A New Approach On Incremental Affinity Propagation Clustering Technique Based On Preference

T. V. Satyavathi, 2 Dr. A. Krishna Mohan
1 (Dept of CSE, Jawaharlal Nehru Technological University, Kakinada (A.P), India)
2 (Ass. Prof In Dept of CSE, Jawaharlal Nehru Technological University, Kakinada (A.P), India)

Abstract:
Many of the clustering algorithms were intended for discovering patterns in static data. Nowadays, more and more data e.g., blogs, Web pages, video surveillance, etc., are come into view in dynamic manner, known as datastreams. Consequently, incremental clustering, evolutionary clustering, and data stream clustering are becoming scorching topics in data mining societies. Characteristics of the dynamic data, or data streams, include their high volume and potentially unbounded size, sequential access, and dynamically evolving nature. This impresses additional requirements to traditional clustering algorithms to hastily process and recap the immense amount of ad infinitum arriving data. It also necessitate the aptitude to adapt to changes in the data distribution, the ability to detect emerging clusters and differentiate them from outliers in the data, and the ability to merge old clusters or discard expired ones. All of these needs make lively data clustering an important confront.

Keywords: Affinity propagation, incremental clustering, K-Medoids, nearest neighbor assignment.

Introduction:
Five popular labelled data sets, actual world time series and a video are used to test the performance of IAPKM and IAPNA. Traditional AP clustering is also put into practice to supply standard performance. Experimental results show that IAPKM and IAPNA can realize equal clustering performance with established AP clustering on all the data sets. For now, the time cost is considerably reduced in IAPKM and IAPNA. Both the effectiveness and the efficiency make IAPKM and IAPNA gifted to be well used in incremental clustering tasks. We point out that the obscurity of extending AP in dynamic data clustering is that, the pre-existing objects have customary definite relationships (non zero responsibilities and non zero availabilities) between each other after affinity propagation, while new objects’ relationships with other objects are motion-less at the initial level (zero responsibilities and zero availabilities). Objects added at different time are at the different statuses, so its firm to find a good exemplar set by just progressing affinity propagation in this case.

Related Work:
Zhang etal. proposed a streaming AP clustering algorithm. In their work new object is allocate to an exemplarif fit principle is satisfied. Otherwise, it is put into areservoir. When the size of reservoir is big enough, traditional AP is re-implemented to empty reservoir. Ott et al. pointed out that, in Zhang et al.’s work, AP clustering was recomputed almost for every recently observed data point. This visibly didn’t work when real-time performance was required. Consequently, they better the competence of streaming AP clustering by limiting the numbers of recomputing. Shi et al. proposed a semi-supervised based incremental AP clustering, and applied it in text clustering.

Literature Survey:
The Author, X.H. Shi, (Et .Al), Aim In [1],a semi-supervised system called incremental affinity propagation clustering is proposed in the paper. In the scheme, the pre-known information is represented by fiddle with similarity matrix. Moreover, an incremental revise is applied to strengthen the prior knowledge. To observe the efficacy of the method we use it to text clustering problem and describe the specific method accordingly. The method is practical to the benchmark dataset Reuters-21578. Numerical results show that the proposed method performs exceptionally well on the data set and has generally compensation over two other generally used clustering methods.

The Author, Adil M. Bagirov Julien Ugon(ET .Al)
Aim In [2],The k-means algorithm and its differences are known to be fast clustering algorithms. Though they are responsive to the choice of starting points and are incompetent for unravel clustering problems in large data sets. Of late incremental approaches have been developed to determine difficulties with the alternative of starting points. The global k-means and the modified global k-means algorithms are based on such an approach. They iteratively add one cluster centre at a time. Numerical experiments show that these algorithms very much improve the k-means algorithm. This makes both algorithms time overwhelming and reminiscence demanding for
clustering even reasonably great datasets. In this paper, a new version of the modified global k-means algorithm is proposed. We set up supplementary cluster function to produce a set of starting points two-faced in different parts of the dataset.

**Problem Definition:**
There are diverse types of clustering. On the other hand nearly every one of the clustering algorithms was intended for determine outline in static data. This induces additional requirements to traditional clustering algorithms to summarily method and précis the massive amount of continually arriving data. It also needs the ability to get used to changes in the data distribution, the aptitude to detect up-and-coming clusters and distinguish them from outliers in the data and the ability to join old clusters or abandon end ones. Clustering or cluster psychiatry is a vital subject in data mining. It plans at separator a dataset into some groups often referred to as clusters such that data points in the same cluster are additional analogous to each other than to those in other clusters.

