Incremental Affinity Propagation Clustering
Based on Message Passing

ABSTRACT:
Affinity Propagation (AP) clustering has been successfully used in a lot of clustering problems. However, most of the applications deal with static data. The affinity propagation based clustering algorithm is then individually applied to each object specific cluster. Using this clustering method, we obtain object specific Exemplars together with a high precision for the data associated with each exemplar. We perform recognition using a majority voting strategy that is weighted by nearest neighbor similarity. This paper considers how to apply AP in incremental clustering problems. Firstly, we point out the difficulties in Incremental Affinity Propagation (IAP) clustering, and then propose two strategies to solve them. Correspondingly, two IAP clustering algorithms are proposed. They are IAP clustering based on K-Medoids (IAPKM) and IAP clustering based on Nearest Neighbor Assignment (IAPNA). Five popular labeled data sets, real world time series and a video are used to test the performance of IAPKM and IAPNA. Traditional AP clustering is also implemented to provide benchmark performance. Experimental results show that IAPKM and IAPNA can achieve comparable clustering performance with traditional AP clustering on all the data sets.
Meanwhile, the time cost is dramatically reduced in IAPKM and IAPNA. Both the effectiveness and the efficiency make IAPKM and IAPNA able to be well used in incremental clustering tasks.

EXISTING SYSTEM:

CLUSTERING, or cluster analysis, is an important subject in data mining. It aims at partitioning a dataset into some groups, often referred to as clusters, such that data points in the same cluster are more similar to each other than to those in other clusters. There are different types of clustering. However, most of the clustering algorithms were designed for discovering patterns in static data. This imposes additional requirements to traditional clustering algorithms to rapidly process and summarize the massive amount of continuously arriving data.

PROBLEM DEFINITION:

✓ It also requires the ability to adapt to changes in the data distribution, the ability to detect emerging clusters and distinguish them from outliers in the data, and the ability to merge old clusters or discard expired ones.

PROPOSED SYSTEM:

we extend a recently proposed clustering algorithm, affinity propagation (AP) clustering, to handle dynamic data. Several experiments have shown its consistent
superiority over the previous algorithms in static data. AP clustering is an exemplar-based method that realized by assigning each data point to its nearest exemplar, where exemplars are identified by passing messages on bipartite graph. There are two kinds of messages passing on bipartite graph. They are responsibility and availability, collectively called ‘affinity’. AP clustering can be seen as an application of belief propagation, which was invented by Pearl to handle inference problems on probability graph. Compared with the previous works, another remarkable feature of our work is that the IAP clustering algorithms are proposed based on a message-passing framework. That’s, each object is a node in a graph, and weighted edges between nodes correspond to pair wise similarity between objects. When a new object is observed, it will be added on the graph and then message passing is implemented to find a new exemplar set. Because that only one, or a few of nodes’ entering will not change the structure of the whole graph a lot, a local adjustment of availabilities and responsibilities is enough. Therefore, messages passing on graphs will re-converge quickly. Based on these features, the IAP clustering algorithms proposed in this paper don’t need to re-implemented AP clustering on the whole data set, nor need to change the similarities between objects.
ADVANTAGES OF PROPOSED SYSTEM:

1. A great deal of time can be saved, which makes AP clustering efficient enough to be used in dynamic environment.

2. The goal of this paper is to propose a dynamic variant of AP clustering, which can achieve comparable clustering performance with traditional AP clustering by just adjusting the current clustering results according to new arriving objects, rather than re-implemented AP clustering on the whole dataset.

HARDWARE REQUIREMENTS:

- Hardware - Pentium
- Speed - 1.1 GHz
- RAM - 1GB
- Hard Disk - 20 GB
- Floppy Drive - 1.44 MB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA
SOFTWARE REQUIREMENTS

✓ Operating System
✓ Windows Technology
✓ Java and J2EE Web Technologies: Html, JavaScript, CSS, IDE
✓ My Eclipse Web Server: Tomcat
✓ Tool kit: Android Phone Database, My SQL Java Version: J2SDK1.5