TEMPORARY COMMERCIAL NETWORK PROGRAMMING FOR BEIJING OLYMPIC GAMES BASED ON DATA MINING AND GENETIC ALGORITHMS

Yan Chen[1], Jun Tian[1], Can Yang[2]
1 The School of Management, Xi'an Jiaotong University,
710049 Xi'an, P.R. China
2 School of Computer Science & Engineering,
South China University of Technology,
510640 Guangzhou, P.R. China

ABSTRACT

This paper is committed to the optimization of the temporary commercial network programming for the 2008 Beijing Olympics. It begins with the information extraction from the Assess database of questionnaire investigations based on data mining techniques including cluster analysis and mining of association rules. Based on the discovered knowledge, we calculate customer traffics and shopping demands of the twenty commercial sites and determine the reasonable construction scales of the 20 commercial sites. In order to reach the three objectives of the programming work—meeting shopping demand, balanced distribution, and making profits—we design an optimization model based on genetic algorithms to optimize the construction scale of each site. Finally, a linear programming model is proposed for choosing the appropriate type and corresponding number of supermarket for each of the 20 sites. All the methods in this paper are innovative tools for the quantitative analysis tasks of the decision-making jobs related to the temporary commercial network of the Beijing Olympic Games.

Keywords: commercial network programming, temporary supermarket, data mining, genetic algorithms.

INTRODUCTION

The 2008 Beijing Olympic Games are approaching. During the Olympic Games, a temporary commercial network, around the stadium facilities, should be constructed to cater to the needs of spectators, visitors, and staff. Its business scope will cover food, Olympic souvenirs, tourism supplies, sports supplies and daily necessities. The programming job of the temporary commercial should meet three principal requirements in terms of location, size and number. They are: first, satisfying the shopping demands during the games; second, distributing evenly among the twenty commercial sites; and third, making profit.

As the simplification of the real map, Figure 1 merely retains the regions and related parts in relevance with the research, including roads (white regions are sidewalks), bus stations, subway stations, taxi stations, private car parking lots, catering facilities, etc. The yellow regions within the graph marked A1-A10, B1-B6, and C1-C4 are the 20 commercial sites required for optimization (CUMCM, 2004a).

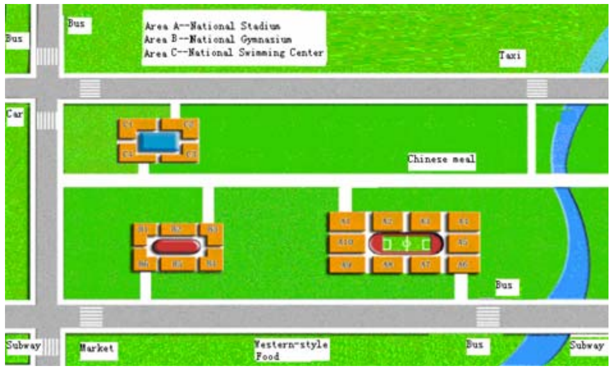

Figure 1. This shows a figure consisting of regions in relevance with the programming of the 20 commercial sites.

Assumed that the capacity of National Stadium A is 100,000; National Gymnasium B is 60,000; and National Swimming Center C is 40,000. Each auditorium of the three stadiums can accommodate 10,000 persons, and the exit of each auditorium directed at one of the 20 commercial sites. All the areas of the 20 commercial sites are the same (CUMCM, 2004a).

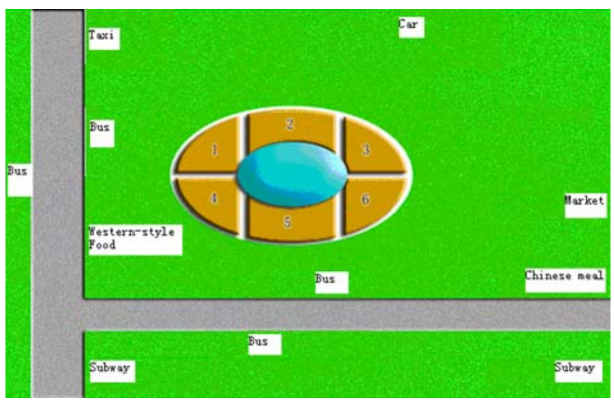

Figure 2. This figure shows the regions where the rehearsal sports meetings take place.

In order to obtain the preferences of the audiences regarding transportation, eating, and shopping, one alternative approach is to hold a rehearsal sports meetings and investigate the audiences during the games, which was done by CUMCM (2004b). Three games were held at the stadium in Figure 2, where CUMCM (2004b) had collected related data through questionnaire survey and recorded in the Access database.

RESEARCH DESIGN

In this section, we will explore the overall architecture of the research. We will divide the whole research into three smaller parts, including data models, shopping demands of commercial sites, and programming models. In this section, we will explore the research methods we will use in these three smaller parts. Now, let us begin with some assumptions we used in the research.

