International Journal of Information Technology & Decision Making
Vol. 5, No. 4 (2006) 605–609
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

NEW OPTIMIZATION MODELS FOR DATA MINING

FRED W. GLOVER
University of Colorado
Campus Box 419, Boulder, Colorado 80309, USA
fred.glover@colorado.edu

GARY KOCHENBERGER
Decision Sciences, University of Colorado
School of Business, Denver, Colorado 80217-3364, USA
gary.kochenberger@cudenver.edu

In recent years modern methods of optimization have contributed greatly to the advances in data mining and related areas. These contributions continue today and promise to further advance the state of the art both in terms of modeling innovations and new solution methodologies. In this paper, we present a new modeling and solution methodology for unsupervised clustering. Preliminary computational experience is given to illustrate the approach. This methodology is part of our current research and offers considerable opportunity for additional investigation to be conducted by other researchers.

Keywords: Clustering; MIP; Tabu search, metaheuristics.

1. Introduction

Data mining, due to its rich variety of important and challenging problems, is proving to be a fruitful research arena for the optimization community. Early in the history of data mining, the dominant methods employed were adopted from standard statistical methodologies. As the optimization community started working on various problems related to data mining, the limitations of classical statistical approaches were increasingly exposed as formal optimization based approaches pointed the way to improved models and solution methods. While contributions have been made in most aspects of data mining, these advances are particularly pronounced in the areas of clustering, classification and feature selection.

In this paper, we highlight a new model for clustering. This model advances the state of the art in this application area and accordingly holds great promise for advancing the practice of data mining. In the following sections, we describe the new model and offer brief comparisons to alternative methods.

2. A New Model for Clustering

Clustering (unsupervised learning) remains one of the key application areas in data mining. Given a distance metric derived from an appropriate function on the items attributes (features), the objective is to group similar items together so that the "within group distances" are small and the "between group distances" are large. Applications of clustering are found in many disciplines ranging from market segmentation studies intended to segregate similar customers to MircoArray data analysis intended to better understand how different genes co-express themselves in response to different treatments.

A natural perspective on clustering is to adopt a graph theoretical point of view where data points are taken as nodes in a graph connected by edges with edge weights denoting the similarity of each pair of nodes in the network. The clustering problem, then, is to partition the graph into cliques with similar characteristics. This clique partitioning problem has long been recognized as a fruitful conceptual approach to adopt for clustering (see, e.g. Mirkin,[1] Grotschel and Wakabayashi,[2] Mehrotra and Trick[3] and Dorndorf and Pesch).[4]

Despite its appealing conceptual qualities, the clique partitioning model has not been widely employed in practice due to the difficulty of solving the standard 0/1 programming model for clique partitioning (CP). Where the model has been applied to small clustering problems, the results have been very effective. For medium to large clustering problems, however, the standard optimization model for CP explodes in size making it difficult if not impossible to solve by standard methods. This computational difficulty has served to preclude the broader use of the clique partitioning model as a tool for clustering.

The alternative model we present here for clique partitioning removes the size and computational obstacles. Consequently, our approach opens the door for the widespread adoption of clique partitioning as a modeling and solution methodology for the general clustering problem. This model has been shown by Wang *et al.*,[5] to have several advantages over the standard K-means procedure that is often used in practice.

2.1. *The CP problem*

Consider a graph $G = (V, E)$ with n vertices and unrestricted edge weights, where V is the vertex set and E is the edge set. The CP problem consists of partitioning the graph into cliques such that the sum of the edges weights over all cliques formed is as large as possible.

Standard 0/1 Linear Model: The standard optimization model for CP (see for instance Oosten *et al.*[6] is given by:

$$\text{CP(edge)}: \ \max \ x_0 = \sum_{(i,j)\in E} w_{ij}x_{ij} \tag{1}$$

$$\text{st} \quad \begin{aligned} & x_{ij} + x_{ir} - x_{jr} \le 1 \quad \forall \text{ distinct } i, j, r \in V. \\ & x_{ij} \in \{0,1\} \end{aligned} \tag{2}$$

The w_{ij} coefficients are unrestricted edge weights and x_{ij} is defined to be 1 if edge (i, j) is in the partition, and equal to 0 otherwise. Note that this is an *edge-based* formulation and even for modest sized graphs, this model explodes in size, having $n(n-1)/2$ variables and $3C_3^n$ constraints. Despite these size characteristics, the dominant methods presented in the literature for solving CP(edge) are exact approaches based on LP methods as illustrated by the cutting plane approaches of Grotschel and Wakabayashi[2] and Oosten *et al.*,[6] and the column generation approach of Mehrotra and Trick.[3] These approaches have proven to be successful on small to moderate size problems. For larger instances, however, the application of these approaches is severely limited due the challenge presented by the large size of CP(edge). For such cases, Metaheuristic methods, coupled with a new formulation, prove to be very effective as illustrated later.

2.2. *New formulation*

The computational challenge posed by CP(edge) for large problems motivates the development of a new formulation that can be readily solved by basic metaheuristic methodologies. We first present the new model and then describe our solution approach.

