Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2004 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

December 2004

An Interactive Clustering Methodology for High Dimensional Data Mining

Haibo Wang
University of Mississippi

Bahram Alidaee
University of Mississippi

Fred Glover
University of Colorado

Gary Kochenberger
University of Colorado at Denver

An Interactive Clustering Methodology for High Dimensional Data Mining

Haibo Wang
University of Mississippi
University, MS 38677
chwang@olemiss.edu

Bahram Alidaee
University of Mississippi
University, MS 38677
balidaee@bus.olemiss.edu

Fred Glover
University of Colorado
Boulder, CO 80304
Fred.Glover@colorado.edu

Gary Kochenberger
University of Colorado at Denver
Denver, CO 80217
Gary.Kochenberger@cudenver.edu

Abstract

This study develops an interactive clustering model and methodology for high dimensional data. The similarity index is calculated with proposed formulation for both continuous-scaled and nominal-scaled attributes. The associated similarity score values are constructed into a graph as clique partitioning problem, which can be reformulated into a form of unconstrained quadratic program model and then solved by a Tabu search heuristic incorporating strategic oscillation with a critical event memory. The complexities of high dimensional data mining are discussed from both mathematical modeling and computational algorithm points of view.

Keywords: Clustering, High-Dimensional, Tabu Search, Clique Partitioning Problem

1. Introduction

Data mining is a process of knowledge discovery and an inter-discipline of science involving machine learning, artificial intelligence, statistics, mathematical programming and optimization. Clustering is the most common approach in data mining. Clustering in data mining is intended to divide objects into groups so that objects within groups are homogeneous and have a high degree of similarity. The motivation of clustering can be either partition data into a pattern which is easy to manage or try to discover the natural subclasses in the data. From a methodology viewpoint, clustering encompasses the methods and techniques from databases, machine learning, statistics, artificial intelligence and optimization. In this study, clustering data mining problem is considered as an application of clique partitioning problem (CPP).

From data mining point of view, CPP is also known as the incapacitated clustering problem (UCP). In the absence of capacities, many such clustering problems can be modeled as CPP. Since CPP is NP-hard in general, heuristic methods are needed to solve large sized clustering problems. In this study objects are represented as nodes on a connectivity graph with the similarity score as the edge weight between each pair of objects. Given this graph, objects are put into clusters with a high degree of similarity by solving the CPP problem.

The model and methodology proposed in this study can be used in data mining, customer relationship management (CRM) and business intelligence (BI). The approach from this study can help business decision support by clustering large amount high dimensional data from CRM or BI applications.

2. Model and Solution Methodology

In this study, the clustering data mining problem is formulated as CPP. The associated CPP model is solved by first re-casting it into the form of unconstrained binary quadratic programming, which is then solved by a tabu search heuristic. The associated CPP model is solved by first re-casting it into the form of unconstrained binary quadratic programming, which is then solved by a tabu search heuristic.

2.1 CPP Model

There are various definitions for CPP. This study will adopt the most commonly accepted definition from Grotschel and Wakabayashi (1990), which is to find a clique partition with maximum weight and can be described as:

Given a complete graph $G = G(V, E)$ with edge weights $w_{ij} \in R$ for all $(i, j) \in E$, a clique partition exists if there is a subset $A \in E$ of the edge set, where a partition of V into nonempty, disjoint sets $V_1 \ldots V_k$ and each $V_p (p = 1 \ldots k)$ forms a complete sub-graph (also known as a clique) and $A = \bigcup_{p=1}^{k} \{\{i, j\} \mid i, j \in V_p, i \neq j\}$, the weight of such a clique partition A is defined as $\sum_{(i,j)\in A} w_{ij}$. The objective is to maximize the sum of $\sum_{(i,j)\in A} w_{ij}$. In mathematical form, CPP is presented as:

$$\max \quad \sum_{(i,j)\in E} w_{ij} x_{ij} \tag{1-1}$$

st

$$x_{ij} + x_{ir} - x_{jr} \leq 1 \quad \forall \textit{ all distinct } i, j.r \in V$$

$$x_{ij} \in \{0,1\} \textit{ for all } \{i, j\} \in E$$

where w_{ij} is edge weight (unrestricted in sign) and $x_{ij} = 1$ when edge (i, j) is in one of clique. The variable in the model is defined as the edge of the graph. The constraints in this model are known as triangle inequalities constraints in the scale of $\frac{n(n-1)(n-2)}{2}$. Grotschel and Wakabayashi (1989) proofed that these constraints can always give integral solution.

