Applied Soft Computing 24 (2014) 316–327

Contents lists available at ScienceDirect

Applied Soft Computing

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

A new knowledge-based constrained clustering approach: Theory and application in direct marketing

Alex Seret [a,*], Thomas Verbraken [a], Bart Baesens [a,b,c]

[a] Department of Decision Sciences and Information Management, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium
[b] School of Management, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom
[c] Vlerick, Leuven-Gent Management School, Reep 1, B-9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 13 November 2012
Received in revised form 3 June 2014
Accepted 3 June 2014
Available online 19 July 2014

Keywords:
Data mining
Constrained clustering
Customer profiling
Business knowledge
Direct marketing

ABSTRACT

Clustering has always been an exploratory but critical step in the knowledge discovery process. Often unsupervised, the clustering task received a huge interest when reinforced by different kinds of inputs provided by the user. This paper presents an approach giving the possibility to incorporate business knowledge in order to guide the clustering algorithm. A formalization of the fact that an intuitive a priori prioritization of the variables might exist, is presented in this paper and applied in a direct marketing context using recent data. By providing the analyst with a new approach offering different clustering perspectives, this paper proposes a straightforward way to apply constrained clustering with soft attribute-level constraints based on feature order preferences.

Introduction

Data mining techniques and tools have been responsible for many of artificial intelligence's recent successes (e.g. [1–3]). Amongst these techniques, clustering has always been an exploratory but critical task in the knowledge discovery process and has been applied in nearly all domains in which the grouping of similar objects makes sense (e.g. [4–6]). Ranging from the most simple techniques, such as the *k*-means algorithm (e.g. [7,8]), to the most advanced approaches, such as kernel methods [9] and spectral approaches [10], clustering techniques have received interest from both the scientific and the business community. The users of such techniques are sometimes in possession of background knowledge and would like to include it into the clustering exercise. The usage of constraints in order to integrate this knowledge into the clustering task, i.e. constrained clustering, is a current active research topic leading to different approaches and techniques (see e.g. [11–13]). Amongst the different constraints' levels, instance-level constraints, based on pairwise information, such as the famous *must-link* and *cannot-link* [11], are widely discussed in the literature [14–17].

Although this kind of constraints is quite known, consider the myriad of papers dealing with semi-supervised clustering, other interesting levels have emerged in the literature and are of great interest when dealing with knowledge integration. In [18], the authors propose a variant of the *k*-means algorithm with cluster-level constraints, ensuring that no empty clusters or clusters with very few data points are obtained. In [19], an algorithm constrained by feature order preferences is presented, in which attribute-level constraints are used as part of the to-be-optimized objective function. Attribute-level constraints are also present in [13] where *must-link* and *cannot-link* are adapted by creating constraints on the attributes' values. Moreover hybrid approaches integrating different levels have been proposed, see e.g. the approaches of [20] in which both instance- and attribute-level constraints are used in order to guide the clustering task.

Another important dimension concerns the degree to which the constraints have to be satisfied, leading to the concepts of hard and soft constraints. On the one hand, hard constraints are constraints that are required to be fully satisfied. For example, the COP-KMEANS algorithm, presented in [11], requires the assignment of each point to a cluster such that instance-level constraints are not violated. If no such cluster exists, the algorithm fails. The same reasoning is used in [13] in which the *mlx-k*-Medoids algorithm is presented which fails if no cluster not violating a set of

* Corresponding author. Tel.: +32 16 326881.
E-mail addresses: alex.seret@kuleuven.be (A. Seret), thomas.verbraken@kuleuven.be (T. Verbraken), bart.baesens@kuleuven.be (B. Baesens).

http://dx.doi.org/10.1016/j.asoc.2014.06.002
1568-4946/

attribute-level constraints can be found. Bradley et al. [18] proposed a cluster-level constrained version of the *k*-means algorithm, which has to satisfy *k* hard constraints imposing that a cluster *k* has to have at least τ_k data points. On the other hand, soft constraints are used to guide the algorithm while accepting a partial violation (satisfaction) of the constraints. For example, Wagstaff [21] introduces the notion of soft-constraints in a soft version of COP-KMEANS, SKOP-MEANS, in which a strength factor α is used to indicate the reliability of a constraint. The objective function is penalized if constraints are violated proportionally to the value of their respective α, allowing a violation of the constraints while trying to minimize it. The same approach is used in [22] in which the authors apply soft-constraints in mixture clustering by using fuzzy constraints and a strength factor γ and optimize an objective function minimizing the constraints' violations. In [19,12], parameters λ and w are respectively used in order to fix the significance of the added constraints. Finally, in [23], the authors are using pairwise judgments of similarity and dissimilarity in a soft way by trying to find a partitioning of the vertices into clusters so that the number of violations is minimized.

