Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications 36 (2009) 155–164

Expert Systems with Applications

www.elsevier.com/locate/eswa

Mining changing customer segments in dynamic markets

Mirko Böttcher [b,*], Martin Spott [a], Detlef Nauck [a], Rudolf Kruse [b]

[a] Intelligent Systems Research Centre, BT Group plc, Adastral Park, IP5 3RE Ipswich, United Kingdom
[b] Institute for Knowledge and Language Engineering, Otto-von-Guericke-University of Magdeburg, Germany

Abstract

Identifying customer segments and tracking their change over time is an important application for enterprises who need to understand what their customers expect from them – now and in the future. This in particular is important for businesses which operate in dynamic markets with customers who, driven by new innovations and competing products, have highly changing demands and attitudes. Customer segmentation is typically done by applying some form of cluster analysis to obtain a set of segments to which future customers are assigned to. In this paper, we present a system for customer segmentation which accounts for the dynamics of today's markets. It employs an approach based on the discovery of frequent itemsets and the analysis of their change over time which, finally, results in a change-based notion of segment interestingness. Our approach allows us to detect arbitrary segments and analyse their temporal development. Thereby, our approach is assumption-free and pro-active and can be run continuously. Newly discovered segments or relevant changes will be reported automatically based on the application of several interestingness measures.

Keywords: Data mining; Frequent itemsets; Customer segmentation; Change mining

1. Introduction

Businesses, especially in the service industry, need to understand their customers in order to serve them best. Understanding customers involves collecting as much data as possible about interactions between customers and the business, analyse this data to turn it into information and finally learn from it and take action. This process is supported by techniques from data warehousing, data quality management, knowledge discovery in databases (or data mining), business intelligence, business process management, etc. In this paper, we will look at a particular aspect of the analytical process – the discovery of changing customer segments.

When we hear about customer segments, we typically think about marketing-driven demographic groups that are defined using a great deal of domain understanding. This approach requires typically running extensive surveys on a significant part of the customer base to learn about their preferences, views, standard of living, consumer behavior, etc. Based on domain understanding a number of segments are then identified and customers are assigned to segments based on some similarity measure. Typically, approaches from cluster analysis are used to initially identify groups in the data which are then interpreted as potential customer segments. The whole process is based on manual analysis and is typically expectation and goal driven. In a nutshell, you would detect the segments you are looking for.

The difficulties of this approach are threefold. Firstly, the employed analytics – clustering – require an underlying similarity measure which typically reduces the data to numeric features. Cluster analyses that can work with symbolic attributes do exist, but are less well known and

* Corresponding author.
E-mail addresses: miboettc@iws.cs.uni-magdeburg.de (M. Böttcher), Martin.Spott@bt.com (M. Spott), Detlef.Nauck@bt.com (D. Nauck), Kruse@iws.cs.uni-magdeburg.de (R. Kruse).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.09.006

typically not supported by commercially available software. The existence of a similarity measure is required, otherwise neither cluster analysis can be applied nor can customers be assigned to clusters.

Secondly, assigning customers to segments is a problem, because survey data that was used to form segments is not available for the vast majority of customers. That means customers are assigned into segments by available information about them which at best contains data about the products and services they use but at worst is based only on rather inadequate data like the postcode, for example.

Thirdly, the segmentation approach is to a large extent goal driven and static. That means the data that is used has been collected with the assumption that it is ultimately relevant for segmentation. Data or attributes not considered to be relevant are dropped from the analytical process early on to make cluster analysis feasible and aid the interpretation of detected clusters which is essential to form meaningful segments. The danger of this approach is that potentially relevant features are ignored meaning certain segments may not be detected. The approach is also static, which means that once segments have been established change in those segments is not monitored because of the practical repercussions like regularly running expensive surveys, etc. This results in missing important trends, threats and opportunities because segments can change in several ways. New groups can appear, disappear, merge, move, shrink or grow.

A promising approach would be to concentrate on data that is actually relevant in describing the relationship between customers and the business, i.e. data about interactions with customers and their usage profile of products and services. The data would be a mixture of symbolic data, like product types, fault codes, complaint reasons, etc., and numeric data on different scales like counts, costs, revenues, frequencies, etc. If data types have to be consolidate it is typically better to discretise numerical data and loose some information instead of turning symbolic values into numbers and thus introducing spurious information like distances and relations.

In this paper, we are looking at using frequent itemset discovery for detecting *interesting* segments in data. We define interesting segments as segments that display some temporal change reflected in the data. We relate growing or shrinking segments to threats and opportunities the business must know about. We explain how tracking the temporal changes of an itemset's support can lead to a notion of interestingness. We will illustrate our approach by applying it to two data sets from customer surveys and network usage.

2. Related work

Customer segmentation is the process of dividing customers into homogeneous groups on the basis of common attributes. In most application customer segmentation is accomplished by defining numerical attributes which describe a customer's value based on economical and market considerations. Cluster algorithms are then commonly employed in order to discover groups of customers with similar attribute values. For example, in Shin and Sohn (2004), three different clustering algorithms are compared to segment stock trading customers based on their amount of trade in different trading scenarios. Segmentation methods based on clustering require a user to carefully select the used attributes by hand in a tedious process. Since the number of used attributes is rather low, commonly only two or three, the analysis of segment change can still be done manually. This might be the reason why to our knowledge no automated approach has been published yet.