The new model proposed for CPP in this study is a quadratic program model and is defined as following:

$$\max \quad \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \sum_{k=1}^{k\_\max} x_{ik} x_{jk}$$

st

$$\sum_{k=1}^{k\_\max} x_{ik} = 1 \qquad i = 1, \ldots, n$$

$$x_{ik} \in \{0,1\}$$

$$x_{ik} + x_{jk} \leq 1, \quad \textit{for all } (i, j) \notin E \quad \textit{and} \quad k = 1, \ldots, K\_\max \tag{2-1}$$

Where $k\_\max$ is the maximum number of cliques allowed, n is the number of nodes w_{ij} is the edge weight and x_{ik} is equal to 1 if node i is assigned to clique k and is 0 otherwise. The number of variables in this model is $n \times k\_\max$. The number of constraints is n in this constrained quadratic program (CQP) model, which is much smaller than the standard linear program model. Some researchers have developed a mechanism to reduce the number of constraints in CQP model by introducing a penalty functions and convert CQP model into an UBQP model (Glover, Kochenberger, Alidaee and Amini, 1999; Beasley, 1998). The simple example of Supplement A illustrates the potential applicability of the transformation for many quadratic and linear program problems.

The new model reduces the number of variables in CPP comparing to the standard formulation for CPP known as triangle inequalities constraints model. The new model motivates new interest for some applications which can be recast into CPP. Then the new GT model can be reformulated into unconstrained binary quadratic program mode.

2.2. Tabu Search with Strategic Oscillation and Critical Event Memory

A Tabu search heuristic method is used to solve the UBQP model and it exploits the tabu search memory on the UBQP by incorporating strategic oscillation with critical event memory. This solution methodology offers several advantages over existing methods. It not only solves the problem faster but also solves larger problems which can not be solved by many existing methods. The quality of the cliques by this method is better than that of other methods in many applications. It also offers the flexibility of solving the CPP while incorporating the domain constraints. In clustering problem, it allows people to cluster data interactively. It combines the features of hierarchical clustering algorithms and partitional clustering algorithms.

Tabu search is a meta-heuristic method introduced by Glover (1997). It directs and manages the search process by incorporating short-term and long-term memories to explore the forbidden regions in the search space. The basic steps for a simple maximization problem with tabu search procedure are:

1. Select an initial seed solution $s \in T$ and set the best solution $s_{best} = s$ and iteration $k = 0$
2. Set $k = k + 1$ and Generate a subset J in neighbor of current solution s_k
3. Find the best solution j in the subset J to satisfy the criteria function
4. If the current new solution is feasible and better than s_{best}, let $s_{best} = s_k$
5. If reach one of the following conditions, stop, otherwise go to step 2:
   C1: s_{best} is optimal solution or within a gap to optimal solution
   C2: Can not find improvement to s_{best} within a number of iterations
   C3: k reaches the certain threshold

Strategic oscillation is the advanced technique for finding solution in forbidden regions by crossing back and forth between feasible and infeasible regions. The critical event memory uses a parameter associated with the search process and strategic oscillation when it reaches a stage of generating a subclass of locally optimal solution.

3. Reformulation of Clustering Data Mining Problem

All data mining related problems in this study come from UCI repository and open literatures. They are stored as comma delimited file originally. The value of the attribute is alphanumeric. There are different types of measurement for data relationship. For the data within the group, there are Euclidian distance (square), manhattan distance, and chebychev distance; For the inter-group data, there are average linkage, centroid linkage, complete linkage, single linkage, average to centroid linkage, hausdorff linkage; for the intra-group data, there are complete diameter, average diameter and centroid diameter. In this study, similarity index is used to measure the relationship of objects in the data set.