As before, n is the number of nodes (vertices) and the w_{ij} are unrestricted edge weights. Without loss of generality we assume here that G is a complete graph. If necessary, artificial edges with negative (penalty) edge weights can be introduced as needed to produce a complete graph in those cases where G is not initially complete. In addition, define

$$K_{\max} = \text{maximum number of cliques allowed (an educated guess)}$$

and

$$x_{ik} = 1 \quad \text{if node } i \text{ is assigned to clique } k; 0 \text{ otherwise.}$$

Then our model is:

$$\text{CP (node):} \;\; \max x_0 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \sum_{k=1}^{K_{\max}} x_{ik} x_{jk} \tag{3}$$

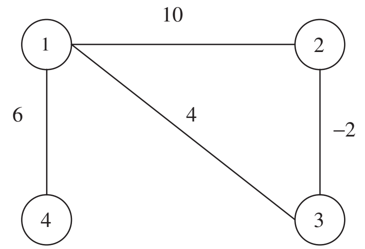
st

$$\sum_{k=1}^{K_{\max}} x_{ik} = 1, \quad i = 1, \ldots, n. \tag{4}$$

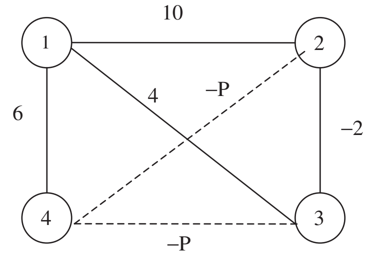
Note that the quadratic terms in the objective function imply that the weight w_{ij} becomes part of the partition weight only when nodes i and j are assigned to the same clique. The constraints of Eq. (4) require that each node is assigned to one of the cliques.

Several remarks about this model are in order: First of all, this is a node-oriented model with many fewer variables than CP(edge), since $n(K_{\max})$ is typically much less than $n(n-1)/2$. Furthermore, the number of constraints (n) is much smaller than the corresponding number ($3C_3^n$) for the edge-oriented model of CP(edge). While CP(edge) is a linear model and CP(node) is quadratic, the size difference enables this quadratic alternative to be used for large instances of clique partitioning problems where the computational burden of CP(edge) precludes its practical use. CP(node) can be effectively solved, even for large instances, by modern meta-heuristic methods such as Tabu Search (see, Glover *et al.*,[7] for details).

Example: We illustrate our approach to clique partitioning via CP(node) by the following example: Consider the graph



Adding artificial edges (2, 4) and (3, 4) with edge weights equal to $-P$, we get the complete graph



Taking (arbitrarily) $K_{\max}$ to be 3, CP(node) becomes:

$$\begin{aligned} \max\ x_0 = {} & 10x_{11}x_{21} + 4x_{11}x_{31} + 6x_{11}x_{41} - 2x_{21}x_{31} - Px_{21}x_{41} - Px_{31}x_{41} \\ & + 10x_{12}x_{22} + 4x_{12}x_{32} + 6x_{12}x_{42} - 2x_{22}x_{32} - Px_{32}x_{42} - Px_{32}x_{42} \\ & + 10x_{13}x_{23} + 4x_{13}x_{33} + 6x_{13}x_{43} - 2x_{23}x_{33} - Px_{23}x_{43} - Px_{33}x_{43} \end{aligned}$$

st

$$\begin{aligned} x_{11} + x_{12} + x_{13} &= 1 \\ x_{21} + x_{22} + x_{23} &= 1 \\ x_{31} + x_{32} + x_{33} &= 1 \\ x_{41} + x_{42} + x_{43} &= 1 \end{aligned}$$

Taking the scalar penalty P to be sufficiently large (like $P = 20$), this model is easily solved to give the non-zero assignments $x_{11} = x_{21} = x_{31} = x_{42} = 1$. Thus the optimal solution to CP(node) consists of two cliques with an objective function value of $x_0 = 12$ with nodes 1, 2 and 3 assigned to one clique and node 4 assigned to a second clique. Note that while we allowed for the possibility of three cliques, only two were used in our solution implying that our initial choice of $K_{\max} = 3$ was more than sufficient.

The computational superiority of CP(node) over CP(edge) is well documented in the literature via substantial computational experimentation as shown in Glover *et al.*[8] This computational advantage, coupled with the natural attractiveness of cliques, makes CP(node) an attractive model to adopt for clustering.

3. Summary and Conclusions

In this paper, we presented a new model for the clique partitioning problem that has significant application potential in the area of clustering. Our model eliminates the computational restrictions associated with the more standard "linear" model for clique partitioning and in doing so facilitates the widespread adoption of clique partitioning as a useful model for clustering. Computational comparisons with standard methods for clustering (such as K-means) indicate that our model CP(node) produces superior clustering results as evaluated by silhouette coefficients and other objective measures of cluster quality.

Acknowledgments

We would like to acknowledge the contributions of our co-workers Bahram Alidaee, Haibo Wang, and Marco Better who have contributed considerably to the ideas given in this paper.

References

1. B. Mirkin, Combinatorial optimization in clustering, in *Handbook of Combinatorial Optimization*, eds. D. Z. Du and P. M. Pardalos (Kluwer, 1998).
2. M. Grotschel and Y. Wakabayashi, A cutting plane algorithm for a clustering problem, *Mathematical Programming* **45** (1989) 59–96.
3. A. Mehrotra and M. Trick, Clique and clustering: A combinatorial approach, *Operations Research Letters* **22** (1998) 1–12.
4. U. Dorndorf and E. Pesch, Fast clustering algorithms, *Journal on Computing* **6**(2) (1994) 141–153.
5. H. Wang, T. Obremski, G. Kochenberger and B. Alidaee, A new clique partitioning model: An alternative to K-means clustering, Working paper, University of Colorado, USA (2006).
6. M. Oosten, J. Rutten and F. Spieksma, The clique partitioning problem: Facets and patching facets, *Networks* **38**(4) (2001) 209–226.
7. F. Glover, G. Kochenberger, B. Alidaee and M. Amini, 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*, eds. S. M. S. Voss, I. Osman and C. Roucairol (Kluwer, 1999), pp. 93–109.
8. F. Glover, G. Kochenberger, B. Alidaee and H. Wang, Clustering microarray data via clique partitioning, *Journal of Combinatorial Optimization* **10** (2005) 77–92.