Analyzing and Modeling the Maximum Diversity Problem by Zero-One Programming*

Ching-Chung Kuo
School of Business Administration, Pennsylvania State University at Harrisburg, Middletown, PA 17057

Fred Glover
Graduate School of Business Administration, University of Colorado at Boulder, Boulder, CO 80309

Krishna S. Dhir
School of Business Administration, Pennsylvania State University at Harrisburg, Middletown, PA 17057

ABSTRACT

The problem of maximizing diversity deals with selecting a set of elements from some larger collection such that the selected elements exhibit the greatest variety of characteristics. A new model is proposed in which the concept of diversity is quantifiable and measurable. A quadratic zero-one model is formulated for diversity maximization. Based upon the formulation, it is shown that the maximum diversity problem is NP-hard. Two equivalent linear integer programs are then presented that offer progressively greater computational efficiency. Another formulation is also introduced which involves a different diversity objective. An example is given to illustrate how additional considerations can be incorporated into the maximum diversity model.

Subject Areas: Discrete Programming, Linear Programming, and Mathematical Programming.

INTRODUCTION

Consider a set of elements (e.g., a group of residents in a small town) and some of their attributes (e.g., gender, age, and religion). For every element in the set, each of its attributes can be in one of several possible states (e.g., male or female; young, middle-aged, or old; Christian, Catholic, Mormon, or Buddhist). The maximum diversity problem is to select a predetermined or bounded number of elements from the set encompassing the greatest variety of attributes states.

The problem of maximizing diversity arises in a wide range of real-world settings. For instance, a pollster desires to survey a representative sample of individuals possessing a wide spectrum of characteristics. Currently, many colleges and universities, in formulating their admissions policies, go beyond test score and class rank and also consider other factors in search of a diverse student body [3] [38]. In the recent U.S. immigration reform, Congress was concerned about promoting the ethnic diversity among the immigrants [33]. In market planning it is frequently

*This research was supported in part by the Joint Air Force Office of Scientific Research and Office of Naval Research Contract No. F49620-90-C-0033 at the University of Colorado.

Decision Sciences
Volume 24 Number 6
Printed in the U.S.A.

desirable to maximize the number and diversity of the strengths in a brand's profile [26]. Other contexts to which the maximum diversity problem may be applicable include plant breeding [37], social problems [40], ecological preservation [35] [44], product design [2] [46], workforce management [43], curriculum design [1] [24], and genetic resource management [14]. Diversity maximization is also an important issue in the following areas: accounting and auditing, chemical experimentation, experimental design, medical research, geological exploration, portfolio selection, and structural engineering [13] [19].

Despite its fundamental importance, the maximum diversity problem is largely unexplored in the literature; research on maximizing diversity and related subjects has been sparse. This paper aims to provide a new model wherein the concept of diversity can be quantified and measured. Several integer programming formulations will be proposed for determining the most diverse set of elements from some larger collection. Additionally, an example will be presented to illustrate the application of the maximum diversity model.

EXISTING WORK

In recent years, diversity analysis has been utilized in empirical studies in economics as well as in other major functional areas of business, that is, accounting, finance, marketing, and production [25] [34] [45]. Typically, an entropy-based index is computed and used to compare the degrees of diversification of several systems with respect to a variable of interest (e.g., income distribution), or to examine the change in the degree of diversification of a system over a period of time. In most cases, the analysis is descriptive in nature. No comprehensive methodology or unifying model has been described that includes diversity maximization as a general objective.

In a sense, statistical clustering [21] offers a limited approach to maximizing diversity since, after identifying a set of clusters, a form of diversity is created by selecting elements from different clusters. However, this approach depends heavily on both the criteria for establishing the desirability of cluster membership and the methodology for taking account of such criteria. It also depends on the rules for selecting elements from various clusters and requires differentiating within cluster diversity from across cluster diversity. To date, no rigorously defined criterion or methodology addressing the issue of maximum diversity has been proposed.

Perhaps the first paper to characterize diversity maximization as a specific goal is that of Glover, Hersh, and McMillan [16]. In this study, a framework for measuring the diversity of a group of items is identified and integrated with a heuristic marginal value approach to find an approximate solution to the maximum diversity problem. Although the approach reportedly outperforms a clustering-based technique in a class of plant breeding programs [39], it suffers from some limitations. For instance, the solution obtained is not necessarily optimal. Further, no flexibility is built in to accommodate the needs of imposing bounds on the number of occurrences of specified attribute states.

More recently, Glover [13] suggested a network-related formulation (netform) model of the maximum diversity problem addressed in [16]. The netform is a generalized network which introduces a visual component that provides fuller insight into the structure of the maximum diversity problem. The netform is also susceptible to being exploited by tabu search methodology [11] [22].

In what follows, we demonstrate how the concept of diversity can be defined and measured explicitly in a new model by formulating a zero-one integer program.