Several approaches have been proposed to analyse changes in customer behavior, for instance in retail marketing (Chen, Chiu, & Chang, 2005), in an internet shopping mall (Kim, Song, Kim, & Kim, 2005; Song, Kim, & Kim, 2001) and in an insurance company (Liu, Hsu, Han, & Xia, 2000). These approaches typically compare two sets of rules generated from datasets of two different periods. For rule representation either decision trees (Kim et al., 2005; Liu et al., 2000) or association rules (Chen et al., 2005; Song et al., 2001) are used. For example, in a telecommunication retail application, such approaches may detect that customers used to order a certain tariff with a certain special option – now they still order this tariff, but seldom with the special option. The aforementioned approaches only detect *what* has changed rather than *how* something changes. Picking up on the last example this means, they cannot spot the declining trend in the ordered special options. Spotting trends, however, is crucial for many cooperations.

In the area of association rules, the discovery of interesting changes has been studied by several authors. In Agrawal and Psaila (1995), a query language for shapes of histories is introduced. Liu, Ma, and Lee (2001) propose a statistical approach to distinguish trend, semi-stable and stable rules with respect to their histories of confidence and support. A fuzzy approach to reveal the regularities in how measures for rules change and to predict future changes was presented by Au and Chan (2005). A framework to monitor the changes in association rule measures based on simple thresholds for support and confidence is described in Spiliopoulou, Baron, and Günther (2003). The issue of how to detect and discard rules which are redundant with respect to their history has been addressed by only a few number of publications. In Liu, Hsu, and Ma (2001), a method to detect so-called *fundamental rule changes* between two time periods is presented. In Böttcher, Spott, and Nauck (2005), it is shown that the approach proposed in Liu et al. (2001) has several short-comings with regard to applicability, reliability and understandability. For this reason, they introduce an alternative approach based on so-called *derivative histories* which detects temporally redundant rules.

3. Preliminaries

3.1. Frequent itemsets

We define a customer segment as a set of customers which have common features or attributes. Given a data set which describes customers any attribute value combination of each subset of its attributes therefore qualifies as a candidate customer segment. However, we are only interested in customer segments which are frequent in relation to the overall population. This means, we do not aim for segments which present only a tiny fraction of customers, but for those which are larger than a user defined frequency threshold.

Customer segments defined this way can be represented by *frequent itemsets*. The discovery of frequent itemsets is a broadly used approach to perform a nearly exhaustive search for patterns within a data set (Agrawal, Imielinski, & Swami, 1993). Its goal is to detect all those attribute values which occur together within a data set and whose relative frequency exceeds a given threshold. The advantage of frequent itemset discovery is the completeness of its results: it finds the exhaustive set of all significant patterns. For this reason, it provides a rather detailed description of a data set's structure. On the other hand, however, the set of discovered itemsets is typically vast.

Formally, frequent itemset discovery is applied to a set $\mathscr{D}$ of *transactions* $\mathscr{T} \in \mathscr{D}$. Every transaction $\mathscr{T}$ is a subset of a set of items $\mathscr{L}$. A subset $\mathscr{X} \subseteq \mathscr{L}$ is called *itemset*. It is said that a transaction $\mathscr{T}$ *supports* an itemset $\mathscr{X}$ if $\mathscr{X} \subseteq \mathscr{T}$. As usual, the frequency of an itemset $\mathscr{X}$ is measured by its *support* $\text{supp}(\mathscr{X})$ which estimates $P(\mathscr{X} \subseteq \mathscr{T})$, or short $P(\mathscr{X})$. For example, suppose that we are given a data set, which contains survey results about customer satisfaction, the following frequent itemset could have been discovered from it:

$$\text{Age} > 50, \quad \text{Satisfied} = \text{Yes}$$

The support of this itemset is the relative frequency of customers that are over 50 years old and satisfied, i.e., it describes the relative size of the customer segment.

In the following, we will use the notions of a customer segment and a frequent itemset synonymously.

3.2. Support histories

The underlying idea of our system is to detect interesting changes in a customer segment, represented by an itemset, by analysing the support of the itemset along the time axis. The starting point of such an approach is as follows: a timestamped data set is partitioned into intervals along the time axis. Frequent itemset discovery is then applied to each of these subsets. This yields sequences – or *histories* – of support for each itemset, which can be analysed further. Of particular interest are regularities in the histories which we call *change patterns*. They allow us to make statements about the future development of a customer segment and thus provide a basis for proactive decision making.

Let $\mathscr{D}$ be a time-stamped data set and $[t_0, t_n]$ the minimum time span that covers all its tuples. The interval $[t_0, t_n]$ is divided into $n > 1$ non-overlapping periods $T_i := [t_{i-1}, t_i]$, such that the corresponding subsets $\mathscr{D}(T_i) \subset \mathscr{D}$ each have a size $|\mathscr{D}(T_i)| \gg 1$. Let $\widehat{T} := \{T_1, \ldots, T_n\}$ be the set of all periods, then for each $T_i \in \widehat{T}$ frequent itemset discovery is applied to the transaction set $\mathscr{D}(T_i)$ to derive itemsets $\mathscr{I}(\mathscr{D}(T_i))$.

Because the support of every itemset $\mathscr{X}$ is now related to a specific transaction set $\mathscr{D}(T_i)$ and thus to a certain time period T_i we need to extend its notation. This is done straightforward and yields $\text{supp}(\mathscr{X}, T_i) \approx P(\mathscr{X}|T_i)$. Each itemset $\mathscr{X} \in \widehat{\mathscr{I}}(\mathscr{D}) := \bigcap_{i=1}^{n} \mathscr{I}(\mathscr{D}(\mathscr{T}_i))$ is therefore described by n values for support. Imposed by the order of time the values form a sequence called *support history* $H(\mathscr{X}) := (\text{supp}(\mathscr{X}, T_1), \ldots, \text{supp}(\mathscr{X}, T_n))$ of the itemset $\mathscr{X}$. These histories are then used in subsequent steps to detect interesting change patterns.