3.1. How to Cast CDM Problem into CPP

There are four steps in the approach of clustering data on the high-dimensional space in this study:

1. Transform the high dimensional data record into one dimensional similarity index matrix.
2. Re-cast the problem into CPP by constructing a graph from similarity scores.
3. Solve the CPP in a form of an UBQP model with a tabu search heuristic.
4. Analyze output and abstract knowledge from the data.

There are nominal-scaled data records, continuous-scaled data records and mixed types data in the data repository. If the values of all attributes (dimensions) are discrete and independent and there is not relationship between the attributes (dimensions), the data record is nominal-scaled. Otherwise it is continuous-scaled attributes.

All models and methodologies are based on the notion of distance, which has different meanings on different types of data records. There are nominal-scaled data records, continuous-scaled data records and mixed types data in the real data. If the values of all attributes (dimensions) are discrete and independent and there is not relationship between the attributes (dimensions), the data record is nominal-scaled. Otherwise it is continuous-scaled attributes. Many high-dimensional data are mixed nominal-scaled and continuous-scaled and very difficult to be clustered. For nominal-scaled attributes, the similarity index (SI) is the common distance measure for the relationship between a pair of objects. SI is the ratio or formula expressing the ratio of one object to another object on the degree of similarity. Euclidean distance is a common measure for continuous-scaled attributes. It works perfectly with one dimensional data. For multiple dimensional data record, the usual Euclidean distance might not represent the data accurately. When the usual Euclidean distance is applied to multiple dimensional data record, the basic assumption is that the range of values in each dimension is same or very close, which is not always true in many data record. When the range of values in each dimension is different, it will guide the clustering method into an unattractive result. In Figure 1, the point 1 and point 3 will be grouped into two different clusters and point 3 and point 4 will be grouped into the same cluster since the range of values in one dimension is large than another dimension. To group point 1 and point 3 into the same group, a standardized Euclidean distance is introduced. When the data has a non-homogeneous distribution where the variance across one dimension is larger than the variance across another dimension, the standardized Euclidean distance will provide a better result. Using standardized Euclidean distance instead of usual Euclidean distance in this study can equalize the variances on different dimensions and produce better clustering result.

There are two approaches to the binary representations by either comparing a pair of objects with logical bit or converting the level of attribute values into multiple binary attributes. The size of binary representation of the first approach is generally larger than the second approach. In the logical bit approach, when two objects have the same values on the same attribute, the

logical bit is 1 and 0 otherwise. Binary data record (with 0 and 1 as the values in all attribute) is a special case of nominal- scaled data. If the number of distinct values of the attribute is two, the data can be treated as either binary data (converting one value in the attribute as '1' and the other as '0') or regular nominal-scaled data (split one original attribute into two binary attributes). Some data records have a mix of binary attributes and non-binary attributes. For binary data records (with all binary attributes), the similarity index is calculated with Hamming Distance method. For the data-record with all non-binary attribute, Erlich et al (2003) proposed a Positive Attribute Distance (PAD) for similarity measurement. They also proposed a Pair Similarity Index (PSI) for similarity measurement (Erlich et al 2002).