MATHEMATICAL FORMULATION

We will begin with a rudimentary formulation of the maximum diversity problem that will subsequently be refined and elaborated. Consider a set of elements and let:

n = total number of elements in the set,
m = number of elements to be selected,
r = number of relevant attributes each element possesses, and
s_{ik} = state or value of attribute k of element i.

Although the definition of s_{ik} as a state indicator suggests that it takes on integer values, in many practical applications it is useful to allow attribute states to take on real values, that is, $sik \in R$, and to normalize them. In order for the maximum diversity problem to be meaningful, we assume that $n \geq m \geq 2$. Given a subset of two elements i and j with respective vectors of attribute states $(s_{i1}, s_{i2}, \dots, s_{ir})$ and $(s_{j1}, s_{j2}, \dots, s_{jr})$, the diversity of the subset may be defined as a normed distance between i and j. In location theory and numerical taxonomy, several distance measures have been used [6]. Depending on the areas of application, one measure may be preferred to others [28]. For example, by a Euclidean distance measure we have,

$$d_{ij} = [\sum_{k=1}^{r} (s_{ik} - s_{jk})^2]^{1/2}. \tag{1}$$

From this, we may measure the diversity of a selected set of elements as the sum of the Euclidean distances between each distinct pair of elements, that is,

$$\text{Diversity} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{ij} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [\sum_{k=1}^{r} (s_{ik} - s_{jk})^2]^{1/2}.$$

Let $x_i=1$ if element i is selected and 0 otherwise, $i=1, 2, \dots, n$. The maximum diversity problem can then be formulated as the following quadratic zero-one integer program.

Formulation (F1)

$$\text{Maximize } Z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{ij} x_i x_j,$$

subject to

$$\sum_{i=1}^{n} x_i = m,$$

$$x_i = 0 \text{ or } 1, \ 1 \leq i \leq n.$$

Some observations about (F1) are in order. Clearly, the problem is trivial if $n \geq m=2$. One simply chooses elements p and q from the set with $p<q$ such that $d_{pq} \geq d_{ij}$, $i=1, 2, \ldots, n-1$; $j=i+1, i+2, \ldots, n$. Nevertheless, the general maximum diversity problem is intractable [8] in that it is unlikely to find any algorithm which guarantees that an optimal solution can be obtained (and verified to be optimal) within a reasonable amount of computer time. The following result characterizes the nature of the maximum diversity problem (see Appendix for the proof of the theorem).

Theorem. Problem (F1) is NP-hard, both with and without restricting the d_{ij} coefficients to non-negative values.

We note that (F1) can easily be modified to compel the number of elements selected to lie between upper and lower bounds. If all $d_{ij} \geq 0$, such a variant is irrelevant here since an optimal solution will always result by selecting $\Sigma_{i=1}^{n} x_i$ to equal its upper bound. However, we will subsequently note that the potential value of introducing additional constraints and, in this case, the use of lower and upper bounds on the range of $\Sigma_{i=1}^{n} x_i$ can be appropriate.

EQUIVALENT LINEAR INTEGER PROGRAMMING MODELS

Accepting that the best possible methods may have to settle for finding solutions with no assurance of optimality, we note that the nonlinearity of (F1) is inconvenient for most existing integer programming approaches. Although several quadratic algorithms have been devised [31] [32] [41] [42], they have not undergone the intensive refinements of linear zero-one methods, nor have they found widespread use in real-world applications.

Methods for converting zero-one polynomial models into equivalent zero-one linear models are well known in the literature. For example, using the approach of Glover and Woolsey [20], (F1) can be transformed into the following linear mixed integer program.

Formulation (F2)

$$\text{Maximize } Z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{ij} y_{ij},$$

subject to

$$\sum_{i=1}^{n} x_i = m$$

$$x_i + x_j - y_{ij} \leq 1, \; 1 \leq i < j \leq n, \tag{2}$$

$$-x_i + y_{ij} \leq 0, \; 1 \leq i < j \leq n, \tag{3}$$

$$-x_j + y_{ij} \leq 0, \; 1 \leq i < j \leq n, \tag{4}$$

$$y_{ij} \geq 0,\ 1 \leq i < j \leq n, \tag{5}$$

$$x_i = 0 \text{ or } 1,\ 1 \leq i \leq n. \tag{6}$$

Observe that in (F2) $y_{ij} \leq 1$ will result, $i=1, 2, \ldots, n-1$; $j=i+1, i+2, \ldots, n$, and the y_{ij} variables will automatically receive zero-one values whenever the x_i variables are assigned such values. Moreover, we note that constraints (2) through (5) also validly model y_{ij} as the product of x_i and x_j in the case where one of x_i and x_j is continuous, relaxing the integer requirement on this variable in (6).