To continue our example, suppose that we may discover that the support of the itemset

$$\text{Age} > 50, \quad \text{Satisfied} = \text{Yes}$$

has a downward trend. This, in turn, can be interpreted as that the group of all satisfied customers over 50 steadily gets smaller.

4. System architecture

As already mentioned above, our approach builds upon the idea of deriving frequent itemsets as representations of customer segments at different points in time, which are then analysed for changes. To derive a history, data sets collected during many consecutive periods have to be analysed for frequent itemsets. After each analysis session the discovered itemsets have to be compared to those discovered in previous periods and their histories have to be extended. On the other hand, history values may be discarded if their age exceeds an application dependent threshold. Therefore, itemsets and histories have to be stored on a long term basis. Taking all of the aforesaid into account the first task of our system is:

1. Frequent itemsets have to be *discovered* and their histories efficiently stored, managed and maintained.
   If histories with a sufficient length are available, the next task is straightforward:
2. Histories that exhibit specific change patterns have to be reliably *detected*.
   Frequent itemset discovery is generally connected with two problems. In the first place, a vast number of itemsets will be detected. Secondly, frequent itemsets may be obvious, already known or not relevant.
   Since a history is derived for each itemset, the first problem also affects our system: it has to deal with a vast number of histories and thus it is likely that many

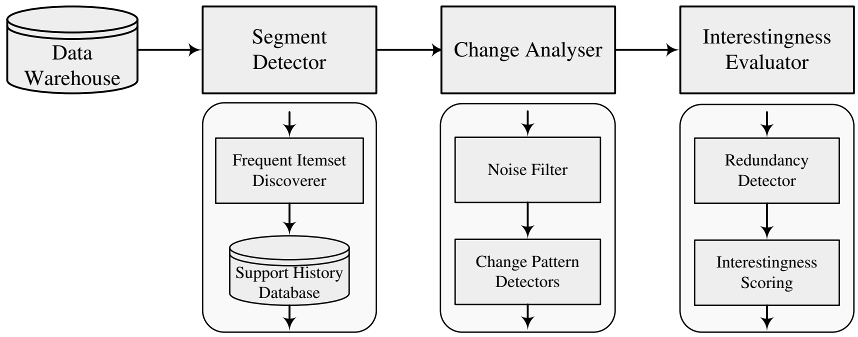

Fig. 1. Architecture of our system for customer segmentation.

change patterns will be detected. Moreover, as we will briefly discuss in Section 5, methods that were developed to deal with this problem for itemsets cannot be used when it comes to analysing change. Furthermore, there is also a quality problem: not all of the detected change patterns are equally interesting to a user and the most interesting are hidden among many irrelevant ones. Overall, the third task is:

3. Histories with a change pattern have to be analysed for redundancies and *evaluated* according to their interestingness.

Because the aforementioned tasks build upon each other, they can be seen as layers which make up a processing flow. According to their task the layers are termed *Segment Detector*, *Change Analyser* and *Interestingness Evaluator*, respectively (see Fig. 1).

5. Segment detector

Given a timestamped data set collected during a certain period, the task of the Segment Detector is to discover and store the customer segments in it. Since in our application each frequent itemset is a potentially interesting customer segment the first component of this layer is an algorithm for frequent itemset discovery, its second component is a database that stores and manages itemsets and their histories. Both components, but also the choice of the time periods, will be explained in the following.

In order to obtain the data set, the period length has to be chosen. Two aspects have to be considered. Long periods lead to many transactions in the individual data sets for the different periods and thus can enhance the reliability of the calculated support. Short periods allow to measure support more frequently, which may lead to a more reliable and earlier detection of change patterns. The choice of periods length should therefore depend on the available amount of data.

After the data set is available, frequent itemset discovery is applied to it. A typical approach may not only consist of the discovery method itself, but also of methods for pruning and constrained mining. Such methods have been developed to cope with the aforementioned problem of a vast amount of discovered itemsets in each period. This itemset quantity problem directly affects our application. A huge number of histories has to be processed and consequently far too many change patterns will be reported. In order to cope with this problem, pruning methods have been developed in order to constrain the itemsets. From the perspective of change discovery such pruning methods treat itemsets generated in different time periods independently from another. However, in our application we process many, temporally ordered itemsets. Thus the itemset property utilized for pruning – in general a measure based on itemset statistics – may vary for some itemsets over time, but still match the pruning criterion in each itemset. Although these variations may render itemsets interesting, they are discarded by existing approaches for itemset pruning. Consequently, we cannot directly use them.

In order to allow fast access and long-term storage, rules and their histories are managed within a database. The most frequent operations on such a database are:

- Determining whether a given rule is contained in the database, or not.
- Retrieving the histories for a given rule.

In our system we use a *Oracle* database management system and store the discovered rules in a database schema which is similar to the *rule cache* proposed in Hipp, Mangold, Güntzer, and Nakhaeizadeh (2002). It differs in two aspects: firstly, we only use the tables storing itemsets and we do not use a table storing rules. Secondly, we introduced a table which stores histories and which is linked to the itemset table using a foreign key association.

6. Change analyzer

The task of the *Change Analyzer* is to discover change patterns in support histories. Here, however, we only discuss how histories are detected that are stable or exhibit a trend. The Change Analyzer fulfills its task by a two step approach. In the first step, a filter is applied to the histories to reduce the noise contained in them. In a second step statistical tests for trend and stability are conducted.

Support histories inherently may contain random noise. Random noise may influence subsequent analysis steps in such a way that wrong and misleading results are produced. To reduce this effect, we use *double exponential smoothing* (Chatfield, 2001) in order to reveal more clearly any trend or stability. It is a simple and fast, yet effective method, which can easily be automated.