Assumptions

Assumption 1. During the Olympic Games, each audience travels by the shortest path twice a day, one get in and out of stadiums, the other is for meal; When there is more than one shortest path, they randomly choose one of them.

Assumption 2. Audiences go shopping twice a day by the way they travel; they randomly choose one of the commercial sites they pass through. Expenditures of the two shopping in a day are the same.

Assumption 3. The behaviours of the audiences do not fluctuate with the economic conditions. The value of the currency does not change with time. Hence, we will not distinct the behaviours of audiences between rehearsal sports games and the 2008 Olympic Games.

Assumption 4. During the Olympics, the overall amount of audiences in the three main stadiums will be 200,000 exactly. That is, each of the 20 auditoriums will hold 10,000 audiences exactly.

Data Models

As for analysis of the survey data, many researchers choose the statistical methods (Chen, 2004; Zhang, 2004; Li, 2004; Meng, 2004; Xue, 2004; Yi, 2006; Tang, 2005; Hu, 2005; Wang, 2005); while others may use data mining techniques (Ai, 2004). Meng (2004) holds the opinions that people should subjectively determine which attributes to be analyzed by the statistical method, hence, the results will be less objective and complete compared with results by data mining methods. Therefore, we choose data mining techniques, including cluster analysis and mining of association rules, to analyze the survey data.

At the beginning of the data mining, clustering will always be a best choice because it can partition a data set into subsets (clusters), so that the data in each subset (ideally) share some common traits. After that, we can analyze characteristics of each cluster; or we may focus on some specific cluster for intensive analysis (Xiang, 2003).

The mining of association rules is used to identify the potential interconnection relations in the database. If two or more items appear together repeatedly with high probability, there will be some connection between these items. Hence, we can establish association rules between these items (Tang, 2002).

By using these two techniques, we can completely obtain the behavioural laws of audiences in terms of travelling, eating and shopping during the rehearsal sports meeting. We assume that the laws we get will reflect behaviours of audiences in 2008 Olympics (Assumption 3). Based on these laws, the customer traffics and shopping demands of the commercial network can be forecasted.

Customer Traffic and Shopping Demands of Commercial Sites

When it comes to customer traffic calculation, researchers' opinions differ greatly. Article Chen (2004) defines it as the total amount of person time that walk through the commercial site. Zhang (2004) uses "consumption flow". Yi (2006) brings up the concept of standardized shopping person.

Since there exists different definitions of customer traffic, using customer traffic as a decision variable to determine the construction scale of each commercial site may incur confusion among researchers. What is worse, it may lead to some errors.

In this paper, we propose a different approach for this job. We use the unit service capacity, and the shopping demand of each commercial site to determine the construction scale of each commercial site.

Optimization Models

When we have equipped with insightful information from the last two steps, we will begin considering the programming of the temporary commercial network. This task will be fulfilled by the following two steps.

First, we design an optimization model based on the three objectives—meeting shopping demand, balanced distribution, and making profits—to figure out the optimal construction scale of each site. Since we are dedicated to develop general methods that can be easily extended to similar commercial network programming problems in other situation, we carefully choose models and algorithms that are scalable, extendible, and flexible. The following reasons explain why we choose genetic algorithms (Sun, 1996).

- Genetic Algorithms use coded string in the search process rather than original variables;
- Genetic Algorithms iterate from group to group, thus reducing the possibility of falling into local optimal solutions;
- The search process is based on probability;
- Genetic Algorithms have no strict requirements (such as connectivity, convexity, etc) on the search space; the search processes require the fitness value only, no other supported information such as derivative is needed.

Second, after the construction scale of each commercial site has been calculated by the genetic algorithm-based model, we need to decide the types and the number of each type for each of the 20 sites. In this setting proposed by CUMCM (2004a), only two different types of size are available. Based on the economics of scale in retailing industry (Zhuang, 2000; Ma, 2002), we have the reasons to believe that the bigger one is more cost-effective. That is, after the construction scale are specified, each commercial site will first consider the big type, only when the remained construction scale is not enough for a big one, the small one is chosen.

SURVEY DATA ANALYSIS BASED ON DATA MINING TECHNIQUES

Data mining, which is a natural evolution of database technology with wide application, is the extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data. The functionalities of data mining include characterization, discrimination, association, classification, clustering, outlier and trend analysis, etc (Han, 2001).

In this section we will use two functionalities of data mining， clustering and association， to analyze the survey data.

Cluster Analysis

Green & Schaffer (1998) compares some alternative approaches on market segmentation and conclude that straightforward clustering of the original data using k-means is a viable approach. When using k-means for clustering, we need to determine the number of clusters at the very beginning. As reported by Green & Schaffer, five clusters is a typical number in market segmentation studies. Miller (1956) researches on the short-term memory capacity of human, and reported that the capacity is 7+2. According to this theory, the feasible number of clusters may between 5 and 9. However, Cowan (2000) reports the short-term memory capacity is 4. We choose the number of clusters not only reference the above theories, but also consider the specific context of the problem.