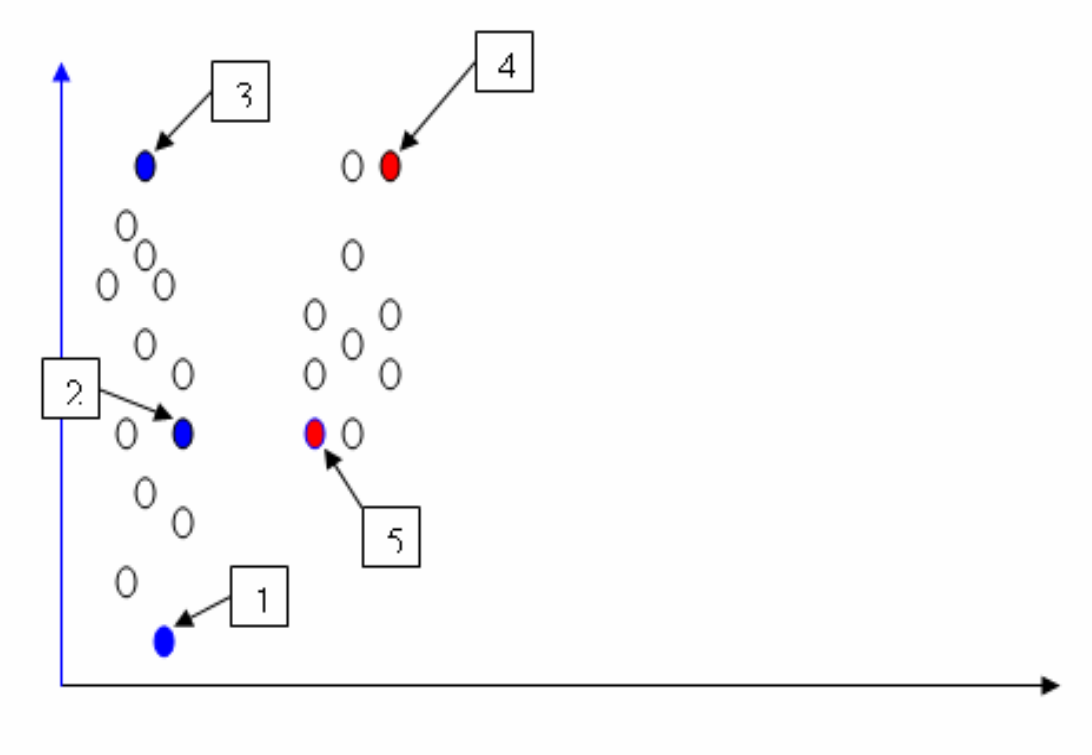

Figure 1. Data in 2-Dimensional Spaces

This study formulates a similarity index with the Original Attribute Distance (OAD). Like PAD and PSI, OAD uses a binary representation of the existence or absence of an attribute in a given object being observed. If there is a non-binary attribute with two distinct values, it can be converted into two binary attributes (chosen by the user and defined as type-1 attributes) or be converted into one binary attribute with one nominal value as '1' and the other nominal value as '0' (defined as type-2 attribute). The resulting binary string representing the object is then used to calculate distance to other strings using the value 1 for '1' bit and 0 for '0' bit on the non-binary attributes (including type-1 attributes), XOR (exclusive OR) logical operation on the binary attributes (including the type-2 attributes), and the total number of original attributes. For the binary representation of each object, there is a binary string $\boldsymbol{s} = \boldsymbol{s}_{ba} + \boldsymbol{s}_{nba}$, where $\boldsymbol{s}_{ba}$ is the substring of $\boldsymbol{s}$ for original binary attributes (including type-2 attributes), and $\boldsymbol{s}_{nba}$ is the substring of $\boldsymbol{s}$ for original non-binary attributes (including type-1 attributes). The OAD similarity index ($\boldsymbol{S}_n$) is defined as:

$$S_n = \frac{V_{ba} + V_{nba}}{N_{na}}$$

$$V_{ba} = L_{ba} - HD(s_{ba1}, s_{ba2})$$

$$HD(s_{ba1}, s_{ba2}) = s_{ba1} \oplus s_{ba2}$$

where V_{nba} is the sum of the values of '1' and '0' bit on both s_{nba1} and s_{nba2} when they appear on the same location. For the logical bit approach, V_{nba} is the sum of the values of '1' and '0' bit for the pair of objects where V_{ba} is zero for logical bit binary representation. V_{ba} is the Hamming Distance similarity index values for substring s_{ba}. L_{ba} is the length of substring s_{ba}. $HD(s_{ba1}, s_{ba2})$ is the Hamming Distance between substrings s_{ba1} and s_{ba2}, $\oplus$ denotes the logical operation XOR. N_{na} is the total number of the original nominal-scaled attributes. It can be used in the all binary attribute data record, non-binary attribute nominal-scaled data record or mixed attribute nominal-scaled data record. It represents the same value no matter which transformation method is chosen by the user. For the high-dimensional data with both continuous-scaled and nominal-scaled attributes, this study proposes a similarity index formulation for a pair of objects as following:

$$S = R_c S_c + R_n S_n$$

$$R_c = \frac{N_{ca}}{N_{ta}}$$

$$R_n = \frac{N_{na}}{N_{ta}}$$

$$S_c = 1 - d'_{se}$$

where N_{ca} is the number of continuous-scaled attributes. N_{ta} is the total number of attributes. N_{na} is the number of nominal-scaled attributes. S_c is the similarity index value for continuous-scaled attributes.

