Neurocomputing 132 (2014) 3–13

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Customer profile classification: To adapt classifiers or to relabel customer profiles?

Edward Apeh [a,*], Bogdan Gabrys [a], Amanda Schierz [b]

[a] Smart Technology Research Centre, School of Design, Engineering and Computing, Bournemouth University, Dorset BH12 5BB, United Kingdom
[b] The Institute of Cancer Research: Royal Cancer Hospital, 15 Cotswold Road, Belmont, Sutton, Surrey SM2 5NG, United Kingdom

ARTICLE INFO

Article history:
Received 30 October 2012
Received in revised form
14 July 2013
Accepted 16 July 2013
Available online 1 December 2013

Keywords:
Adaptive algorithms
Data mining
Decision support systems
Classification algorithms
Industry applications
Labelling

ABSTRACT

Customer profiles are, by definition, made up of factual and transactional data. It is often the case that due to reasons such as high cost of data acquisition and/or protection, only the transactional data are available for data mining operations. Transactional data, however, tend to be highly sparse and skewed due to a large proportion of customers engaging in very few transactions. This can result in a bias in the prediction accuracy of classifiers built using them. The problem is even more so when identifying and classifying changing customer profiles whose classification may change either due to a concept drift or due to a change in buying behaviour. This paper presents a comparative investigation of 4 approaches for classifying dynamic customer profiles built using evolving transactional data over time. The changing class values of the customer profiles were analysed together with the challenging problem of deciding whether to change the class label or adapt the classifier. The results from the experiments we conducted on a highly sparse and skewed real-world transactional data show that adapting the classifiers leads to more stable classification of customer profiles in the shorter time windows; while relabelling the changed customer profile classes leads to more accurate and stable classification in the longer time windows.

1. Introduction

The problem of identifying different types of customers in order to adequately meet their product/service needs is of great importance to businesses who want to remain competitive in the current economic climate.

Algorithms from data mining have been used extensively for explorative and predictive purposes to model business problems. For instance, clustering techniques are often the first techniques used in market segmentation [1] while classification techniques such as *K*-NN have been used in customer profile personalization in on-line retail systems [2].

Pre-processing techniques are used in almost all data mining applications to efficiently perform the data mining task. The decision to apply a pre-processing technique may be driven by the need to generate a model from a dataset that is too large to process in full (data reduction) [3], handle missing values/inconsistent data (data cleansing) [4], combine data from multiple sources into a coherent store (data integration) [5], and normalize data so that it can be more efficiently processed (data transformation) [3].

The type of pre-processing techniques employed in identifying and classifying customer profiles based on transactional data needs to be carefully designed as customers with sparse transactions, which tend to make up the bulk of transactional data, are difficult to distinguish and accurately classify. This problem is even more pronounced when the customers' transactional data is evolving and the sparse transactions are mixed with dense transactions, as the classifier performance tends to be biased towards the larger number of customers with sparse transactions.

Our work in [6] presented an investigation of a data mining approach that combines the unsupervised data binning pre-processing technique with classification to identify different types of customer profiles using their transactions. Our proposed approach groups customer profiles into bins on the basis of the number of items transacted over 30 months so as to more accurately and confidently classify a customer given their transactions.

The experiments we conducted using static classifier models built using the transactional data aggregated over 30 months showed that classifier models based on customers who bought more items performed better than classifier models based on customers who bought fewer items.

However, it is often the case that a business needs to identify their customers as soon as possible and might not have the luxury

* Corresponding author. Tel.: +447550643038.
E-mail addresses: eapeh@bournemouth.ac.uk (E. Apeh), bgabrys@bournemouth.ac.uk (B. Gabrys), aschierz@icr.ac.uk (A. Schierz).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.07.048

of time to build classifier models of customer profiles based on transactional data accumulated over a long period of time.

Meeting the need for a timely identification of reliable and stable customers is challenging due to the dynamic nature of customer profiles built using transactional data.

In this paper we extend our work in [6] by further investigating the problem of classifying evolving customer profiles based on transactional data and the challenging problem of discerning between the change in the customer profile (which may necessitate the effective change of the customer's label) and the change in the performance of the model(s) (which may necessitate changing or adapting the model(s)).

The problems of sparseness and skewness inherent in transactional data which make mining transactional data challenging are described in greater detail and our use of data binning algorithm in handling the sparseness and skewness in transactional data is also further described.