A trend is present if a history exhibits steady upward growth or a downward decline over its whole length. This definition is rather loose, but in fact there exists no fully satisfactory definition for trend (Chatfield, 2001). From a data mining perspective a trend describes the pattern that each value is likely to be larger or smaller than all its predecessors within a sequence, depending on whether the trend is upward or downward. Hence, it is a qualitative statement about the current and likely future development of a history. However, taking aspects of interpretability and usefulness into account, such a statement is sufficient in the case of our application. When faced with a vast number of customer segments and their histories, a user often has a basic expectation whether they should exhibit a trend and of what kind. By comparing his expectations with reality he will mostly be able to roughly assess the implications for his business. On the other hand, a user will rarely know in advance how trends should look like quantitatively, for example, their shape or target values. Thus he may be unable to exploit the advantages of more sophisticated trend descriptions, like regression models.

To choose a method for trend detection, it has to be taken into account that the number of histories to examine is huge. Whenever a trend is reported the user is basically forced to rely on the correctness of this statement, because it is infeasible for him to verify each trend manually. In addition to the requirement of reliable detection, the method should incorporate no assumptions about any underlying model, because it is very unlikely that it will hold for all or at least most sequences. Therefore, nonparametric statistical tests are the appropriate choice for trend detection.

Within our system we provide two statistical tests for trend, the *Mann–Kendall test* (Mann, 1945) and the *Cox–Stuart test* (Cox & Stuart, 1955). The Cox–Stuart test exploits fewer features of the history, leading to a computational effort that increases linearly with the history length. Although this may render the Cox–Stuart test susceptible to noise, because the influence of artefacts on the test result is stronger, it is considerably faster for long histories. In contrast to this, the Mann–Kendall test is much more robust, but its computational effort increases quadratically with the history length. Therefore, it has to be determined which of the two issues – speed or robustness – is more important depending on the actual characteristics of the data used.

Roughly speaking, a history is considered stable if its mean level and variance are constant over time and the variance is reasonably small. Similar to trends, a clear definition of stability is difficult. For example, a history may exhibits a cyclical variation, but may nevertheless be stable on a long term scale. Depending on the actual interest of a user, either the one or the other may have to be emphasised. From a data mining perspective stability describes the pattern that each value is likely to be close to a constant value, estimated by the mean of its predecessors. Thus it is, like a trend, a qualitative statement about the future development of a history. However, in contrast to a trend, it can easily be modeled in an interpretable and useful way, e.g. by the sample mean and variance. Generally, stable customer segments are more reliable and can be trusted – an eminently useful and desirable property for long term business planning.

To test for stability we use a method based on the well-known χ^2 test. However, since the χ^2 test does not take the inherent order of a history's values into account, our method may infrequently also classify histories as stable, which actually exhibit a trend. Therefore, we chose to perform the stability test as the last one in our sequence of tests for change patterns.

7. Interestingness evaluator

Since usually a vast number of change patterns for customer segments will be detected, it is essential to provide methods which reduce their number and identify potentially interesting ones. This is the task of the *Interestingness Evaluator*. To reduce the number of candidate segments the Interestingness Evaluator contains a redundancy detection approach, based on so-called derivative histories (Böttcher et al., 2005). Although this approach proves to be very effective, the number of temporally non-redundant customer segments may still be too large for manual examination. Therefore, a component for interestingness evaluation is provided, which contains a set of interestingness measures.

7.1. Redundancy detection

Generally, most changes captured in a segment's history – and consequently also change patterns – are simply the snowball effect of the changes of other segments. Suppose we are looking at customer satisfaction surveys and our system would discover that the support of the segment

$$\mathscr{X}_1 : \text{Age} > 50, \quad \text{Satisfied} = \text{Yes}$$

shows an upward trend. That is, the fraction of customers over 50 who are satisfied increases. However, if the fraction of males among all over 50 year old satisfied customers is stable over time, the history of

$$\mathscr{X}_2 : \text{Age} > 50, \quad \text{Gender} = \text{Male}, \quad \text{Satisfied} = \text{Yes}$$

shows qualitatively the same trend. In fact, the history of segment $\mathscr{X}_2$ can be *derived* from the one of $\mathscr{X}_1$ by multiplying it with a gender related constant factor. For this reason, the segment $\mathscr{X}_2$ is *temporally redundant* with respect to its support history.

It is reasonable to assume that a user will generally be interested in customer segments with non-derivative and thus non-redundant histories, because they are likely key drivers for changes. Moreover, derivative segments may lead to wrong business decisions. In the above example, a decision based on the change in segment $\mathcal{X}_2$ would account for the gender as one significant factor for the observed trend. In fact, the gender is completely irrelevant. Therefore, the aim is to find segments that are non-redundant in the sense that their history is not a derivative of related segments' histories. In a way, the approach is searching and discarding segments that are not the root cause of a change pattern which, in turn, can be seen as a form of pruning. In order to find derivative segments we have to answer the following questions. Firstly, what is meant by *related* itemsets (segments, respectively), and secondly, what makes a history a *derivative* of other histories. Regarding the first question, we use the superset relation to define *related itemsets*: an itemset $\mathcal{Y}$ is related to an itemset $\mathcal{X}$ iff $\mathcal{X} \prec \mathcal{Y} := \mathcal{X} \supset \mathcal{Y}$. We also say that $\mathcal{Y}$ is *more general* than $\mathcal{X}$ because its supporting transaction set is larger. In the following, we write $\mathcal{X}\mathcal{Y}$ for $\mathcal{X} \cup \mathcal{Y}$. We then define:

Definition 1. Let $\mathcal{X}, \mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_p$ be itemsets with $\mathcal{X} \prec \mathcal{X}_i$ for all i and $p > 0$. Let the $\mathcal{X}_i$ be pairwise disjoint. Let supp the support, $\mathrm{supp}(T) := \mathrm{supp}(\mathcal{X}, T)$ and $\mathrm{supp}_i(T) := \mathrm{supp}(\mathcal{X}_i, T)$ its functions over time and $\mathcal{M} := \{g : \mathbb{R} \to \mathbb{R}\}$ be the set of real-valued functions over time. The history $H(\mathcal{X})$ is called *derivative* iff a function $f : \mathcal{M}^p \to \mathcal{M}$ exists such that for all $T \in \widehat{T}$

$$\mathrm{supp}(T) = f(\mathrm{supp}_1, \mathrm{supp}_2, \ldots, \mathrm{supp}_p)(T) \tag{1}$$

For simplicity, we call an itemset *derivative* iff its history is derivative.

The main idea behind the above definition is that the history of an itemset is derivative, if it can be constructed as a mapping of the histories of more general itemsets. To compute the value $\mathrm{supp}(\mathcal{X}, T)$ the values $\mathrm{supp}(\mathcal{X}_i, T)$ are thereby considered. The definition above does not allow for a pointwise definition of f on just the $T \in \widehat{T}$, but instead states a general relationship between the support values independent from the point in time. It can therefore be used to predict the value of $\mathrm{supp}(\mathcal{X})$ given future values of the $\mathrm{supp}(\mathcal{X}_i)$. A simple example we will see below is $\mathrm{supp} = f(\mathrm{supp}_1) = c\,\mathrm{supp}_1$, i.e. the support history can be obtained by multiplying the support history of a more general itemset with a constant c.

In the following, we introduce two criteria for detecting derivative support histories which can be used in combination or independently from another. The functions f are quite simple and we make sure that they are intuitive.

The first criterion checks if the support of an itemset can be explained with the support of exactly one less specific itemset.

Criterion 1. The term $\mathrm{supp}(\mathcal{X}\mathcal{Y}, T)/\mathrm{supp}(\mathcal{Y}, T)$ is constant over $T \in \widehat{T}$ given disjoint itemsets $\mathcal{X}$ and $\mathcal{Y}$.

The meaning of the criterion becomes clear when being rewritten as

$$c = \mathrm{supp}(\mathcal{X}\mathcal{Y}, T)/\mathrm{supp}(\mathcal{Y}, T) = P(\mathcal{X}\mathcal{Y}|T)/P(\mathcal{Y}|T) = P(\mathcal{X}|\mathcal{Y}T)$$

with a constant c. The probability of $\mathcal{X}$ is required to be constant over time given $\mathcal{Y}$, so the fraction of transactions containing $\mathcal{X}$ additionally to $\mathcal{Y}$ constantly grows in the same proportion as $\mathcal{Y}$. For this reason, the influence of $\mathcal{X}$ in the itemset $\mathcal{X}\mathcal{Y}$ on the support history is not important. Due to

$$\mathrm{supp}(\mathcal{X}\mathcal{Y}, T) = c \cdot \mathrm{supp}(\mathcal{Y}, T) \tag{2}$$

with $c = \mathrm{supp}(\mathcal{X}\mathcal{Y}, T)/\mathrm{supp}(\mathcal{Y}, T)$ for any $T \in \widehat{T}$, $\mathcal{X}\mathcal{Y}$ is obviously a derivative of $\mathcal{Y}$ with respect to support history as defined in Definition 1.

Figs. 2 and 3 show an example of a derivative support history which we obtained from the survey data set used in Section 8. Fig. 2 shows the support histories of the less specific itemset at the top and the more specific itemset underneath over 20 time periods. The shape of the two curves is obviously very similar and it turns out that the history of the more specific itemset can be approximately reconstructed using the less specific one based on (2). As shown in Fig. 3, the reconstruction is not exact due to noise.

Opposed to the criterion above, the following is based on the idea of explaining the support of an itemset with the support values of two subsets.

Criterion 2. The term $\frac{\mathrm{supp}(\mathcal{X}\mathcal{Y},T)}{\mathrm{supp}(\mathcal{X},T)\mathrm{supp}(\mathcal{Y},T)}$ is constant over $T \in \widehat{T}$ given disjoint itemsets $\mathcal{X}$ and $\mathcal{Y}$.

$\mathrm{supp}(\mathcal{X}\mathcal{Y}, T)$ measures the probability of the itemset $\mathcal{X}\mathcal{Y}$ in period T which is $P(\mathcal{X}\mathcal{Y}|T)$. The term $\frac{\mathrm{supp}(\mathcal{X}\mathcal{Y},T)}{\mathrm{supp}(\mathcal{X},T),\mathrm{supp}(\mathcal{Y},T)} = \frac{P(\mathcal{X}\mathcal{Y}|T)}{P(\mathcal{X}|T)P(\mathcal{Y}|T)}$ is quite extensively used in data mining to measure the degree of dependence of $\mathcal{X}$ and $\mathcal{Y}$ at time T. Particularly in association rule mining this measure is also known as *lift* (Webb, 2000), or *interest factor* (Silverstein, Brin, & Motwani, 1998). The criterion therefore expresses that the degree of dependence between both itemsets is constant over time.