The term similarity score bases on the notation of similarity indexes. If $S = 1$, two objects are identical. In another word, these two objects have 100% degree of similarity. If $S = 0$, two objects have absolute diversity. These two objects have 0% degree of similarity. Similarity Threshold (ST) can be chosen by the user based on their domain knowledge and computational experiences. (The average value of similarity index can be a good starting point.) If the similarity index of a pair of objects is larger than ST, these two objects have a high degree of similarity and are intended to be grouped into a cluster. Otherwise they have less likelihood to be grouped into a cluster. The Similarity Score (SS) is the percentage value of the difference between a degree of similarity of a pair of objects and the similarity threshold. After the similarity indexes are calculated from high dimensional data, it is transformed into similarity score with this equation:

$$ST = \frac{\sum_{i=1}^{k} S_i}{k}$$

$$SS = (S - ST) * 100$$

$$k = \frac{n(n-1)}{2}$$

4. Clustering Result and Output Validation

There are 19 data mining problems from UCI repository and open literatures are tested by the new model and methodology. The computational time and number of final clusters are reported in Table 1. The computation time depends on the number of variables, which is the product of number of objects and k_max and does not depend on the number of attributes. It can solve the problem very quickly with very large numbers of patterns such as millions of patterns but smaller number of objects. The data from medical science, chemical grouping, computational biology, web document grouping and marketing might fit into this model very well. For example, 100 DNA strings with over 10000 nucleotides units have one million patterns. This model can cluster such data very quickly. A web document might have several thousand keywords, one thousand such documents have more than millions of patterns. In Table 1, there are different types of data. Some data set only have binary attributes, which can be measured by HD index. Some data set only have non-binary attribute, which can be measured by PAD, PSI or OAD. Some data set have mixed binary and non-binary attributes, which can be measured by OAD. Some data set only have continuous-scaled attributes, which can be measured by normalized Euclidean distance. Some data set have both continuous and nominal-scaled attributes, which can be measured by the proposed formulation. The computational time depends on the number of variables and does not depend on the number of dimensions. The data set with large number of objects but small number of attributes tends to be difficult to cluster.

Table 1. Clustering Results on Data Mining Problems

Data Source Name	Number of Objects	Number of Dimensions (Attributes)	Number of Patterns	k-max	Number of Variables	Number of Clusters	Time (seconds)
Balance-scale**	625	5	3100	12	7500	12	1300
Balloon1**	16	5	80	4	64	2	0.28
Balloon2**	20	5	100	4	80	2	0.42
Balloon3**	20	5	100	4	80	2	0.43
Balloon4**	20	5	100	4	80	2	0.40
Breast-cancer**	699	10	6990	12	8388	8	1600
Cleveland***	303	8	2424	20	6060	15	560
DNA**	1186	61	72346	7	8302	7	1400
Flare1***	323	13	4199	10	3230	7	230
Flare2***	1066	13	13858	8	8528	5	1800
House-1984**	425	17	7225	10	4250	5	260
@House-2000**	435	21	9135	10	4350	4	460
@House-2002**	436	21	9156	10	4360	3	470
Lenses**	24	5	120	4	96	3	0.54
Lung-cancer***	32	57	1824	10	320	4	2.7
SPECT*	187	23	4301	20	3740	15	200
SPECTF*	80	23	1840	10	800	3	2.2
Sponge-data**	76	46	3496	10	760	2	11
Zoo***	101	17	1717	10	1010	7	17

***** binary attribute data record
****** non-binary attribute data record
*** mixed binary and non-binary attribute data record
@ data from ADA website (www.ada.org)