In this paper, a new approach based on business considerations is proposed in order to incorporate business knowledge into the clustering task in an easy and efficient way. The goal of this approach is to focus on the perceived value of the partitioning resulting from the clustering task and not only on the statistical aspects of it (see e.g. [12]). This approach is based on the fact that business people, experts or not, have some insights regarding the importance of the variables before actually starting the analysis. If this insight is limited, an unconstrained approach has to be used, which has already been widely discussed in the literature. However, if this insight is considered sufficient, new approaches are needed in order to consider the a priori knowledge as a critical input for the clustering task. By incorporating this knowledge, the comprehensibility and the perceived value of the clustering will increase. From a more technical point of view, we propose a straightforward approach to transform background knowledge about features' importance into a metric that is used, as an example, to constrain the Self-Organizing Map algorithm in a soft way, leading to a soft-constrained attribute-level clustering approach based on metric learning. In a related work, Sun et al. [19] propose a solution to a problem which looks quite similar to the one this paper is tackling but which, in fact, is totally different. Sun et al. [19] propose a formulation of a clustering objective function penalizing the violation of feature order preferences making use of background knowledge about the importance of the features in order to create soft attribute-level constraints by parameterizing a weighted distortion measure. This objective function is further incorporated into a prototype-based clustering algorithm in which an iterative approach is used to converge to an accurate partitioning of the data points. Although the approach of Sun et al. [19] and the one proposed in this paper make use of feature order preference, the purposes of both methods are different. In [19], the approach and the algorithm lead to a better capturing of the ground truth if some background knowledge about the importance of the variables is available. This notion of importance is objective and is purely data-driven. Therefore, during their experiments, the "true" number of clusters is provided and simulated feature order preferences are generated using the ground truth class information. This idea of ground truth information is also exploited in [24], where a semi-supervised clustering method for incorporating instance-level and attribute-level information is proposed. In their work, attribute-level constraints are in the form of order preferences, generated using ground truth class information which enables the calculation of rough estimates of the optimal attribute weights leading to the order preferences. For their experiments, 6 UCI data sets with known class information are used to evaluate the proposed methods. In contrast, the proposed paper introduces a subjective, goal-driven notion of importance. Indeed, the soft attribute-level constraints of this paper are based on feature order preferences that reflect the importance of the variables as perceived by the analyst, hence introducing a bias guiding the algorithm and providing the analyst with a powerful exploratory knowledge-based tool.

The remainder of this paper is structured as follows. In section "Techniques used", the necessary background for this paper is introduced. Section "Prioritization approach" presents the approach for the prioritization of the variables in the clustering task. In section "Methodology implementing the prioritization approach", a 5-step methodology implementing the proposed prioritization approach is described. In section "Application", the theory is illustrated in a direct marketing context by a comparison between the results of the proposed approach and the traditional approach. The main conclusions are summarized in section "Conclusion".

Techniques used

In this section, the SOM algorithm, the *k*-means algorithm and the concept of salient dimension extraction are presented.

Self-organizing maps

Kohonen maps, also called self-organizing maps (SOM), have been introduced in 1981 by Kohonen. Fields like data exploratory analysis, web usage mining [25], industrial and medical diagnostics [26], and corruption analysis [8] are contemporary examples of SOM analysis applications and successes. This section is based on [27] and aims at giving a theoretical background to the reader, whereas an application of the technique can be found in section "Application". The main objective of the SOM algorithm is the representation of a high dimensional input dataset on lower dimensional maps. This enables to explore the data and to use techniques like visual correlation analysis or clustering analysis in an intuitive manner. In the first step, a feedforward Neural Network (NN) is trained on the input data. The output layer is a map with a lower dimensionality and a given number of neurons. During each iteration of the algorithm, an input data vector n_i is compared with the neurons m_r of the output map using Euclidian distances. The neuron m_c with the smallest distance with regard to the input vector is identified as the Best Matching Unit (BMU):

$$\| n_i - m_c \| = \min_r \{\| n_i - m_r \|\}. \quad (1)$$

The weights of the BMU are then modified in the direction of the input vector, leading to a self-organizing structure of the neurons. A learning rate $\alpha(t)$ and a neighborhood function $h_{cr}(t)$ are defined as parameters of the learning function:

$$m_r(t+1) = m_r(t) + \alpha(t) h_{cr}(t)[n_i(t) - m_r(t)]. \quad (2)$$

The learning-rate will influence the magnitude of the BMU's adaptation after matching with an input vector n_i, whereas the neighborhood function defines the range of influence of the adaptation. In order to guarantee the stability of the final output map, decreasing learning rates and neighborhood functions are often used at the end of the training. An exhaustive discussion of the influence of the parameters such as the number of neurons, the shape of the map, or the initial weights of the neurons is to be found in [27].

k-Means

The *k*-means algorithm is a typical iterative distance-based clustering approach which iteratively creates *k* clusters based on the