We can improve further on (F2) using the results of Glover [9]. Let L_i and U_i be, respectively, lower and upper bounds on the quantity $\Sigma_{j=i+1}^{n} d_{ij}x_j$; for example, $L_i = \Sigma_{j=i+1}^{n} \min(0, d_{ij})$ and $U_i = \Sigma_{j=i+1}^{n} \max(0, d_{ij})$. Then the method of [9] yields the following formulation equivalent to (F1).

Formulation (F3)

$$\text{Maximize } Z = \sum_{i=1}^{n-1} w_i$$

subject to

$$\sum_{i=1}^{n} x_i = m,$$

$$-U_i x_i + w_i \leq 0,\ 1 \leq i \leq n-1,$$

$$-\sum_{j=i+1}^{n} d_{ij}x_j + L_i(1 - x_i) + w_i \leq 0,\ 1 \leq i \leq n-1,$$

$$x_i = 0 \text{ or } 1,\ 1 \leq i \leq n.$$

Compared to the $n(n-1)/2$ new variables and $3n(n-1)/2$ new inequalities introduced in (F2), there are only $n-1$ new variables and $2(n-1)$ new inequalities in (F3). Consequently, the continuous linear programming relaxation of (F3) can be solved more efficiently than (F2). It should be pointed out, however, that (F3) does not yield an LP formulation as restrictive as (F2).

For the purpose of computer implementation, the preceding formulation can be improved using the observation of [10] by introducing a nonnegative slack variable u_i in the first w_i inequality to give $w_i = U_i x_i - u_i$. Then the inequality $-U_i x_i + w_i \leq 0$ can be replaced by $u_i \geq 0$ and w_i can be replaced throughout the remainder of the formulation by $U_i x_i - u_i$. The effect is to yield $(n-1)$ inequalities together with the nonnegativity condition for u_i. (Corresponding observations apply to introducing a slack variable for the other inequality involving w_i.) The formulation by Kettani and Oral [27], which claims to improve the results of [9], in fact requires twice as many inequalities as required in our formulation.

The following gives an additional diversity model based on a goal of maximizing the minimum separation among the elements selected.

MAXIMIN DIVERSITY MODEL

Formulation (F4)

$$\text{Maximize } Z = w,$$

subject to

$$\sum_{i=1}^{n} x_i = m,$$

$$(M - d_{ij})y_{ij} + w \le M,\ 1 \le i < j \le n,$$

$$x_i + x_j - y_{ij} \le 1,\ 1 \le i < j \le n,$$

$$-x_i + y_{ij} \le 0,\ 1 \le i < j \le n,$$

$$-x_j + y_{ij} \le 0,\ 1 \le i < j \le n,$$

$$y_{ij} \ge 0,\ 1 \le i < j \le n,$$

$$x_i = 0 \text{ or } 1,\ 1 \le i \le n.$$

While w is unrestricted in sign in the above formulation, the constraint $w \ge 0$ may be included to facilitate the solution of (F4) if $d_{ij} \ge 0$, $i=1, 2, \ldots, n-1$; $j=i+1, i+2, \ldots, n$, since this is a maximization problem. Furthermore, the objective of (F4) can incorporate tie breaking by seeking to maximize $Mw + \Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} d_{ij} y_{ij}$, where, as before, M is an extremely large positive number. This hybrid model incorporating both "maximin" and "maxisum" goals has proven to be useful in linear programming approaches to discriminant analysis [12] [17]. We anticipate that such an objective will also be important for a variety of other applications. However, it creates a more difficult optimization problem to wrestle with. As a result, specialized methods such as those discussed in [15] and [18] are desirable.

We note that all of the preceding formulations can be accompanied by additional constraints, such as resource limitations restricting combinations of elements that can be selected together, and input or output requirements compelling certain subclasses of elements to be represented in the set constructed. This type of elaboration can be used to assure that the diversified collection will include elements from clusters generated in advance. Thus, our formulations permit cluster-based approaches to be embedded within a broader framework.

For the sake of concreteness, an example in Plane and McMillan [36] is adapted and illustrated. We use the simplest (most direct) form of this model, (F2), and then show how additional relevant constraints can be introduced to address particular concerns.

AN ILLUSTRATIVE EXAMPLE

Suppose the president of a state university has been given a list of ten names by the governor of the state. From this list, the president has been asked to select five people to serve as the university's governing regents for the next year. The information about the ten nominees has been summarized in Table 1 with respect to six attributes: gender, race, geographical region, education, occupation, and political party affiliation. The goal is to have a diverse board of regents with the attributes states exhibiting the greatest variety.

In order to subject the problem to the application of the maximum diversity model, we need to compute the between-nominee distances. For each of the six attributes, we assign a positive integer to each of the possible states representing the amount, degree, or category associated with the state with respect to the attribute. This is accomplished by utilizing the coding scheme given in Table 2, and the results are displayed in Table 3. The distance between each distinct pair of nominees is then computed using the formula in (1). For example, the Euclidean distance between Baum and Inman is $d_{29}=[(1-1)^2+(1-1)^2+(1-1)^2+(4-3)^2+(3-3)^2+(2-1)^2]^{1/2}\approx 1.4142$. The symmetric distance matrix is shown in Table 4.

With all the between-nominee distances on hand, the problem of determining the most diverse governing board of regents can be formulated as the following mixed linear integer program based upon (F2).

$$\text{Maximize } Z = 2.6458y_{12} + 2.8284y_{13} + \dots + 2.8284y_{9,10},$$

subject to

$$x_1 + x_2 + \dots + x_{10} = 5,$$

$$x_1 + x_2 - y_{12} \le 1,$$

$$\vdots$$

$$x_9 + x_{10} - y_{9,10} \le 1,$$

$$-x_1 + y_{12} \le 0,$$

$$\vdots$$

$$-x_9 + y_{9,10} \le 0,$$

$$-x_2 + y_{12} \le 0,$$

$$\vdots$$

$$-x_{10} + y_{9,10} \le 0,$$

$$y_{12}, y_{13}, \dots, y_{9,10} \ge 0,$$

$$x_1, x_2, \dots, x_{10} = 0 \text{ or } 1.$$

Table 1: Information about nominees and their relevant attributes.

Number	Name	Gender	Race	Region	Education	Occupation	Political
1	Adams	Female	White	West	Some college	Middle	Democratic
2	Baum	Male	White	East	College	Middle	Nonpartisan
3	Cain	Male	Black	West	High school	Lower middle	Republican
4	Dunn	Female	Black	Central	Elementary school	Lower middle	Democratic
5	Evans	Female	White	Central	College	Upper	Nonpartisan
6	Frey	Male	Black	East	College	Middle	Nonpartisan
7	Gill	Male	White	West	Some college	Upper	Republican
8	Huss	Male	White	Central	High school	Lower	Democratic
9	Inman	Male	White	East	Some college	Middle	Republican
10	Jones	Female	Black	East	College	Upper	Democratic

Table 2: Coding scheme.

	State and Index			
Attribute	1	2	3	4
Gender	Male	Female		
Race	White	Black		
Region	East	Central	West	
Education	Elementary school	High school	Some college	College
Occupation	Lower	Lower middle	Middle	Upper
Political	Republican	Nonpartisan	Democratic	

Table 3: Coded data.

Number	Name	Gender	Race	Region	Education	Occupation	Political
1	Adams	2	1	3	3	3	3
2	Baum	1	1	1	4	3	2
3	Cain	1	2	3	2	2	1
4	Dunn	2	2	2	1	2	3
5	Evans	2	1	2	4	4	2
6	Frey	1	2	1	4	3	2
7	Gill	1	1	3	3	4	1
8	Huss	1	1	2	2	1	3
9	Inman	1	1	1	3	3	1
10	Jones	2	2	1	4	4	3

Table 4: Distance matrix.

	1	2	3	4	5	6	7	8	9	10
1	.0000	2.6458	2.8284	2.6458	2.0000	2.8284	2.4495	2.6458	3.0000	2.6458
2	2.6458	.0000	3.3166	3.7417	1.7321	1.0000	2.6458	3.1623	1.4142	2.0000
3	2.8284	3.3166	.0000	2.6458	3.4641	3.1623	2.4495	2.6458	2.6458	4.1231
4	2.6458	3.7417	2.6458	.0000	3.8730	3.6056	3.8730	2.0000	3.4641	3.7417
5	2.0000	1.7321	3.4641	3.8730	.0000	2.0000	2.0000	3.8730	2.2361	1.7321
6	2.8284	1.0000	3.1623	3.6056	2.0000	.0000	2.8284	3.3166	1.7321	1.7321
7	2.4495	2.6458	2.4495	3.8730	2.0000	2.8284	.0000	3.8730	2.2361	3.3166
8	2.6458	3.1623	2.6458	2.0000	3.8730	3.3166	3.8730	.0000	3.1623	4.0000
9	3.0000	1.4142	2.6458	3.4641	2.2361	1.7321	2.2361	3.1623	.0000	2.8284
10	2.6458	2.0000	4.1231	3.7417	1.7321	1.7321	3.3166	4.0000	2.8284	.0000

Because this example problem, which is called (P), is quite small, we may use a standard zero-one optimization procedure to solve it with a good chance of obtaining a verified optimal solution. We have done this to obtain an optimal solution given by $(x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*, x_7^*, x_8^*, x_9^*, x_{10}^*)=(0, 0, 1, 1, 0, 0, 1, 1, 0, 1)$ with a total diversity of $Z^*=32.6685$. In other words, to maximize the diversity of the governing board, the president should select Cain, Dunn, Gill, Huss, and Jones as the university's regents. If it is desirable to choose at least one middle-class person, the constraint $x_1+x_2+x_6+x_9 \geq 1$ may be added to (P). The solution to this problem shows that the following five people will serve as the governing regents for the next year with a total diversity of $Z^*=32.4952$: Dunn, Gill, Huss, Inman, and Jones. If, alternatively, it is required that no more than two regents be Democrats or Republicans, the constraints $x_1+x_4+x_8+x_{10} \leq 2$ and $x_3+x_7+x_9 \leq 2$ may be included in (P). The solution in this case shows that the board will consist of Baum, Dunn, Gill, Huss, and Jones and the total diversity becomes $Z^*=32.3541$. Other modeling considerations can be accommodated in a similar way.

ATTRIBUTE CHARACTERISTICS

We have developed several integer programming formulations for the maximum diversity problem, and have also provided an example to illustrate the application of the model. However, an issue pertaining to attribute characteristics deserves further elaboration.

We have assumed in the above illustration that all of the six attributes are of equal importance to the president of the state university in the decision-making process. In case the president is more concerned about gender than the other five attributes, the state indexes for male and female in the coding scheme can be changed from 1 and 2 to 1 and 3 or 1 and 5, respectively. This will magnify the distance between any two nominees along the dimension of gender; this will in turn make the attribute of gender a more influential factor in determining the most diverse governing board of regents for the next year.

In the previous example, although it seems trivial to quantify the various states of each attribute, there are many attributes in real life that are nonmetric in nature. These include, among others, religion, ethnicity, hobby, color, and shape. In order to deal with diversity problems involving these kinds of qualitative characteristics, a special approach needs to be taken to quantify the states. Taking the previous illustration as an example, it was implicitly assumed that the attribute of political party affiliation can be measured on a scale of 1 to 3 depending on a person's attitude toward social change and/or reform. Suppose the president agrees that the Democrats are traditionally more liberal, the Republicans tend to be more conservative, and the nonpartisans are generally considered to be moderate. As a result, in the coding scheme, the numbers 1, 2, and 3 are assigned to the states Republican, nonpartisan, and Democratic, respectively, according to the degree of conservativeness. The same line of argument can be applied to other nonmetric traits. However, it is important that the decision maker's judgment and preference be properly reflected in the coding scheme.

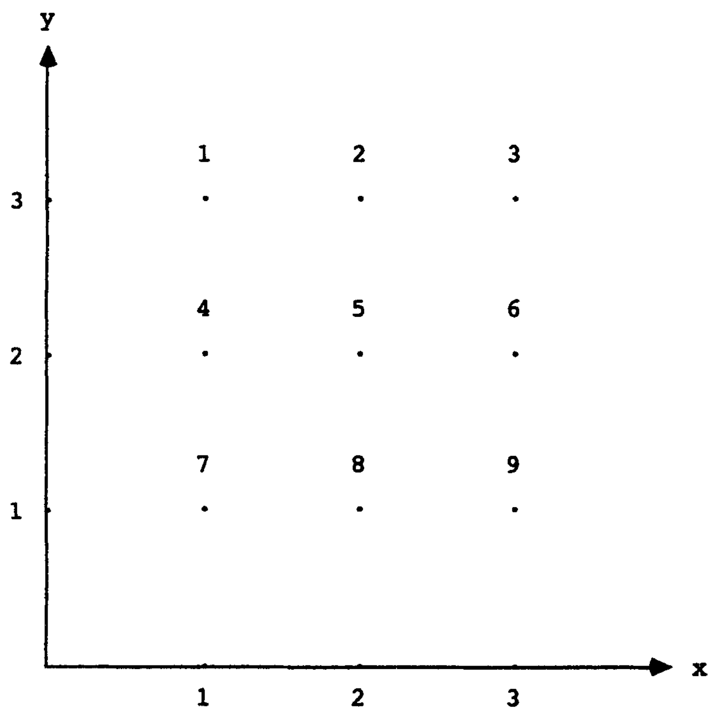
CONCLUSION

In this exploratory study of the maximum diversity problem, a new approach to measuring diversity is proposed. A quadratic zero-one model is formulated to maximize diversity, which is proven to be NP-hard. In order to solve the problem more effectively, we transform the nonlinear integer program into two equivalent formulations of progressively greater efficiency. Moreover, another formulation based upon a different diversity objective is presented. We conclude with an example illustrating the application of the maximum diversity model.

We have demonstrated the trade-offs in efficiency versus the ability to encompass more complex goals in maximizing diversity. Since many fewer variables are used in (F3) than in (F2), this is a significant improvement with respect to computational efficiency. However, (F2) underlies the more complex objective of (F4). Moreover, each of the maximum diversity formulations is very flexible in allowing a decision maker to incorporate a wide range of additional considerations, as illustrated by the supplementary constraints introduced in the example discussed previously. Also, the coding scheme can be easily modified to reflect the decision maker's preference as well as value judgment.

The treatment of the maximum diversity problem described in this paper suggests an approach to another class of problems in which the objective is to minimize diversity. For example, we may seek minimum diversity when compliance to specifications is of paramount importance, such as in precision industries. In experimental design, the degree of homogeneity within a population or between experimental and control populations may be enhanced through the minimization of diversity. Another potential application of the concept of minimum diversity is in group technology [7], where parts are classified into families with similar characteristics and processed in various cells to increase production efficiency. We note that most of the results for diversity maximization presented in this work may be modified and applied to diversity minimization. This problem is discussed further in [30].

Figure 1: A graphical representation of a maximum diversity problem.



In survey design, it is often desirable to draw a representative sample to ensure more reliable results [5]. The concept of representivity is particularly important in statistical stratified sampling [4] [29]. Although diversity plausibly has the connotation of representivity, diversification does not necessarily imply representation, and vice versa. Generally, maximum diversity and maximum representivity may not be achieved simultaneously, as evidenced by the decrease in total diversity in the illustrative example when the additional consideration of representivity by middle-class people was to be accommodated. We conjecture that the issue of representivity can be better addressed by (F4) due to the criterion of "maximin" distance used in the model. To see this, we treat an element possessing r attributes as a point in the r-dimensional space and consider the maximum diversity problem in Figure 1, where $n=9$, $m=5$, $r=2$, $s_{11}=1$, $s_{12}=3$, $s_{21}=2$, $s_{22}=3$, etc. Based on (F1) (and hence (F2) and (F3)), the maximally diverse subset of points can be any of the following: {1, 2, 3, 7, 9}, {1, 3, 4, 7, 9}, {1, 3, 6, 7, 9}, and {1, 3, 7, 8, 9}. However, based on (F4), the subset in which the minimum between-element diversity is maximized is {1, 3, 5, 7, 9}. In this instance, the maximin subset appears to be more representative than the maxisum subset.

This type of representivity can be achieved at a more refined level by establishing a hierarchy of objectives, as proposed in [18], to progressively maximize the second smallest distance, then the third smallest distance, and so on, in a strictly preemptive order. Such a hierarchical approach can be created out of combinations of the maxisum and maximin objectives, and the consequences of such variants for creating different forms of diversity (that embody different aspects of representivity) provide an inviting realm for future study. [Received: September 28, 1992. Accepted: August 18, 1993.]

REFERENCES

[1] AACSB. *Final report: The AACSB accreditation project.* St. Louis, MO: The American Assembly of Collegiate Schools of Business, 1991.

[2] Besanko, D., Perry, M. K., & Spady, R. H. The logit model of monopolistic competition: Brand diversity. *Journal of Industrial Economics*, 1990, *38*(4), 397-415.

[3] Bunzel, J. H., & Au, J. K. D. Diversity or discrimination? Asian Americans in college. *Public Interest*, 1987, *87*, 49-62.

[4] Chaudhuri, A., & Vos, J. W. E. *Unified theory and strategies of survey sampling.* Amsterdam, the Netherlands: North-Holland, 1988.

[5] Dent, T. How to design for a more reliable customer sample. *Business Marketing Digest*, 1992, *17*(2), 73-76.

[6] Francis, R. L., McGinnis, L. F., & White, J. A. *Facility layout and location: An analytical approach* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, 1992.

[7] Gallagher, C. C., & Knight, W. A. *Group technology production methods in manufacture.* Chichester, England: Ellis Howard, 1986.

[8] Garey, M. R., & Johnson, D. S. *Computers and intractability: A guide to the theory of NP-completeness.* New York: W. H. Freeman, 1979.

[9] Glover, F. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 1975, *22*(4), 455-460.

[10] Glover, F. An improved MIP formulation for products of discrete and continuous variables. *Journal of Information & Optimization Sciences*, 1984, *5*(1), 69-71.

[11] Glover, F. Future paths for integer programming and links to artificial intelligence. *Computers and Operations Research*, 1986, *13*(5), 533-549.

[12] Glover, F. Improved linear programming models for discriminant analysis. *Decision Sciences*, 1990, *21*(4), 771-785.

[13] Glover, F. Advanced netform models for the maximum diversity problem. Working paper, Graduate School of Business Administration, University of Colorado at Boulder, 1991.

[14] Glover, F. Formulation and illustrative use of the diversity problem in genetic resource management. Working paper, Graduate School of Business Administration, University of Colorado at Boulder, 1992.

[15] Glover, F. Tabu search for nonlinear and parametric optimization (with links to genetic algorithms). Forthcoming in *Discrete Applied Mathematics.*

[16] Glover, F., Hersh, G., & McMillan, C. Selecting subsets of maximum diversity. MS/IS Report No. 77-9, University of Colorado at Boulder, 1977.

[17] Glover, F., Keene, S., & Duea, B. A new class of models for the discriminant problem. *Decision Sciences*, 1988, *19*(2), 269-280.

[18] Glover, F., & Laguna, M. Tabu search. In C. R. Reeves (Ed.), *Modern heuristic techniques for combinatorial problems.* Cambridge, MA: Blackwell Scientific Publications, 1993.

[19] Glover, F., & McMillan, C. Optimizing diversity. Working paper, Graduate School of Business Administration, University of Colorado at Boulder, 1978.

[20] Glover, F., & Woolsey, R. E. Converting the 0-1 polynomial programming problem to a 0-1 linear program. *Operations Research*, 1974, *22*(1), 180-182.

[21] Hair, J. F., Anderson, R. E., & Tatham, R. L. *Multivariate data analysis with readings* (2nd ed.). New York: Macmillan, 1987.

[22] Hansen, P. The steepest ascent mildest descent heuristic for combinatorial programming. Presented at the Congress on Numerical Methods in Combinatorial Optimization, Capri, Italy, 1986.

[23] Harary, F. *Graph theory.* Reading, MA: Addison-Wesley, 1969.

[24] Jackson, J. L. Campuses debate multiculturalism. *Focus*, 1991, *19*, 4-5.

[25] Jones, L. E. Optimum product diversity and the incentives for entry in natural oligopolies. *Quarterly Journal of Economics*, 1987, *102*(3), 595-613.

[26] Keely, A. From experience—Maxi-niching the way to a strong brand: Positioning according to systemic dynamics. *Journal of Product Innovation Management*, 1989, *6*(3), 202-206.
[27] Kettani, O., & Oral, M. Equivalent formulations of nonlinear integer problems for efficient optimization. *Management Science*, 1990, *36*(1), 115-119.
[28] Klastorian, T. D. The *p*-median problem for cluster analysis: A comparative test using the mixture model approach. *Management Science*, 1985, *31*(1), 84-95.
[29] Konijn, H. S. Statistical theory of sample survey design and analysis. Amsterdam, the Netherlands: North-Holland, 1973.
[30] Kuo, C.-C., Glover, F., & Dhir, K. S. A discrete optimization model for the minimum diversity problem. Working paper, School of Business Administration, Pennsylvania State University at Harrisburg, 1993.
[31] Laughhunn, D. J. Quadratic binary programming with application to capital-budgeting problems. *Operations Research*, 1970, *18*(3), 454-461.
[32] Lawler, E. L., & Bell, M. D. A method for solving discrete optimization problems. *Operations Research*, 1966, *14*(6), 1098-1112.
[33] McConnell, S. The new battle over immigration. *Fortune*, 1988, *117*(10), 89-102.
[34] Nayak, T. K., & Gastworth, J. L. The use of diversity analysis to assess the relative influence of factors affecting the income distribution. *Journal of Business & Economic Statistics*, 1989, *7*(4), 453-460.
[35] Pearce, D. Economics and genetic diversity. *Futures*, 1987, *19*(6), 710-712.
[36] Plane, D. R., & McMillan, C. *Discrete optimization: Integer programming and network analysis for management decisions*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[37] Porter, W. M., Rawal, K. M., Rachie, K. O., Wien, H. C., & Williams, R. C. *Cowpea germplasm catalog No. 1*. International Institute of Tropical Agriculture, Ibadan, Nigeria, 1975.
[38] Ramirez, A. Cream of the crop? *Wall Street Journal*, April 6, 1979, *1*, 29.
[39] Rawal, K. M., Kaltenhauser, J. C., & Snyder, M. J. A clustering technique for data analysis in plant genetic resources research. Presented at the AAAS Annual Meeting, Denver, Colorado, February 1977.
[40] Swierenga, R. P. Ethnicity in historical perspective. *Social Science*, 1977, *52*(1), 31-44.
[41] Taha, H. A. A Balasian-based algorithm for zero-one polynomial programming. *Management Science*, 1972, *18*(6), B328-B343.
[42] Taha, H. A. Further improvements in the polynomial zero-one algorithm. *Management Science*, 1972, *19*, B226-B227.
[43] Thomas, R. R. From affirmative action to affirming diversity. *Harvard Business Review*, 1990, *68*(2), 107-117.
[44] Unkel, W. C. Natural diversity and national forest planning. *Natural Areas Journal*, 1985, *5*, October, 8-13.
[45] Varadarajan, P., & Ramanujam, V. Diversification and performance: A reexamination using a new two-dimensional conceptualization of diversity in firms. *Academy of Management Journal*, 1987, *30*(2), 380-393.
[46] von Ghyczy, T. Product diversity and proliferation as a new mode of competing in the motor car. *International Journal of Technology Management*, 1986, *1*(3/4), 515-518.

APPENDIX

Proof of Theorem

We will prove the theorem by showing that the clique problem [23], which is NP-complete [8], is reducible to (F1). It is sufficient to consider the case where the d_{ij} coefficients are restricted to be non-negative since this is the most limiting situation. We begin with stating the clique problem.

> *Clique Problem*: Given a graph $G=(V, E)$ and a positive integer $K \leq |V|$, does G contain a clique of size K or more, that is, a subset V' of V with $|V'| \geq K$ such that every two vertices in V' are joined by an edge in E?

Given any instance of the clique problem, we construct the following instance of (F1) in which $n \geq m \geq 3$ and $d > 0$:

$$d_{ij} \in \{0, d\}, i = 1, 2, \ldots, n-1; j = i+1, i+2, \ldots, n$$

$$\{1, 2, \ldots, n\} = V,$$

$$\{(i,j) \in V \times V^{\wedge} | ^{\wedge} i < j \text{ and } d_{ij} = d\} = E,$$

$$m = K.$$

Obviously, $K = m \leq n = |V|$ or $K \leq |V|$. We will show that the clique problem has a solution if and only if there exists a feasible solution to (F1) with a total diversity of $dm(m-1)/2$.

Suppose there exists a feasible solution $(x_1, x_2, \ldots, x_n)$ to (F1) for which the total diversity is $d_m(m-1)/2$. Let $V' = \{i | x_i = 1, 1 \leq i \leq n\}$ and $E' = \{(i,j) \in V' \times V' | i < j\}$. Since $\Sigma_{i=1}^{n} x_i = m$ and $x_i \in \{0, 1\}$, i=1, 2, ..., n, we have $|V'| = m$ and $|E'| = m(m-1)/2$. Thus V' is a subset of V and $|V'| = m = K$ or $|V'| \geq K$. Moreover, $dm(m-1)/2 = \Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} d_{ij} x_i x_j = \Sigma_{(i,j) \in E'} d_{ij} x_i x_j = \Sigma_{(i,j) \in E'} d_{ij} \leq \Sigma_{(i,j) \in E'} d = d|E'| = dm(m-1)/2$ or $dm(m-1)/2 = \Sigma_{(i,j) \in E'} d_{ij} \leq dm(m-1)/2$. This implies that $\Sigma_{(i,j) \in E'} d_{ij} = dm(m-1)/2$ or, equivalently, $d_{ij} = d$ for all $(i,j) \in E'$. Hence, E' is a subset of E. As $G' = (V', E')$, the induced subgraph of V on V', is a complete graph, V' is a clique contained in V. Therefore, the clique problem has a solution.

Conversely, suppose the clique problem has a solution, that is, there exists a subset V' of V with $|V'| \geq K$ such that every two vertices in V' are joined by an edge in E. Let V'' be a subset of V' with $|V''| = K$ and $E'' = \{(i,j) \in E | i, j \in V''\}$. We see that $|E''| = |V''|(|V''| - 1)/2 = K(K-1)/2$ and $G'' = (V'', E'')$ is a complete subgraph of G. Consider $(x_1, x_2, \ldots, x_n)$ with $x_i = 1$ if $i \in V''$ and 0 otherwise. Note that $\Sigma_{i=1}^{n} x_i = \Sigma_{i \in V''} x_i = \Sigma_{i \in V''} 1 = |V''| = K = m$ or $\Sigma_{i=1}^{n} x_i = m$, which implies that $(x_1, x_2, \ldots, x_n)$ is a feasible solution to (F1). Furthermore, $\Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} d_{ij} x_i x_j = \Sigma_{(i,j) \in E''} d_{ij} x_i x_j = \Sigma_{(i,j) \in E''} d_{ij} = \Sigma_{(i,j) \in E''} d = d|E''| = dK(K-1)/2 = dm(m-1)/2$ or $\Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} d_{ij} x_i x_j = dm(m-1)/2$. Therefore, there exists a feasible solution to (F1) for which the total diversity is exactly $dm(m-1)/2$.

Given any instance of the clique problem, the instance of (F1) can be constructed in polynomial time. We then conclude that the problem of (F1) is NP-hard.

Ching-Chung Kuo is Assistant Professor of Operations Management at The Pennsylvania State University at Harrisburg. He received his Ph.D. in industrial engineering and management sciences from Northwestern University. His current research interests include mathematical modeling and applied operations research. Dr. Kuo's work has appeared in the *European Journal of Operational Research*.

Fred Glover is the US West Chaired Professor in Systems Science at the University of Colorado, Boulder. He has authored or co-authored more than two hundred and twenty published articles in the fields of mathematical optimization, computer science, and artificial intelligence with particular emphasis on practical applications in industry and government. In addition to holding editorial posts for journals in the U.S. and abroad, Dr. Glover has been featured as a National Visiting Lecturer by the Institute of Management Science and

the Operations Research Society of America, and has served as a host and lecturer in the U.S. National Academy of Sciences Program of Scientific Exchange.

Krishna S. Dhir is the S.B.A. Professor of Business Administration and the Director of the School of Business Administration at the Pennsylvania State University at Harrisburg. He received his Ph.D. in administrative policy and management science from the University of Colorado. His research focuses on policy analysis. He has published in *IEEE Transactions on Engineering Management*, *Journal of Marriage and the Family*, *Business and Society*, *International Journal of Public Administration*, and *Handbook of Behavioral Economics*.