There is no standard measure for clustering problem and it all depends on the context of the study object. For example, the intra-cluster similarity or inter-cluster similarity might not be able to measure the solution quality for DNA microarray data clustering since biochemist is interested in the objects outside any clusters (the lone object). Any of the existing measures

might not be able to measure it. Since the objective of clustering is to group the objects based on their similarity, the inter-cluster similarity and intra-cluster similarity can be used to measure how objects similar to each other in the cluster. One of the reasons for using a general similarity index for clustering data is that both nominal-scaled and continuous-scaled attributes can be measured by similarity. For the continuous data with the relationship between attributes, such similarity index might not be valid. Based on the information from open literatures, the idea of using the ratio between intra-cluster similarity and inter-cluster similarity is new. The intra-cluster similarity measures how close the objects in the cluster are and the inter-cluster similarity measures how the cluster separate from each other. The ratio between these two similarities can measure how an object is placed in the cluster. The objective in clustering data mining problem is to partition the objects into clusters by reducing the inter-cluster similarity and increasing the intra-cluster similarity.

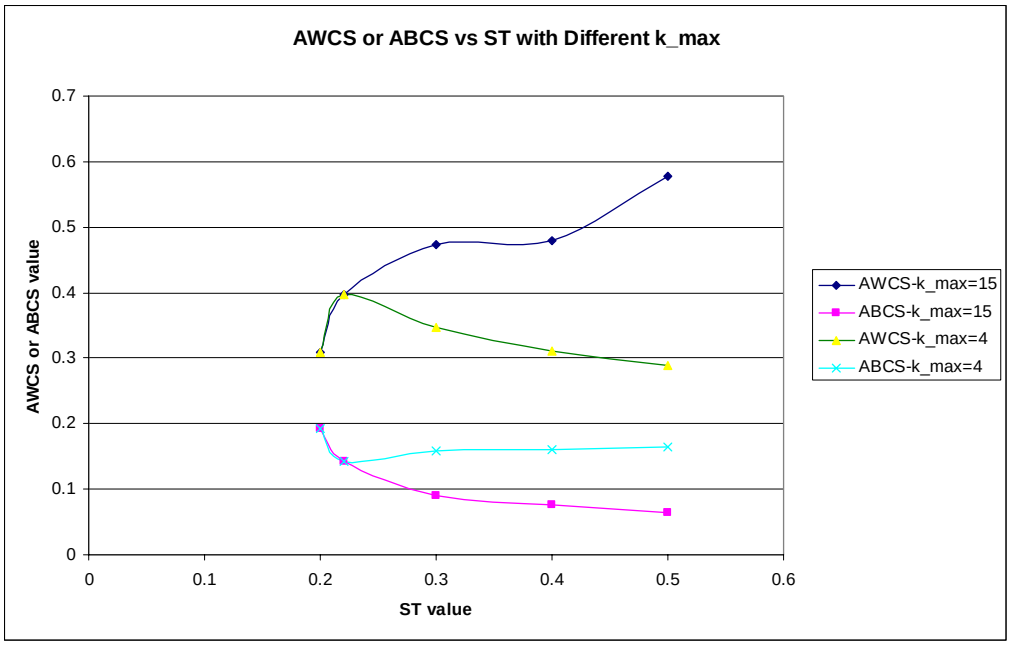

Figure 2. Correlation Among ST value, Average Within-Cluster Similarity and Average Between-Cluster Similarity with Different k_max Values

Different similarity thresholds produce different clusters and it will help researchers to validate the cluster and compare the quality of the cluster. This can also allow the researchers to choose different values of k_max and it can produce different set of clusters, which can be treated as a process of grouping small clusters into big clusters. When a good combination of k_max and similarity threshold is chosen, the researcher can abstract information and knowledge from the clusters which cannot be found with other clustering techniques. Figure 2 & 3 show the relationship among k_max, ST, number of resulting clusters and similarity. When k_max is large than the number of resulting clusters in optimal solution, a larger ST value will produce more clusters with better compactness within the clusters and better separation between the clusters. When k_max is small, a larger ST value can only produce the number of clusters equal to k_max. In both case of k_max value, the average value of ST, which is 0.22 in this data set, produces a small number of clusters with better quality. When ST is less than 0.22, it produces smaller number of clusters with a worse quality. When

ST value is larger than 0.22, the number of clusters depends on k_max. The quality of clusters has a positive correlation with k_max.