**Proposed Approach:**
AP clustering is an exemplar-based method that realized by handover each data point to its nearest exemplar, where exemplars are recognized by passing messages on bipartite graph. There are two kinds of messages passing on bipartite graph. They are responsibility and availability, jointly called ‘affinity’. AP clustering can be seen as a request of belief propagation, which was pretend by Pearl to grip deduction problems on probability graph. The objective of this paper is to offer a active alternative of AP clustering, which can attain comparable clustering recital with traditional AP clustering by just regulating the current clustering results according to new arriving objects, somewhat than re-implemented AP clustering on the whole dataset. We lengthen a recently proposed clustering algorithm, affinity propagation (AP) clustering, to grip dynamic data. Several experiments have shown its reliable advantage over the preceding algorithms in static data.

**Proposed Methodology:**
**Admin:**
Clustering the dynamic objects which are inserted by user. Firstly picking out some special objects that called exemplars, and then associating each left object to its nearest exemplar.

**Incremental AP Clustering Based on K-Medoids:**
Objects arriving at different time step are at the different statuses, so it is not likely to find the correct exemplar set by simply continuing affinity propagation.

**Incremental AP Clustering Based on Nearest Neighbor Assignment:**
A technique of Nearest-neighbor Assignment is employed to construct the relationships between the new arriving objects and the previous objects. NA means that the responsibilities and availabilities of the new arriving objects should be assigned referring to their nearest neighbors. NA is proposed based on such a fact that if two objects are similar, they should not only be clustered into the same group, but also have the same relationships.

**User:**
After authentication of user adding dynamic objects. These dynamic objects are clustered according to proposed technique.

**Test By Labeled Data Sets:**
The sum of similarities is one of the most widely used criteria. In some cases, different clustering result can obtain comparable external dispersity and internal dispersity. Therefore, we use labeled data sets to evaluate the proposed algorithms. An advantage is that we can not only evaluate the clustering algorithms by dispersity, but also by some other indicators, e.g., mutual information, clustering accuracy. According to the object function of exemplar-based clustering.

**Distance Calculation:**
The measure of similarity between two objects is also an important. In this project negative square root of Euclidean distance is adopted.

**Euclidean Distance:**
\[ s(i, j) = -\sqrt{\|x_i - x_j\|^2}. \]

**Preference Range Computing Algorithm**
- **Input:** \( s(i,k) \) : The similarity between point \( i \) and point \( k \) ( \( i \neq k \) )
- **Output:** The maximal value and minimal value of preferences: \( p_{\text{max}} \), \( p_{\text{min}} \)
- **Step 1:** Initialize \( s(k, k) \) to zero: \( s(k, k) = 0 \)
Step 2: Compute the maximal value of Preferences:
\[ P_{\text{max}} = \max \{ s(i, k) \} \]

Step 3: Compute the minimal value of Preferences

Step 3.1: Compute the net similarity when the number of clusters is 1:
\[ \text{dpsim}_1 = \max \{ \sum s(i,j) \} \]

Step 3.2: Compute the net similarity when the number of clusters is 2:
\[ \text{dpsim}_2 = \max \{ \sum \max \{ s(i,k), s(j,k) \} \} \]

Step 3.3 Compute the minimal value of Preferences:
\[ P_{\text{min}} = \text{dpsim}_1 - \text{dpsim}_2 \]

Adaptive Affinity Propagation Clustering Algorithm

Input: \( s(i,k) \): the similarity between point \( i \) and point \( k \) (\( i \neq k \))

Output: The clustering result

Step 1: Apply Preferences Range algorithm to computing the range of preferences:
\[ [P_{\text{min}}, P_{\text{max}}] \]

Step 2: Initialize the preferences: preference = \[ P_{\text{min}} - P_{\text{step}} \]

Step 3: Update the preferences: step preference = preference + \( p \)

Step 4: Apply Incremental Affinity Propagation algorithm to generating \( K \) clusters

Step 5: Terminate until Sill is largest.

Results:

The generated result graph specifies the performance of projected algorithms and it takes less number of iterations for better clustering.

II. Enhancement:

Affinity propagation clustering algorithm suffers from one limitation that it is hard to know the value of the parameter preference which can yield an optimal clustering solution. This limitation can be overcome by a method named, adaptive affinity propagation. The technique first finds out the range of preference, then searches the space of preference to find a good value which can optimize the clustering result.

Conclusion:

The proposal of NA is based on such an idea that” if two objects are alike, they should not only be clustered in to the same group, but also have the same statuses”. Both the two ideas are important, and will be extremely obliging in lively clustering intend. Incremental clustering is only a bough of dynamic data clustering. Some other dynamic clustering problems are also of immense meaning. How to apply the two ideas to other dynamic data clustering tasks, such as streaming data clustering, is a note worthy future work.

Future Work:

Dynamic data clustering tasks and streaming data clustering is significant future work. Improve performance to measure similarity between objects is also great importance.

References:


