Knowledge-Based Systems 000 (2018) 1–15

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A new weighting k-means type clustering framework with an l^2-norm regularization

Xiaohui Huang[a,b,*], Xiaofei Yang[b], Junhui Zhao[a], Liyan Xiong[a], Yunming Ye[b]

[a] School of Information Engineering Department, East China Jiaotong University, Nanchang 330013, China
[b] Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China

ARTICLE INFO

Article history:
Received 22 September 2017
Revised 17 March 2018
Accepted 21 March 2018
Available online xxx

Keywords:
Clustering
k-means algorithm
Feature weighting
l^2-norm regularization

ABSTRACT

k-Means algorithm has been proven an effective technique for clustering a large-scale data set. However, traditional k-means type clustering algorithms cannot effectively distinguish the discriminative capabilities of features in the clustering process. In this paper, we present a new k-means type clustering framework by extending W-k-means with an l^2-norm regularization to the weights of features. Based on the framework, we propose the l^2-Wkmeans algorithm by using conventional means as the centroids for clustering numerical data sets and present the l^2-NOF and l^2-NDM algorithms by using two different smooth modes representatives for clustering categorical data sets. At first, a new objective function is developed for the clustering framework. Then, the corresponding updating rules of the centroids, the membership matrix, and the weights of the features, are derived theoretically for the new algorithms. We conduct extensive experimental verifications to evaluate the performances of our proposed algorithms on numerical data sets and categorical data sets. Experimental studies demonstrate that our proposed algorithms delivers consistently promising results in comparison to the other comparative approaches, such basic k-means, W-k-means, MKM_NOF, MKM_NDM etc., with respects to four metrics: *Accuracy, RandIndex, Fscore*, and Normal Mutual Information (*NMI*).

1. Introduction

Clustering is an unsupervised learning technique which aims to partitioning a data set into several disjoint subsets such that the objects within a subset have high similarities and the objects in different subsets are dissimilar by certain pre-defined criteria [1]. It is one of most widely used approaches for exploratory data analysis, with applications ranging from protein sequence analysis [2], community discover [3] to image segmentation [4] or astronomic data analysis [5].

k-Means clustering is a kind of partitioning clustering method. Since its simplicity and effectiveness, k-means clustering algorithm has been widely used to solve various real-life problems. All the features have equivalent effects in the basic k-means [6] clustering process. However, as a matter of fact, different features may have different discriminative capabilities for clustering a high-dimensional data set in real-life application. For instance, in the sentence "London is the first city to have hosted the modern Games of three Olympiads", the keywords "London, Olympiads" have more discriminative capabilities than the keywords "city, modern" in sport news. To distinguish the importance of different features in the clustering process, many weighting feature methods [1,7–11] were proposed by using the framework of k-means type clustering. These approaches can be classified two categories: (1) using the βth power to constrain feature weights [1,8,11]; (2) using the entropy to constrain feature weights [7,9,10,12,13]. Huang et al. proposed the method of using the βth power to constrain feature weights in the W-k-means algorithm [1,8] which is able to automatically weight features based on the importance of the features in the clustering process by adding a new step of calculating the weights to the basic k-means algorithm. However, when a cluster includes one feature on which the scatter is zero, only the weight of this feature is assigned to one and the weights of the other features are allocated to zeros. That means that only the feature of zero scatter is employed in the W-k-means clustering process. It is unreasonable for clustering a high-dimensional data set. Jing et al. proposed the EWkmeans algorithm [7] which uses the entropy to constrain feature weights under the k-means clustering framework by adding an entropy term in its objective function to stimulate more features to contribute to the identification of clusters. This method must calculate a groups of exponents with the negative scatters of the clusters on every feature as the exponential variables, i.e. $e^{-scatter}$. The scatter is usually large for a large

* Corresponding author.
E-mail address: hxh016@hotmail.com (X. Huang).

https://doi.org/10.1016/j.knosys.2018.03.028
0950-7051/

scale data set. Thus, $e^{-scatter}$ will tend to zero and often overflow in the implementation of the algorithm.

In this paper, we proposed a k-means type clustering framework by using a new fashion to weight features with an l^2-norm regularization. Based on the method of weighting features, a new clustering algorithm, named l^2-Wkmeans, is proposed by extending the W-k-means algorithm [1,8] for clustering numerical data sets and two new algorithms, named l^2-NOF and l^2-NDM, are proposed by extending MKM_NOF and MKM_NDM [14] for clustering categorical data sets, respectively. By combining the l^2-norm regularization and the nonnegative constraint to feature weights, our proposed methods can effectively select discriminative features and reduce the effects of noisy features in the clustering process. In addition, different from MKM_NOF and MKM_NDM have no capability of feature selection and W-k-means calculates a weight for every feature in a whole data set as the feature selection, our proposed methods calculate a weight for every feature in a cluster as the feature selection, which results in our proposed method are more robust than the original methods against noise. Then, we achieve the iterative updating rules of the three algorithms by minimizing the corresponding objective functions. Extensive experiments on both numerical and categorical data sets corroborate that our proposed methods can improve clustering results by more effective feature selection with an l^2-norm regularization. The main contributions of this work are twofold:

- We propose a k-means type clustering framework by using a new fashion to weight features with an l^2-norm regularization. On the basis of the framework, we propose three clustering algorithms: the l^2-Wkmeans algorithm for clustering numerical data sets, the l^2-NOF and l^2-NDM algorithms for clustering categorical data sets.
- We develop an objective function for the weighting k-means type clustering framework and give the complete proof of the convergence of our proposed algorithms by optimizing the corresponding objective functions.