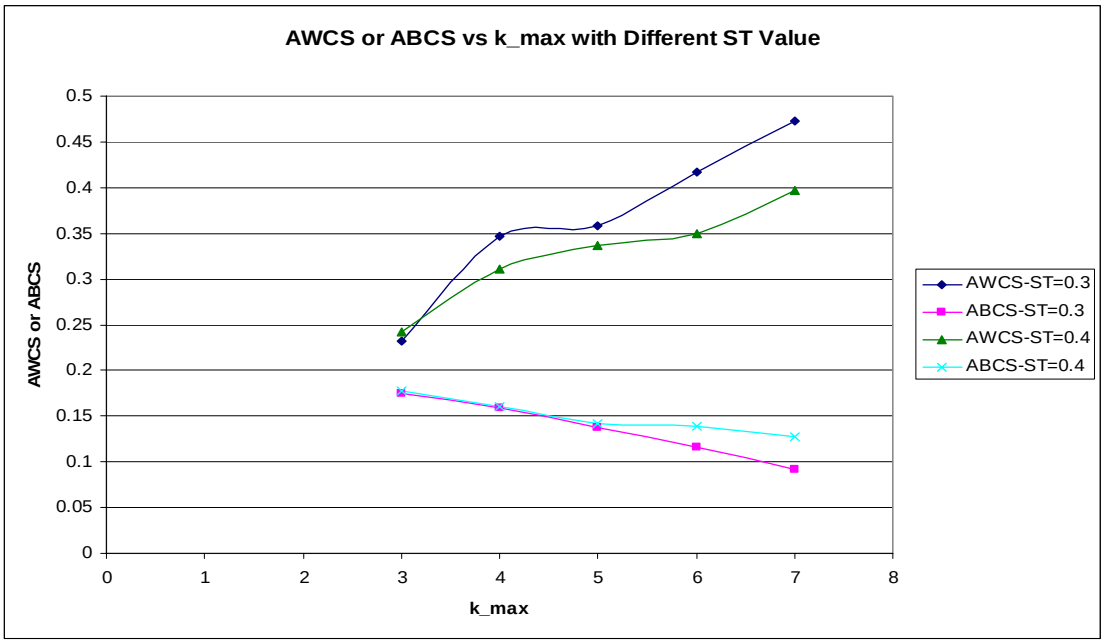

Figure 3. Correlation Among Values of k_max, Average Within-Cluster Similarity and Average Between-Cluster Similarity with Different ST values

5. Concluding Remarks

Computational results show that the new model and methodology generate attractive results on a vast range of applications. New approach has been implemented to recasting these applications into CPP model. For the large problems in general, the new model can find an optimal solution in a matter of minutes as opposed to the weeks it took the commercial software to solve. Sometimes even the LP relaxation of these problems can take hours using previous models. This study provides several interesting findings: the computing time for the CPP depends on the number of variables in the $\hat{Q}$ matrix and choice of tabu tenure value. The maximum number of cliques allowed in the problem can be estimated from an existing method. The number of variables in the $\hat{Q}$ matrix depends on the maximum number of cliques. A close estimation of maximum number of clique can improve the computational time for the model. In clustering data mining problem, the size of $\hat{Q}$ matrix is determined by the number of objects and predetermination on the domain knowledge. The size of $\hat{Q}$ matrix is independent from the number of attributes. This model can solve problems with very large numbers of attributes too difficult to be solved by other models. But it is not so efficient to solve the problem with a large number of objects and small number of attributes. The missing values in the real world data set are common and missing value has very big impact on the low dimensional data clustering. Many techniques have been implemented to predict the value of missing data set. In high dimensional data, the missing values have a smaller contribution to the results. The higher the dimension (number of attributes) is, the less contribution the missing values has to the similarity index values. The approach in this study can cluster data

interactively with two steps: First, the model can allow the researchers to choose different similarity threshold values based on their experience and domain knowledge (incorporating domain constraints to the weight). Different similarity thresholds produce different clusters and the new model will help researchers to validate the cluster and compare the quality of the cluster. Secondly, this model can allow the researchers to choose different values of k_max and it can produce different set of clusters, which can be treated as a process of grouping small clusters into big clusters. This process can be very useful for some applications. When a good combination of k_max and similarity threshold is chosen, the researcher can abstract information and knowledge from the clusters which cannot be found with other clustering techniques. A visualization tool could be developed to help identify attractive clusters. This can be very useful for computational biology and drug design.

