DATA CLUSTERING USING KERNEL BASED ALGORITHM

Mahdi'eh Motamedi¹ and Hasan Naderi²

¹MS Software Enineering, Science and Research Branch, Islamic Azad University, Lorestan, Iran
²Faculty of Software Engineering. Science and Technology, Tehran, Iran

Abstract

In recent machine learning community, there is a trend of constructing a linear logarithm version of nonlinear version through the ‘kernel method’ for example kernel principal component analysis, kernel fisher discriminant analysis, support Vector Machines (SVMs), and the current kernel clustering algorithms. Typically, in unsupervised methods of clustering algorithms utilizing kernel method, a nonlinear mapping is operated initially in order to map the data into a much higher space feature, and then clustering is executed. A hitch of these kernel clustering algorithms is that the clustering prototype resides in increased features specs of dimensions and therefore lack intuitive and clear descriptions without utilizing added approximation of projection from the specs to the data as executed in the literature presented. This paper aims to utilize the ‘kernel method’, a novel clustering algorithm, founded on the conventional fuzzy clustering algorithm (FCM) is anticipated and known as kernel fuzzy c-means algorithm (KFCM). This method embraces a novel kernel-induced metric in the space of data in order to interchange the novel Euclidean matric norm in cluster prototype and fuzzy clustering algorithm still reside in the space of data so that the results of clustering could be interpreted and reformulated in the spaces which are original. This property is used for clustering incomplete data. Execution on supposed data illustrate that KFCM has improved performance of clustering and stout as compare to other transformations of FCM for clustering incomplete data.

Keywords:
Kernel method, kernel clustering algorithms, clustering prototype, fuzzy clustering algorithm, kernel-induced metric.

1. INTRODUCTION

There are three different forms of grouping (group membership) is possible. In the non-overlapping groups, each object, only a group (segment clusters) is allocated to an object may be assigned to multiple groups at the overlapping groups, and wherein the fuzzy groups element belongs to each group with a certain degree of applying. Hard methods (e.g. k-means, spectral clustering, kernel PCA) arrange each data point to exactly one cluster, whereas in soft methods (e.g. EM algorithm with mixture-of-Gaussians model) to each data point for each cluster a level is assigned, with which this data point can be associated with this cluster. Soft methods are especially useful when the data points are relatively homogeneously distributed in space and the clusters are only as regions with increased data point density in appearance, that is, if there are, for example, transitions between the clusters or background noise (hard methods in this case useless) (Zhang & Lu, 2009).
The objective of clustering is to partition the point of data into standardized groups, emerges in many fields including machine learning, pattern recognition, image processing and determining. Means is one of the most popular clustering algorithm, where cluster errors are minimized in order to define the total of the distance of the squared Euclidean between every set of data point and the analogous cluster center (Jain, 2010). This algorithm has two key shortcomings including the dependency of solutions on the prior position of the cluster centers, rendering in low minima, and it is capable of calculating only clusters, which are separable linearly.

A popular and simple effort to overcome first shortcoming is the execution of multiple restarts, in which the centers of the clusters are placed randomly at distinct prior places, better local minima, therefore, could be calculated. One still has to take decision of the quantity of the restarts and one cannot certain about the initializations done are ample in order to attain a near-optimal minimum. The worldwide k-means algorithm has been projected in order to deal with the concern (Kulis, Basu, Dhillon & Mooney, 2009). This implies that it is a local search process. This Kernel is an expansion of the standard κ-means algorithm, which draws points of data from the spaces of input in order to feature space by means of a nonlinear minimizes, and transformation the errors in cluster in feature spaces. Clusters separated linearly, therefore, in input spaces are attained, surpassing the second limitation of k-means.

This paper proposes the global kernel means algorithm, optimization of deterministic algorithm for optimization of the cluster error, which utilizes kernel means as a process of local search for solving the problem of clustering. In an instrumental fashion, the algorithm functions and solves every intermediate problem with cluster utilizing kernel means. In order to reduce the complexity, two schemes of spreading up are recommended known as global kernel means and fast global kernel means with models of convex mixture. For each intermediate concern, the first variant placed the points of data set which assurances the highest minimization in errors of clustering when initializing the new cluster in order to add this point and perform kernel means just one time from the initialization. Nevertheless, the other variant is positioned in the set of data, by placing a model of convex mixture (Filippone, Camastra, Masulli & Rovetta, 2008). For every intermediate problem of mixture then tries just examples as an initialization possible for the cluster which are new rather than of the complete set of data.

The graphs presented emphasize on the nodes clustering. For optimization, spectral procedures are applied efficiently and effectively. A number of graphs reduce objectives such as normalized cut and association. The paper presents experimental outcomes, which contrast global kernel means, kernel means with multiple starts on artificial digits, data, graphs, and face images. The outcomes support the claim that global kernel means places near optimal solutions since it is executed followed by multiple restarts in the form of clustering error.

2. MATERIAL and METHOD

2.1 K-Means

The example of K-means is taken from (EM) Exception-Maximization algorithm. In the second and the third step respectively, the steps of maximization and exception are included. Owing to the fact that exception-maximization is algorithm is converted to a local maximum. It is noted that the K-means is, generally, a batch algorithm, such as every input is assessed before adaptation process, dissimilar to on-line algorithm, which is followed by code book modification. K-means lack robustness, which is a major drawback of the k-means. The concern can be resolved easily by monitoring the effect of outliers in the execution of means in data set. In the k-
means algorithm is the number K-set of clusters before the start. A function for calculating the
distance between two measurements must be added to the average value and compatible.

