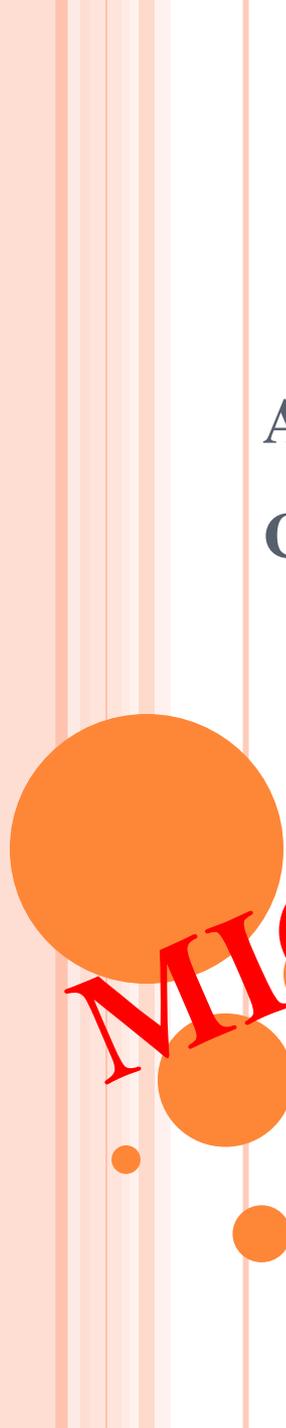


**ADAPTIVE ENSEMBLING OF SEMI-SUPERVISED  
CLUSTERING SOLUTIONS**

**MICANS INFOTECH**



# EXISTING SYSTEM

- Conventional cluster ensemble approaches have several limitations: they do not consider how to make use of prior knowledge given by experts, which is represented by pair wise constraints.
- Pair wise constraints are often defined as the must-link constraints and the cannot-link constraints.
- The must link constraint means that two feature vectors should be assigned to the same cluster, while the cannot-link constraints means that two feature vectors cannot be assigned to the same cluster.
- Most of the cluster ensemble methods cannot achieve satisfactory results on high dimensional datasets. Not all the ensemble members contribute to the result.



# ABSTRACT

- Conventional semi-supervised clustering approaches have several shortcomings, such as not fully utilizing all useful Must-link and cannot-link constraints, not considering how to deal with high dimensional data with noise, and not fully addressing
- The need to use an adaptive process to further improve the performance of the algorithm. In this paper, we first propose the transitive closure based constraint propagation approach, which makes use of the transitive closure operator and the affinity propagation to address the first limitation.
- Then, the random subspace based semi-supervised clustering ensemble framework with a set of proposed



## CONT..

- Confidence factors is designed to address the second limitation and provide more stable, robust and accurate results
- Next, the adaptive semi-supervised clustering ensemble framework is proposed to address the third limitation, which adopts a newly designed adaptive process to search for the optimal subspace set finally, we adopt a set of nonparametric tests to compare different semi-supervised clustering ensemble approaches over multiple datasets.
- The experimental results on 20 real high dimensional cancer datasets with noisy genes and 10 datasets from UCI datasets and keel datasets.
- The proposed approaches work well on most of the real-world datasets. It outperforms other state-of-the-art approaches on 12 out of 20 cancer datasets, and 8 out of 10 UCI machine



# DISADVANTAGE

- They do not consider how to make full use of must-link constraints and cannot-link constraints.
- Some methods do not take into account how to deal with high dimensional data with noise.

**MICANS INFOTECH**



# PROPOSE SYSTEM

- Cluster ensemble, is referred to as consensus clustering, one of the important research directions in ensemble learning.
- It can be divided into two stages: the first stage aims at generating a set of various ensemble members, while the objective.
- The second stage is to choose a suitable consensus function to summarize the ensemble members and search for an optimal unified clustering solution.
- To attain these objectives, we use a knowledge reuse framework which integrates multiple clustering solutions into a unified one.



## CONT..

- While there are various kinds of cluster ensemble techniques, some of them consider how to handle high dimensional data clustering,
- How to make use of prior knowledge of the given data sets most of the conventional cluster ensemble methods do not consider how to handle the over fitting problem, and cannot obtain satisfactory results when handling high dimensional data

**MICANS INFOTECH**



# ADVANTAGES

- The research works are efficiency or scalability of semi-supervised clustering.
- A number of researchers also address the semi supervised clustering problem using nonnegative matrix factorization



# SOFTWARE REQUIREMENT

## ○ **HARDWARE REQUIREMENT:**

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Sony.
- Ram : 512 Mb.

**MICANS INFOTECH**



## CONT...

- **SOFTWARE REQUIREMENT:**

- Operating system : Windows XP.
- Coding Language : ASP. Net with C#
- Data Base : SQL Server 2005.

**MICANS INFOTECH**



# CONCLUSION

- This paper has proposed an adaptive semi-supervised clustering ensemble framework (A-RSEMICE) for high dimensional data clustering.
- When compared with traditional semisupervised clustering approaches A-RSEMICE is characterized with the following three properties
- A new proposed transitive closure based constraint propagation approach adopted to make use of the transitive closure operator and the constrain propagation to fully explore how to use all useful must-link and cannot-link constraints.



## CONT...

- A-RSEMICE adopts the random subspace based semi-supervise clustering ensemble framework to integrate the clustering solutions obtained by different transitive closure operators from multiple datasets into a unified clustering solution.
- A newly designed adaptive process is adopted to search for the optimal subspace set. We performed a thorough analysis of the properties of A-RSEMICE in the experiments, and draw conclusion

MICANS INFOTECH