The support history of $\mathcal{X}\mathcal{Y}$ can then be constructed using

$$\mathrm{supp}(\mathcal{X}\mathcal{Y}, T) = c \cdot \mathrm{supp}(\mathcal{X}, T)\mathrm{supp}(\mathcal{Y}, T) \tag{3}$$

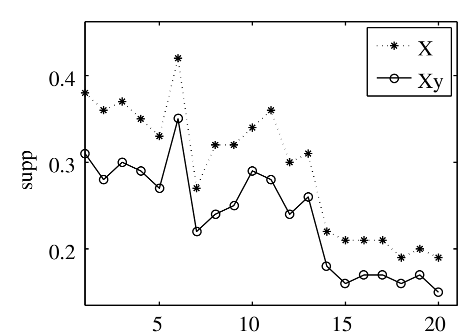

Fig. 2. Histories of the segment $\mathcal{X}$ and its derivative segment $\mathcal{X}y$.

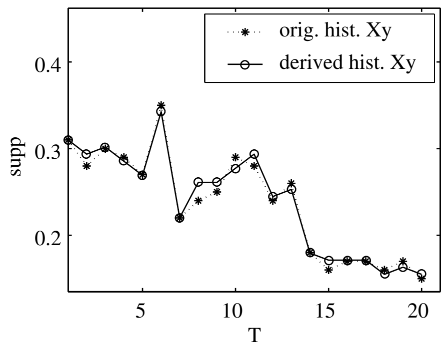

Fig. 3. Reconstructed history of $\mathcal{X}y$ using the history of $\mathcal{X}$.

with $c = \text{supp}(\mathcal{X}\mathcal{Y}, T)/(\text{supp}(\mathcal{X}, T)\text{supp}(\mathcal{Y}, T))$ for any $T \in \widehat{T}$, that is the individual support values of the less specific itemsets are used corrected with the constant degree of dependence on another. According to Definition 1 the support history of $\mathcal{X}\mathcal{Y}$ is therefore derivative.

Overall, an itemset is considered derivative if more general itemsets can be found, such that at least one of Criterion 1 or 2 holds.

7.1.1. Testing the criteria

To check if the history of an itemset $\mathcal{X}\mathcal{Y}$ is derivative, we need to test if the criteria of the previous section hold. Due to data usually being noisy, we will not check Criterions 1 and 2 directly, but instead statistically test their validity. Also, we rewrite the criteria in an equivalent form in order to account for the order of values over time in the histories. Our experiments have shown that direct use of the criteria counterintuitively marked some histories as derivative when they were noisy.

Let $\Delta_i\text{supp}(\mathcal{X}) := \frac{\text{supp}(\mathcal{X}, T_i)}{\text{supp}(\mathcal{X}, T_{i-1})}$ be the relative change in support for itemset $\mathcal{X}$ between two periods T_{i-1} and T_i. Then Criterion 1 holds, iff for any $T_i \in \widehat{T} \setminus \{T_1\}$

$$\Delta_i\text{supp}(\mathcal{X}\mathcal{Y}) = \Delta_i\text{supp}(\mathcal{Y}) \tag{4}$$

Likewise, Criterion 2 holds, iff for any $T_i \in \widehat{T} \setminus \{T_1\}$

$$\Delta_i\text{supp}(\mathcal{X}\mathcal{Y}) = \Delta_i\text{supp}(\mathcal{X})\Delta_i\text{supp}(\mathcal{Y}) \tag{5}$$

This means that if Criterion 1 holds for an itemset $\mathcal{X}\mathcal{Y}$ then the relative changes in its history are equal to the temporally related relative changes in the history of a more general itemset $\mathcal{X}$. If Criterion 2 holds, then the relative changes in the history of $\mathcal{X}\mathcal{Y}$ are equal to the product of the corresponding relative changes in the histories of $\mathcal{X}$ and $\mathcal{Y}$.

Obviously, (4) and (5) are following the same general scheme $y_i = x_i$; $i = 2, \ldots, n$, whereas the quantities y_i and x_i stand for the left and accordingly right-hand side of the equations.

It is convenient for the following discussion to imagine x_i and y_i in a plot, whereby y_i is – as implied by Definition 1 – the dependent quantity. If $y_i = x_i$ holds, then all points in the plot should be on a straight line with slope 1 and intercept 0. In practice this equality will rarely hold due to noise. In fact, the underlying relationship will be $y_i = x_i + \epsilon$ where ϵ is a random error with zero mean and unknown, but low variance.

Under the assumption that the dependency of y_i from x_i can be generally described by $y_i = a\ x_i + b + \epsilon$, we fit a regression line $y = \hat{a}x + \hat{b}$. We then test if x_i is statistically equal to y_i by carrying out the following to two steps:

(1) Based on the estimates $\hat{a}$ and $\hat{b}$ we test the hypothesis that the true parameters of the model are $a = 1$ and $b = 0$ using a standard t-test (Montgomery & Runger, 2002).
(2) Additionally, we test if the variance of ϵ is small, i.e. if the points (x_i, y_i) are sufficiently close to the regression line by setting a threshold $\tilde{r}$ for Pearson's correlation coefficient r.

Fig. 4 illustrates the testing procedure. It shows the scatter plot of the relative changes of the support histories from Fig. 2. The regression line is $y = 1.0107x - 0.0103$ and the correlation coefficient $r \approx 0.97$. The above test procedure using a significance level of 0.05 and $\tilde{r} = 0.95$ shows that the more specific itemset is indeed derivative with respect to the history of the less specific one.

