding window on the sequences generated from Jul. 2007 to Aug. 2008, we get 11 windows listed in Table 2.

Following the framework proposed in Section 4, the classification in our previous work [13] is firstly applied on W_0 and the initial classifier CSP_0 is generated. By discriminative pattern boosting with W_1, CSP_0 is refined to CSP_1 and then is applied to make debt prediction on W_2. Here we still use confidence as the base interestingness measure. The debt prediction performance on W_2 is illustrated in the first graph in Fig. 3. Thereafter, CSP_1 is boosted with sequence data in W_2, and the generated CSP_2 is applied on W_3 to predict debt occurrence. As the procedure goes on continuously, the debt prediction performance on all the

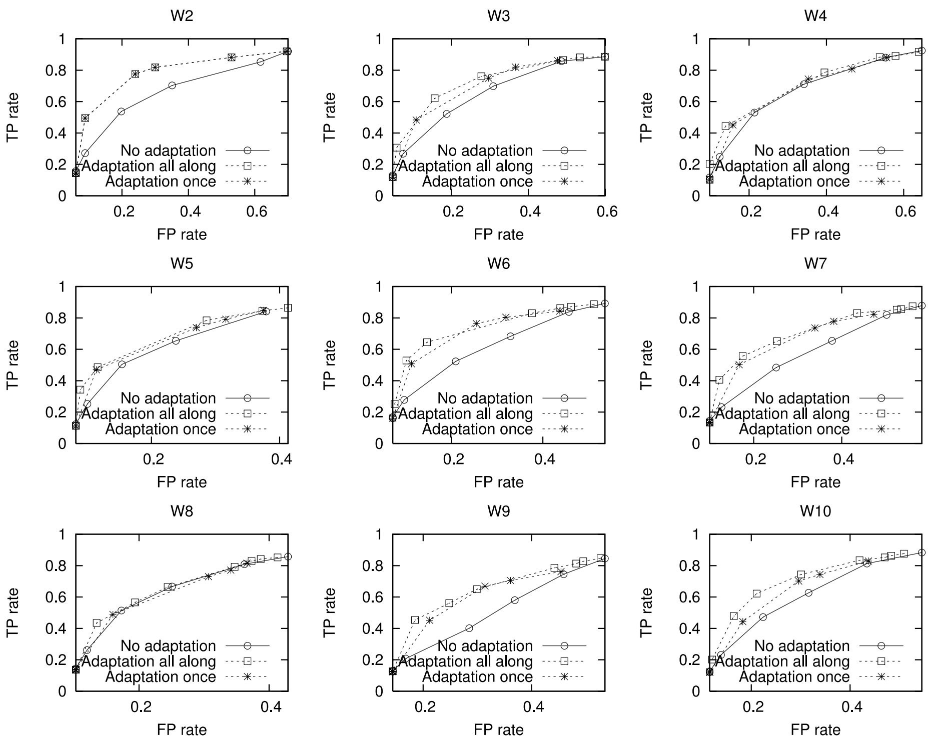

Fig. 3. RoC curves of Adaptive Sequence Classification Framework

following windows are listed in Fig. 3, which is represented by the ROC curves labelled *Adaptation_all_along*. In order to evaluate the performance of adaptive sequence classification framework, debt prediction on each window is also performed with initial classifier CSP_0, whose performance is denoted by the ROC curves labelled *No_adaptation*. According to Fig. 3, we can tell that the proposed adaptive framework outperforms the initial classifier in debt prediction on continuously coming datasets. Since the classifier is continuously updated with the latest data, it catches up the pattern variation in the new dataset and then works well on the debt prediction on the oncoming dataset. Meanwhile, we apply CSP_1, which is boosted once based on initial classifier, to each of the windows and get the performance denoted by the curves labelled *Adaptation_once*. The classifier boosted once still outperforms the initial classifier. While it does not contains the pattern information in the latest datasets, its performance is always worse than that of *Adaptation_all_along* strategy.

Above all, conclusion could be drawn that our proposed adaptive sequence classification framework updates the classifier with new data, includes the sequence pattern variation in the new data, and performs effectively on the continuously arriving data.

6 Related Work

6.1 Sequence Classification

There have been several researchers working towards building sequence classifiers based on frequent sequential patterns. Lesh et al. [2] proposed an algorithm for sequence classification using frequent patterns as features in the classifier. In their algorithm, subsequences are extracted and transformed into sets of features. After feature extraction, general classification algorithms such as Naïve Bayes, SVM or neural network can be used for classification. Their algorithm is the first try on the combination of classification and sequential pattern mining. However, a huge amount of sequential patterns are mined in the sequential mining procedure. Although pruning algorithm is used for the post-processing, there are still a large amount of sequential patterns constructing the feature space. Tseng and Lee [3] proposed an Classify-By-Sequence (CBS) algorithm to combine sequential pattern mining and classification. In their paper, two algorithms, CBS_Class and CBS_All were proposed. In CBS_Class, the database is divided into a number of sub-databases according to the class label of each instance. Then sequential pattern mining is implemented on each sub-database. In CBS_All, conventional sequential pattern mining algorithm is used on the whole dataset. Weighted scoring is used in both algorithms. Exarchos [4] proposed to combine sequential pattern mining and classification followed by an optimization algorithm. The accuracy of their algorithm is higher than that of CBS. However optimization is a very time-consuming procedure.

6.2 Adaptive Classification

Some applications similar to debt detection are financial crime detection, network intrusion detection and spam detection. Bonchi et al. [7] proposed a classification-based methodology for planning audit strategies in fraud detection. The models are constructed by analyzing historical audit data. Then the models are used to plan future audits for the detection of tax evasion. A decision tree algorithm, C5.0, was used in their case study. Although the target problem is similar with ours, the data used are different. What we used are transactional data that record activities related to customers. Because the time order in activities is important for predicting debt occurrences, sequence classifiers instead of decision trees are used in our application. Rosset et al. [8] studied the fraud detection in telecommunications and presented a two-stage system based on C4.5 to find fraud rules. They adapted the C4.5 algorithm for generating rules from bi-level data, i.e., customer data and behaviour-level data. However, the behaviour data they used is the statistics in a short time frame, which is different from sequential patterns in our techniques.

Fawcett and Provost [9] addressed the issue of adaptive fraud detection in telecommunication services. They use a rule-learning program to uncover indicators of fraudulent behaviour from a large database of customer transactions. Then the indicators are used to create a set of monitors, which profile legitimate customer behaviour and indicate anomalies. Xu et al. [10] utilize FP-tree growth algroithm to extract the associations among features from transactions during a certain period in order to profile the user's behavior adaptively. Lu et al. [11] used deviations from the expected Benfords Law distributions as an indicators of anomalous behaviour that are strong indicators of fraud. The adaptive Benfords Law adapts to the incomplete data records, compared with classic Benfords Law. Lee et al. [12] used meta-learning mechanism by combining existing models with new models trained on new intrusion data or new normal data to make intrusion detection models adaptive. Different from transactional fraud detection that attempts to classify a transaction or event as being legal or fraud, our techniques try to predict the likelihood of a customer being fraud based on his/her past activities. It is at customer level instead of transaction level.

7 Conclusion and Future Work

In this paper we proposed a novel adaptive sequence classification framework for long running sequence classification in the circumstance of time-varying sequence patterns. In order to make the classifier catch up with the latest sequence pattern variation, we introduce a discriminative pattern boosting strategy, which boosts discriminative patterns and depresses less discriminative patterns based on the latest sequential data. The proposed methods are tested on a real-world dataset, and the case study shows the effectiveness of the proposed strategy.

Our current adaptive framework refines the classifier round by round, and in each round the adaptation is based on the classifier generated in the last round. Though it tracks the evolvement of sequential patterns, the latest pattern variation is given the same consideration as the previous ones. In our future work, we will study how to put tilted weight to the historical data, which may include more latest sequence pattern characteristics into the classification model.

Acknowledgements. We are grateful to Mr. Peter Newbigin and Mr. Brett D. Clark from Payment Reviews Branch, Business Integrity Division, Centrelink, Australia, for extracting data, providing domain knowledge and helpful suggestions.

References

1. Centrelink. Centrelink annual report 2004-2005. Technical report, Centrelink, Australia (2005)
2. Lesh, N., Zaki, M.J., Ogihara, M.: Mining Features for Sequence Classification. In: Proc.of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, California, USA, August 1999, pp. 342–346 (1999)
3. Tseng, V.S.M., Lee, C.-H.: CBS: A new classification method by using sequential patterns. In: Proc. of SIAM International Conference on Data Mining (SDM 2005), pp. 596–600 (2005)
4. Exarchos, T.P., Tsipouras, M.G., Papaloukas, C., Fotiadis, D.I.: A two-stage methodology for sequence classification based on sequential pattern mining and optimization. Data and Knowledge Engineering 66(3), 467–487 (2008)
5. Cheng, H., Yan, X., Han, J., Hsu, C.-W.: Discriminative Frequent Pattern Analysis for Effective Classification. In: Proc. of IEEE International Conference on Data Engineering (ICDE 2007), April 2007, pp. 716–725 (2007)
6. Han, J., Cheng, H., Xin, D., Yan, X.: Frequent pattern mining: current status and future directions. Data Mining and Knowledge Discovery 15(1), 55–86 (2007)
7. Bonchi, F., Giannotti, F., Mainetto, G., Pedreschi, D.: A classification-based methodology for planning audit strategies in fraud detection. In: Proc. of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 1999, pp. 175–184 (1999)
8. Rosset, S., Murad, U., Neumann, E., Idan, Y., Pinkas, G.: Discovery of fraud rules for telecommunications - challenges and solutions. In: Proc. of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 1999, pp. 409–413 (1999)
9. Fawcett, T., Provost, F.: Adaptive fraud detection. Data Mining and Knowledge Discovery 1, 291–316 (1997)
10. Xu, J., Sung, A.H., Liu, Q.: Tree based behavior monitoring for adaptive fraud detection. In: Proc. of the 18th International Conference on Pattern Recognition, Washington, DC, USA, pp. 1208–1211 (2006)
11. Lu, F., Boritz, J.E., Covvey, D.: Adaptive fraud detection using Benfords Law. In: Lamontagne, L., Marchand, M. (eds.) Canadian AI 2006. LNCS (LNAI), vol. 4013, pp. 347–358. Springer, Heidelberg (2006)
12. Lee, W., Stolfo, S.J., Mok, K.W.: Adaptive intrusion detection: a data mining approach. Artificial Intelligence Review 14, 533–567 (2000)
13. Zhang, H., Zhao, Y., Cao, L., Zhang, C., Bohlscheid, H.: Customer activity sequence classification for debt prevention in social security. Accepted by Journal of Computer Science and Technology (2009)