The rest of the paper is organized as follows: a brief review related to k-means type clustering algorithms on a numerical data set and a categorical data set is presented in Section 2. Section 3 introduces the details of our proposed algorithms. Experiments on both numerical data sets and categorical data sets are presented in Section 4. Finally, we discuss the features of our proposed algorithms in Section 5 and conclude this paper in Section 6.

2. Related work

This section gives a briefly survey of the previous works which involve the k-means type clustering algorithms on numerical data sets and categorical data sets.

2.1. k-means type clustering on numerical data sets

The last decades have witnessed the rapid development of k-means type clustering algorithms which range from no weighting k-means type algorithms to various weighting k-means algorithms.

2.1.1. No weighting k-means type clustering algorithms

Basic k-means algorithm aims at finding a partition such that the sum of the squared distances between the empirical means of the clusters and the objects in the clusters is minimized. Basic k-means algorithm has been improved from the different aspects to overcome its weaknesses. Since k-means type algorithms are sensitive to initial centroids, Arthur and Vassilvitskii proposed k-means++ algorithm [15] which chooses the initial centroids by maximizing the distances among them. Another weakness of k-means type algorithms is to require manually tuning the parameter k (the number of clusters). Pelleg and Moore proposed X-means [16] to automatically seek the number of clusters by optimizing a criterion such as Akaike Information Criterion or Bayesian Information Criterion. Basic k-means algorithm and most of its variations usually get stuck at local optima. Therefore, Bagirov proposed a global k-means algorithm [17,18] which dynamically adds one cluster center at a time and uses each data point as a candidate for the kth cluster center. Since the global k-means mentioned above requires a large amount of computational cost, Bai et al. [19] proposed an acceleration mechanism for the production of new cluster centroids by applying local geometrical information to describe approximately the set of objects.

Moreover, in order to cater for the demands of clustering data sets in different applications, several variations [20–25] of k-means algorithms are proposed. Bradley et al. [20] presented a fast scalable and single pass version of k-means that is able to solve the clustering problem of data stream. Dhillon and Modha studied a certain spherical k-means algorithm [21] for clustering a document data set. To cluster a large-scale data set, the triangle inequality is employed to accelerate the standard k-means algorithm [22–24]. Similar to our proposed method, Regularized k-means [25] aims at improving clustering results on high-dimensional data sets by using the penalty term for eliminating the effects of redundant features. However, Regularized k-means [25] use l^1-norm penalty term to centroids instead of l^2-norm penalty term to feature weights in our proposed method.

This type of k-means algorithms equally treats all the features and has no capability to feature selection. Therefore, the clustering results produced by this type of algorithms are usually not promising when a data set includes noisy features.

2.1.2. Weighting k-means type clustering algorithms

All the features are employed equally in the clustering process of no weighting k-means type clustering algorithms. In actual practice, a useful clustering pattern usually hides in a subspace defined by a subset of all the features. Therefore, many researchers [1,7,8] are devoted to study various types of weighting feature ways to find the hidden subspaces. Huang et al. proposed the W-k-means [1] algorithm by using the βth power to constrain the feature weights to tune the distribution of feature weights. In W-k-means, the same feature in the different clusters shares a single value of feature weight. As a matter of fact, the same feature in the different clusters usually has different importance in real applications. For the sake of solving the problem mentioned above, Chan et al. proposed the Attributes-Weighting clustering Algorithm (AWA) [8] by using the similar weighting feature technique as W-k-means to calculate a weight for every feature in each cluster.

Let $X = \{X_1, X_2, \ldots, X_n\}$ be a set of n objects. Object $X_i = \{x_{i1}, x_{i2}, \ldots, x_{im}\}$ is characterized by a set of m features. The membership matrix U is a $n \times k$ binary matrix, where $u_{ip} = 1$ indicates that object i is assigned to cluster p, otherwise, it is not assigned to cluster p. $Z = \{Z_1, Z_2, \ldots, Z_k\}$ is a set of k vectors representing the centroids of k clusters. W is a weighting matrix where each row $W_p = \{w_{p1}, w_{p2}, \ldots, w_{pm}\}$ denotes a weight vector of all the features in a cluster. The attributes-weighting clustering algorithm can be formulated as

$$P(U, W, Z) = \sum_{p=1}^{k} \sum_{i=1}^{n} u_{ip} \sum_{j=1}^{m} w_{pj}^{\beta} (x_{ij} - z_{pj})^2, \tag{1}$$

subject to

$$u_{ip} \in \{0, 1\}, \sum_{p=1}^{k} u_{ip} = 1, \sum_{j=1}^{m} w_{pj} = 1, 0 \le w_{pj} \le 1, \tag{2}$$

where β is parameter to tune the distribution of feature weights. $\beta > 1$, and the larger of the $\beta's$ value, the more similar among the weights of different features in a cluster is. The AWA algorithm can

Please cite this article as: X. Huang et al., A new weighting k-means type clustering framework with an l^2-norm regularization, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.03.028