This approach combines the features from hierarchical clustering and partitional clustering techniques. Compared to the k-means method, this methodology is robust. In the k-mean method, many researchers have to use principle components analysis to find the value of k at first. There is no standard to validate cluster and evaluate output, it varies from application to application. A reasonable interpretation of the output requires a balance of the implicit assumption and interesting information abstracted from the clusters. The approach in this study allows researchers to validate the cluster from two different types of quality measures. These findings are particularly important in light of the cluster validation for different kinds of data.

References

Beasley, E. "Heuristic Algorithms for the Unconstrained Binary Quadratic Programming Problem", Working Paper, Imperial College, 1998.

Erlich, Z., Gelbard, R. and Spiegler, I. "Data mining by means of binary representation: a model for similarity and clustering," *Information Systems Frontiers* (4:2), 2002, pp.187-197.

Erlich, Z., Gelbard, R. and Spiegler, I. "Evaluating a positive attribute clustering model for data mining," *Journal of Computer Information Systems* (43:3), 2003, pp.100-108.

Glover, F. and Laguna, M. *Tabu Search*, Kluwer Academic Publishers, 1997.

Glover, F., Kochenberger, G., Alidaee, B. and Amini, M. "Tabu search with critical event memory: an enhanced application for binary quadratic programs" In: *Meta-Heuristics, Advances and Trends in Local Search Paradigms for Optimization,* S. Voss, S. Martello, I. Osman, and C. Roucairol (Eds.), Kluwer Academic Publishers, 1999.

Grotschel, M. and Wakabayashi, Y. "A cutting plane algorithm for a clustering problem," *Mathematical Programming* (45), 1989, pp.59-96.

Grotschel, M. and Wakabayashi, Y. "Facets of the clique partitioning polytope," *Mathematical Programming* (47), 1990, pp.367-387.

Supplement

A. An example for 4 nodes and 2 cliques with the node-based model.

$$\max f(x) = w_{12}(x_{11}x_{21} + x_{12}x_{22}) + w_{13}(x_{11}x_{31} + x_{12}x_{32}) + w_{14}(x_{11}x_{41} + x_{12}x_{42})$$
$$+ w_{23}(x_{21}x_{31} + x_{22}x_{32}) + w_{24}(x_{21}x_{41} + x_{22}x_{42}) + w_{34}(x_{31}x_{41} + x_{32}x_{42})$$

st

$$x_{11} + x_{12} = 1$$
$$x_{21} + x_{22} = 1$$
$$x_{31} + x_{32} = 1$$
$$x_{41} + x_{42} = 1$$
$$x_{ij} \in \{0,1\} \quad \textit{for } i = 1,2,3,4 \textit{ and } j = 1,2$$

The UBQP formulation for this example is:

$$\max f(x) = x\hat{Q}x^t$$

where

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{21} & x_{22} & x_{31} & x_{32} & x_{41} & x_{42} \end{bmatrix} \quad x^t = \begin{bmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \\ x_{31} \\ x_{32} \\ x_{41} \\ x_{42} \end{bmatrix}$$

$$\hat{Q} = \begin{bmatrix} P & -P & w_{12} & 0 & w_{13} & 0 & w_{14} & 0 \\ -P & P & 0 & w_{12} & 0 & w_{13} & 0 & w_{14} \\ w_{12} & 0 & P & -P & w_{23} & 0 & w_{24} & 0 \\ 0 & w_{12} & -P & P & 0 & w_{23} & 0 & w_{24} \\ w_{13} & 0 & w_{23} & 0 & P & -P & w_{34} & 0 \\ 0 & w_{13} & 0 & w_{23} & -P & P & 0 & w_{34} \\ w_{14} & 0 & w_{24} & 0 & w_{34} & 0 & P & -P \\ 0 & w_{14} & 0 & w_{24} & 0 & w_{34} & -P & P \end{bmatrix}$$

$$x_{ij} \in \{0,1\} \quad \textit{for } i = 1,2,3,4 \textit{ and } j = 1,2$$