

JIGSAW ACTIVITY (INFS 755, Fall 2008- Huzefa Rangwala)

As part of this activity we learn about the various clustering algorithms. Each one of us will be assigned to read one of the five [A – E] research papers describing a clustering algorithm. Please read the paper thoroughly, keeping the following questions in mind. Try to make notes, if you feel like.

1. How does the particular algorithm work? What is the type of this algorithm?
2. Performance of the algorithm. Was the algorithm compared to others ? How does it compare to the k-means algorithm ? Would there be any advantage of using this algorithm ? What about complexity (run-time as well as memory) ?
3. Are there any applications discussed in the paper ? What are the datasets used to evaluate the algorithm ?
4. Think of ways to improve upon the algorithm. We always like to do better !!

Below are the list of papers. You are responsible for the paper highlighted in your sheet. The paper will also be attached to your sheet for convenience.

- A. DBSCAN: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise; Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu (KDD96)
- B. CURE: An Efficient Clustering Algorithm for Large Databases; Sudipto Guha, Rajeev Rastogi, Kyuseok Shim (ACM 1998)
- C. Chameleon: Hierarchical Clustering Using Dynamic Modeling; George Karypis, Sam Han, Vipin Kumar (IEEE 1999)
- D. CLUTO: A Comparison of Document Clustering Techniques; Michael Steinbach, George Karypis, Vipin Kumar (SIAM 2000)
- E. **BIRCH: An Efficient Data Clustering Method for Very Large Databases; Tan Zhang, Raghu Ramakrishnan, Miron Livny (SIGMOD 1996)**

Once you have read the papers thoroughly, you will be deemed as an expert for your assigned paper.

On Monday 10/27/2008, we will discuss these papers in details. As part of phase 1, we will meet in our expert groups i.e students reading the same paper will come together and discuss the key points about the paper and answer a few questions posed. This will strengthen our ideas about the paper read. [15 min]

Having studied the paper and discussed with our fellow experts, we will gather in groups such that each one of you will be from a different expert group. Now we will share very the ideas about our method studied, try contrasting them, think about the best methods, and also try to come up with new methods. [30 min]

Finally we will come together as a class to discuss some of the points that stand out. In this was we have a fair understanding of the various methods. Thanks for participating and be ready to discuss, debate and question.

Based on your participation or enthusiasm, I will award 5% credit towards the final grade.

-Huzefa