A comparative investigation of 4 approaches for classifying dynamic customer profiles built using evolving transactional data over time is also presented.

The rest of this paper is organized as follows: Section 2 commences with the description of the problem of constructing customer profiles from transactional data, and provides background knowledge of our work. An overview of the problem of concept drift and label switching in the context of customer profile classification is presented in Section 2.3. Section 2 concludes with an overview of the techniques for adaptation and relabelling together with related works covered in Sections 2.4 and 2.5 respectively.

Section 3 gives a description of our proposed approach while Section 4 gives a description of the data used for the proof-of-concept experiments as well as the experiment evaluation measures used. The results of the experiments in which the effect of model adaptation on classification performance in contrast to the effect of re-labelling are also presented and analysed in Section 4.

The paper concludes in Section 5 with a discussion of the problem of model adaptation versus instances re-labelling in a dynamic transactional data setting. In the context of customer profile classification using transactional data, this introduces the directions for future work in addressing the important research question of whether to adapt (or change) the model or to incorporate a new labelling scheme.

2. Problem description and background knowledge

The construction of profiles for individual customers is a major concern for businesses that desire to build and effectively manage the relationship they have with their valued customers. Well constructed customer profiles provide businesses with vital information such as [7]:

1. who their valuable customers are, and
2. how they behave.

Classifying the constructed customer profiles into these insightful and valuable groups is however challenging as the customer profiles and their constituents are constantly changing. For example, a customer profile of Gardening may undergo a considerable change in value depending on the season or the change in individual customer's buying behaviour.

This paper presents an approach for analysing transactions of customer profiles for the purpose of classification. For our experiments, we use verified customer profiles for Electricians and Plumbers/Heat fitters (referred to henceforth as PlumbHeaters) sourced from Screwfix – a large UK retailer.

2.1. Problem statement

Formally, given a set of transactions $\mathbf{T}$, containing N transactions categorized into d product item topics:

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,d} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,d} \\ \vdots & \vdots & \cdots & \vdots \\ t_{N,1} & t_{N,2} & \cdots & t_{N,d} \end{bmatrix}$$

we define a set of M customer profiles,

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,d} \\ \vdots & \vdots & \cdots & \vdots \\ p_{i,1} & p_{i,2} & \cdots & p_{i,d} \\ \vdots & \vdots & \cdots & \vdots \\ p_{M,1} & p_{M,2} & \cdots & p_{M,d} \end{bmatrix}$$

with

$$p_{i,j} = \sum_{k=i_1}^{i_{n_i}} t_{k,j}, \quad \forall i = 1, \ldots, M \ \forall j = 1, \ldots, d$$

where $k \in \{i_1, i_2, \ldots, i_{n_i}\}$ is a set of indexes referring to the n_i transactions of the i-th customer in the set of transactions $\mathbf{T}$.

Goal: We seek to build classifiers using distinctive groups (defined by the number of items purchased), for which the predictive error of classifying unseen customer profiles over time is minimal.

2.2. The issue with transactional data

Transactional data apart from being traditionally large is also inherently sparse, due mainly to the underlying process from which they are generated. For example, in retail transactional data, where it is usual for customers to purchase only a very small fraction of products, the average size of a basket (i.e. the collection of items that a customer purchases in a typical transaction) might contain just 3–4 products out of 1000s of products in the retailer's catalogue/inventory. Such a transaction when represented in an attribute-vector representation will have an average of 3–4 out of 1000s of product attributes that are not null. This implies that the fraction of non-zero attributes on the table (i.e. the sparsity factor) will be 3/1000–4/1000, or 0.3–0.4%. The sparsity factor of the transactional data used in our work here can be seen from the relationship between the number of transactions and the number of items transacted in Table 2.

Pre-processing techniques such as sampling, clustering, and data binning [3]. are often used to prepare the dataset used in building and maintaining data mining based business models.

For sparse data, conventional sampling may not work well, because most of the samples are zeros [8]. Likewise, sampling fixed dataset columns from the dataset, as is done in some cases [9], is also inflexible because different rows may have very different sparsity factors leading to each sampled data instance conveying little or no information for accurate inference.

In order to address the issue of sparsity present at an individual transaction level in our work here, the transactions over the 30 months period for a given customer were aggregated.

2.3. Concept drift and customer profile class switching problem

Customer profiles based on real-world transactional data tend to change over time as customers change their buying behaviour in reaction to the change in their circumstances, the market, the business, etc.