The algorithm is as follows:
1. Initialization: (random) selection of K-Cluster centers
2. Assignment: Each object is assigned closest cluster center
3. Recalculation: There are for each cluster, the cluster centers recalculated
4. Repeat: If now the assignment of objects changes, go to step 2, otherwise abort

2.2 Peculiarities k-Means Algorithm

The k-means algorithm delivers for different starting positions of the cluster centers may have
different results. It may be that a cluster remains empty in one step and thus (for lack of
predictability of a cluster center) can no longer be filled. It must be appropriately chosen k, and
the quality of the result depends strongly on the starting positions from. To find an optimal
clustering belongs to the complexity class NP . The k-means algorithm is not essentially the
optimal solution, but is very fast (Graves & Pedrycz, 2010).

To avoid these problems, we simply start the k-means algorithm new in the hope that in the next
run by other random cluster centers a different result is delivered. Because of the above
shortcomings of the theoretical k-means algorithm is considered heuristic because it still often
provides useful results.

3. EXTENSIONS and SPECIAL CASES

Extensions of the K-means algorithm are the k-median algorithm, the k-means algorithm ++ or
the Fuzzy c-means . The Iso-data algorithm can be viewed as a special case of K-means. This
method has a typical k-Means, k-Medoids (Tran, Wehrens, & Buydens, 2006). The essential
impression of k-Means algorithm is the n data objects into k classes (which all kinds of pre-k
unknown), so that each class after the division of the data points to a minimum distance of such
centers.

K-means algorithm is essentially to achieve a basic idea of clustering: class within the data points
as close as possible, the class between data points farther the better (Riesen & Bunke, 2010). In
most cases, the above algorithm will eventually get the optimal clustering results of convergence,
but there are also likely to fall into local optima. There is still a problem; the above algorithm has
a premise that the specified value k, i.e. the number of clusters. In practical applications, k value
is not given in advance, so another focus of k-means algorithm is to find a suitable k, square error
count reaches a minimum. The general practice is to try a number of k, the square error (distance)
of the minimum k value.

K-Means algorithm step is to calculate the mean of the current cluster center of each cluster as a
new cluster center, which is the k-Means algorithm named reasons (Ayvaz, Karahan, & Aral,
2007). Calculated as

$$\mu_k = \frac{1}{N_k} \sum_{x_i \in \text{cluster}_k} x_i$$
Standards are not common standards whether the convergence detection steps, because to achieve full convergence may not be realistic.

4. RESULT AND DISCUSSION

The set one of data is a 2D set developed by 424 points that is consisted of non-separable liner classes. The two modules, therefore, are not separated by K-means utilizing just two code vectors as shown in figure 1 (a).

Like SOM, K-means describes the limitations with respect to other clustering algorithms (Wu, Kumar, Quinlan, Ghosh, Yang, Motoda & Steinberg, 2008). The study then executed kernel method on the set of data utilizing two distance centers.
Figure 1b shows that two clusters can be separated by algorithm. It is crucial to note that the center counter images in the input space are not present.

Figure 2

![Figure 2](image)

According to figure 2, the behavior of algorithm thought out the stages needed for convergence. The data 2 is the most popular benchmark data employed in Machine Learning (Dong, Pei, Dong, & Pei, 2007). Linearly, one class is separable from the rest of the two, whereas, the rest of the two are not. Usually, IRIS data is epitomized by the projection of 4D data and the two key elements. The study contrasted the attainable results against the Ng-Jordan algorithm, which is a spectral clustering algorithm.

![Figure 3](image)

Figure 3 shows the results obtained utilizing K-Means and algorithm. In table 1, the second column illustrates the mean performance of K-Means on 20 runs of SOM, Ng-Jordan, and Neural Gas the methods
of the study attained with distinct parameter and initialization. The figures displayed show that this algorithm seemed to function better as compare to the other algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Iris Data</th>
<th>Wisconsin Database</th>
<th>Spam Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOM</td>
<td>121.5 ± 1.5 (81.0%)</td>
<td>600.5 ± 0.5 (96.7%)</td>
<td>1210 ± 30 (78.9%)</td>
</tr>
<tr>
<td>K-Means</td>
<td>133.5 ± 0.5 (89.0%)</td>
<td>656.5 ± 0.5 (96.1%)</td>
<td>1083 ± 153 (70.6%)</td>
</tr>
<tr>
<td>Neural Gas</td>
<td>137.5 ± 1.5 (91.7%)</td>
<td>606.5 ± 0.5 (96.1%)</td>
<td>1030 ± 120 (88.4%)</td>
</tr>
<tr>
<td>Na-Jordan Algorithm</td>
<td>126.5 ± 7.5 (84.3%)</td>
<td>652 ± 2 (95.5%)</td>
<td>929 ± 9 (90.6%)</td>
</tr>
<tr>
<td>Our kernel method</td>
<td>142 ± 1 (94.7%)</td>
<td>662.5 ± 0.5 (97.0%)</td>
<td>1247 ± 3 (81.3%)</td>
</tr>
</tbody>
</table>

5. Conclusion

This paper presents a practical approach to the evaluation of clustering algorithms and their performance on various high-dimensional and sparse data sets; whereas this sets high demands on the algorithms in terms of computational complexity and assumptions that must be made. The paper discussed approaches for solving this optimization and multi-stage problem. Distance matrices and recommender systems have been used to reduce the complexity of the problem and to calculate missing values. The study focused in the comparison of the different methods in terms of the similarity of the results, with the aim to find similar behavior. Another focus was on the flexibility of the algorithms with respect to the records, as well as the sparseness of the data and the dimensionalities have a major impact on the problem. In conclusion, it has been achieved with a combination of recommender systems, hierarchical methods, and Affinity Propagation good results. Kernel-based algorithms were sensitive with respect to changes in the output data set.

References