7.1.2. Implementation issues

The proposed criteria rely on a search over the set of related (more general) itemsets. Generally, this search is exhaustive and thus a potentially exponential number of comparisons is required (e.g. for every frequent itemset all subsets have to be enumerated in the worst case). The approach's apparent complexity may evoke questions about its feasibility. Our experiments have been conducted on real business data and computation time was reasonable, in particular considering that change mining for business data are typically carried out on a weekly or monthly basis. Nonetheless, in some domains complexity may be an issue. In this case, the number of comparisons can be considerably reduced by the following simplification adopted from Liu et al. (2001): instead of all related itemsets only *closely related* ones are considered. For Criterion 1, this means that for an itemset $\mathcal{X}$ and any $y \in \mathcal{X}$ the itemset

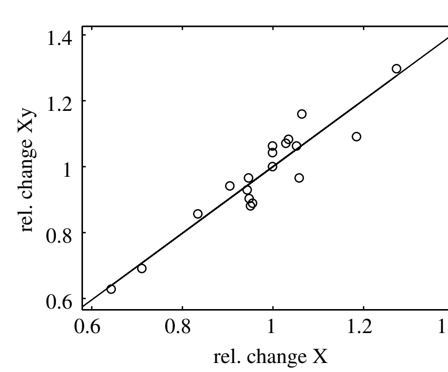

Fig. 4. Scatter plot of the relative changes of the curves shown in Fig. 2.

$\mathscr{X} \setminus \{y\}$ is considered, and for Criterion 2 that the itemsets $\mathscr{X} \setminus \{y\}$ and $\{y\}$ are considered.

7.2. Interestingness scoring

To assess the interestingness of detected trends and stabilities it has to be considered that each history is linked to a segment which itself has a certain relevance to a user. The detection of a specific change pattern may significantly influence this prior relevance. However, there is no broadly accepted and reliable way of measuring an itemset's interestingness up to now (Tan, Kumar, & Srivastava, 2004). Therefore, we consider any statement about the interestingness of a history also as a statement about the interestingness of its related itemset.

To assess stable histories two things should be considered: in the first place, most data mining methods typically assume that the domain under consideration is stable over time. Secondly, support is an interestingness measure for itemsets themselves. Taking all this into account, a stable history is in some way consistent with the above-mentioned assumption of data mining. It is summarised by the mean of its values, which in turn can then be treated as an objective interestingness measure. Here the variance of the history can be neglected, since it is constrained by the stability detection method.

To develop objective interestingness measures for trends is more complex due to their richness of features. For identifying salient features of a given trend, it is essential to provide reference points for comparison. As such we chose the assumptions a user naively makes in the absence of any knowledge about the changes in support histories. From a psychological perspective they can be seen as the anchors relative to which histories with a trend are assessed: a trend becomes more interesting with increasing inconsistency between its features and the user's naive assumptions. We identified three such assumptions and defined heuristic measures for the discrepancy between a history and an assumption:

- *Stability*: Unless other information is provided, a user assumes that histories are stable over time. This assumption does not mean that he expects no trends at all, but expresses his naive expectations in the absence of precise knowledge about a trend. It should be noted that this is consistent with many data mining approaches, which implicitly assumes that the patterns hidden in the data are stable over time. The histories of the segment $\mathscr{X}\mathscr{Y}$ in Fig. 5 would violate the stability assumption because its trend is very clear.
- *Non-rapid change*: Since a user shapes its business, he will be aware that the domain under consideration changes over time. However, he will assume that any change is continuous in its direction and moderate in its value. For example, if a business starts a new campaign, it will probably assume that the desired effect on the customers evolves moderately, because, for instance, not all people will see a commercial immediately. On the other hand, a rapid change in this context attracts more attention, because it may hint at an overwhelming success or an undesired side effect. For example, the history of the segment $\mathscr{Y}$ in Fig. 5 would be very interesting according to the non-rapid change assumption because the depicted trend is very pronounced and steep.
- *Homogeneous change*: If the support of an itemset changes over time, it is assumed that the rate and direction of changes in the support of all its supersets are the same. This basically means that the observed change in the itemset does not depend on further items. For example, a user may know that the fraction of satisfied customers increases. The homogeneous change assumptions states that the observed change in satisfaction affects all customers and not only selected subpopulations, e.g. customers over 50. For example, the fraction of satisfied males among all customers may increase. According to the homogeneous change assumption a user would conclude that the fraction of all satisfied married male customers increases at the same rate. For example, the history of the segments $\mathscr{X}\mathscr{Y}$ in Fig. 5 would be very interesting because its shape is completetly different from those of its more general segments.

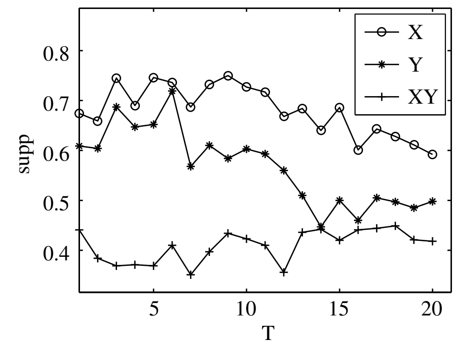

Fig. 5. Examples of interesting histories which exhibit a trend.

8. Experimental evaluation

To evaluate our system, we chose two representative real-life dataset. One contains answers of residential customers to a survey collected over a period of 40 weeks. The other contains network usage data of business customers collected over a period of 9 months. We transformed each dataset into a transaction set by recoding every (attribute, attribute value) combination as an item.

In the survey dataset, each tuple is described by 19 nominal attributes with a domain size between 2 and 10. We split the transaction set into 20 subsets, each corresponding to a period of two weeks. The subsets contain between 829 and 1049 transactions. From each subset we derived frequent itemsets (customer segments, respectively) with a support greater than 0.04 and not more than five describing attributes per segment. From the obtained 20 frequent

Table 1
Absolute number of segments which exhibit a trend or are stable differentiated by non-redundancy

	Segments		Trend down		Trend up		Stable	
	All	Non-redundant	All	Non-redundant	All	Non-redundant	All	Non-redundant
Surveys	1202	457	50	31	147	50	830	307
Network	8984	1909	3030	294	100	43	5854	1572

Table 2
Relative number of segments which exhibit a trend or are stable differentiated by non-redundancy

	Segments		Trend down		Trend up		Stable	
	All	Non-redundant	All	Non-redundant	All	Non-redundant	All	Non-redundant
Surveys (%)	100.0	38.0	4.2	2.6	12.2	4.2	69.1	25.5
Network (%)	100.0	21.2	33.7	3.3	1.1	0.5	65.2	7.5

itemsets we created a compound itemset by intersecting them. Its size is 1202.

The network usage dataset is described by 24 nominal attributes with a domain size of 5. We split the transaction set into nine subsets each covering a period of one month and having a size of 37 transactions. From each subset we derived frequent itemsets (customer segments, respectively) with a support greater than 0.1 and not more than five describing attributes per segment. The intersection of these itemset has a size of 8984.

Subsequently, we run our proposed system using the Mann–Kendall test for trend detection. Thereby two objectives are linked with our evaluation. Firstly, the number of segments exhibiting trends or stabilities has to be determined. Secondly, the number of derivative support histories has to be determined. The results of our analysis are shown in Tables 1 and 2.

As we can see the number of segments which exhibit a trend or stability strongly depends on the data set. For example, in the survey data set approximately 4.2% of the segments show an upward trend compared to 33.7% for the network usage data. It shows, however, that segments which exhibit some kind of regular change exist and that they can be rather frequent. Looking in the columns for non-redundant changes we can see that only a small fraction of changing segments cannot be explained by the change of more general segments. As we discussed earlier, segments with redundant changes can lead to suboptimal business decisions. As we see in our results they also significantly increase the number of changing segments. This, again, underlines the need for redundancy detection in our system for which we provided a powerful method.

9. Conclusions

We have shown how frequent itemset discovery, combined with tracking the temporal development of support and the application of an change-based interestingness notion can be used for detecting and monitoring customer segments. This is a very important challenge for customer-focussed enterprises facing very dynamic markets. Many businesses regularly collect huge volumes of time-stamped data about all kinds of customer interactions. This data reflects changes in customer behavior. It is crucial for the success of most businesses to detect these changes, correctly interpret their causes and finally to adapt or react to them. Hence, there is a significant need for data mining approaches that are capable of finding the most relevant and interesting changes in a data set.

We have proposed a system for our approach that can provide detailed knowledge about how customer behavior evolves over time. We successfully applied our system to two problem domains which are very significant for a telecommunications company: customer analytics, to investigate what is likely to drive customer satisfaction in the future, and network usage, to understand the drivers of change in customer behavior when they are using services.

References

Agrawal, R., & Psaila, G. (1995). Active data mining. In *Proceedings of the first international conference on knowledge discovery and data mining (KDD-95)* (pp. 3–8).

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD international conference on management of data* (pp. 207–216). Washington, DC.

Au, W.-H., & Chan, K. (2005). Mining changes in association rules: A fuzzy approach. *Fuzzy Sets and Systems, 149*(1), 87–104.

Böttcher, M., Spott, M., & Nauck, D. (2005). Detecting temporally redundant association rules. In *Proceedings of fourth international conference on machine learning and applications* (pp. 397–403). Los Angeles, CA: IEEE Computer Society.

Chatfield, C. (2001). *Time-series forecasting*. London/Boca Raton, FL: Chapman & Hall/CRC Press.

Chen, M.-C., Chiu, A.-L., & Chang, H.-H. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications, 28*(4), 773–781.

Cox, D., & Stuart, A. (1955). Some quick sign tests for trend in location and dispersion. *Biometrika, 42*, 80–95.

Hipp, J., Mangold, C., Güntzer, U., & Nakhaeizadeh, G. (2002). Efficient rule retrieval and postponed restrict operations for association rule mining. In *Proceedings of the 6th Pacific-Asia conference on advances in knowledge discovery and data mining* (pp. 52–65). Berlin: Springer.

Kim, J. K., Song, H. S., Kim, T. S., & Kim, H. K. (2005). Detecting the change of customer behavior based on decision tree analysis. *Expert Systems, 22*(4), 193–205.

Liu, B., Hsu, W., & Ma, Y. (2001). Discovering the set of fundamental rule changes. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 335–340).

Liu, B., Ma, Y., & Lee, R. (2001). Analyzing the interestingness of association rules from the temporal dimension. In *Proceedings of the 2001 IEEE international conference on data mining* (pp. 377–384).

Liu, B., Hsu, W., Han, H.-S., & Xia, Y. (2000). Mining changes for real-life applications. In *Proceedings of the 2nd international conference on data warehousing and knowledge discovery* (pp. 337–346). London: Springer.

Mann, H. (1945). Nonparametric tests against trend. *Econometrica, 13*, 245–259.

Montgomery, D., & Runger, G. (2002). *Applied statistics and probability for engineers*. New York: Wiley.

Shin, H., & Sohn, S. (2004). Segmentation of stock trading customers according to potential value. *Expert Systems with Applications, 27*(1), 27–33.

Silverstein, C., Brin, S., & Motwani, R. (1998). Beyond market baskets: Generalizing association rules to dependence rules. *Data Mining and Knowledge Discovery, 2*(1), 39–68.

Song, H. S., Kim, J. K., & Kim, S. (2001). Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications, 21*(3), 157–168.

Spiliopoulou, M., Baron, S., & Günther, O. (2003). Efficient monitoring of patterns in data mining environments. In *Advances in databases and information systems, 7th East European conference (ADBIS 2003)* (pp. 253–265). Berlin: Springer.

Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems, 29*(4), 293–313.

Webb, G. I. (2000). Efficient search for association rules. In *Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 99–107).