By using k-means in the software package—Clementine (SPSS, 2003)—we can get the clustering results in Table 1.

Table 1. The Clustering results of 2 runs of the k-means in Clementine. Sign "/" denotes fields that do not contain in the run.

	Data		Transporting						Eating		
Fields	Type	Values	C1	C2	C3	C4	C5	C6	C1	C2	C3
Bus(E&W)	set	0,1	1	0	0	0	0	0	/	/	/
Bus(S&N)	set	0,1	0	1	0	0	0	0	/	/	/
Subway(E)	set	0,1	0	0	1	0	0	0	/	/	/
Subway(W)	set	0,1	0	0	0	1	0	0	/	/	/
Taxi	set	0,1	0	0	0	0	1	0	/	/	/
Private Car	set	0,1	0	0	0	0	0	1	/	/	/
Chinese Food	set	0,1	/	/	/	/	/	/	1	0	0
Western Food	set	0,1	/	/	/	/	/	/	0	1	0
Supermarket	set	0,1	/	/	/	/	/	/	0	0	1
Age	Range	[1,4]	/	/	/	/	/	/	2.67	2.17	2.25
Expenditure	Range	[1,6]	2.52	2.66	2.43	2.43	2.55	2.53	2.35	2.59	2.50
Male(%)			63.4%	64.5%	55.4%	56.3%	33.7%	33.1%	53.3%	52.1%	52.1%
Headcount			1828	1774	2006	2024	2010	958	2382	5567	2651
Percentage			17.3%	16.7%	18.9%	19.1%	19.0%	9.0%	22.5%	52.5%	25.0%

From Table 1 we can learn that the first run segments audiences into 6 clusters by their transporting preferences; the second run segments audiences into 3 clusters by their eating habits. From the clustering results we get characteristics of each cluster, and treat audiences by clusters in further studies.

Association Rules Mining

We employ the Apriori model in Clementine (SPSS, 2003) to mine the association rules in the database of survey data. We set parameters as follows:

Minimun rule support	20%
Minimun rule confidence	60%
Maximun number of antecedents	5

Table 2. Strong association rules in the database of survey data. Note that Support refers to the percentage of records in the data for which the antecedents (the "if" part of the rule) are true (SPSS, 2003). This definition of support differs from the usual one.

No	Antecedent 1	Antecedent 2	Consequent	Support	Confidence
1	Age = 2		Western Food	58	61.9
2	Expenditure = 2		Sex = Male	24.8	61.8
3	Sex = Female	Age = 2	Western Food	28.3	62.4
4	Expenditure = 3	Age = 2	Western Food	32.4	61.5
5	Sex = Male	Age = 2	Western Food	29.7	61.5

Based on these association rules, we can effectively target services for audiences and enhance the overall service quality.

CUSTOMER TRAFFIC AND SHOPPING DEMAND OF COMMERCIAL SITE

In this section, we begin with working out the travelling paths of each cluster of customers. Based on these paths, we calculate the customer traffic and shopping demand of each commercial site.

Routes of Customer Traffic

The problem of finding the shortest path which customers travel is converted into a problem in graph theory. The commercial sites, taxi stations, bus stations, subway stations, parking lots, Chinese food restaurants, Western food restaurants, and supermarkets (for eating) in Figure 1 are saw as nodes of graph; roads are regarded as edges of the graph; edges are weighted according to the length between nodes.

Finding the shortest paths in a graph is a classical problem in Graph Theory and we solve it by applying the Dijkstra Algorithm (Yin, 1999).

Customer Traffic of Each Commercial Site

We propose the concept of shopping traffic in calculating the customer traffic of the commercial network. We define it as follows.

Shopping Traffic. One travel of each customer contributes 1 unit of shopping traffic to the commercial network. Customers contribute shopping traffic to the commercial sites they shop. And if a customer shops at n commercial sites in a travel, he contributes to the n sites in proportion to his spending, while his overall contribution to the network is 1 unit.

According to the definition and the shortest paths we have figured out in 4.1, the shopping traffic of each commercial site can be obtained.

Table 3. It shows the Shopping traffic of each of the 20 commercial sites.

Commer Site	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Shopping Traffic	27498	13006	12925	15075	19994	50503	19994	15075	12925	13006
Commer Site	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4
Shopping Traffic	15779	12554	23067	12554	15779	40267	13183	15415	13183	38219

Shopping Demand of Each Commercial Site

We have assumed that customers buying things by the way they travel, and they shop twice a day while the spending of the two are the same. Thus, based on the shopping traffic of each commercial site and the characteristics of each cluster of customers, we can figure out the shopping demand of each commercial site.

Table 4. It shows the Shopping demand (by hundred yuan) of each of the 20 commercial sites.

Commer Site	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Shopping Demand	34080	16204	16206	19002	25291	63999	25291	19002	16206	16204
Commer Site	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4
Shopping Demand	19911	15716	28634	15716	19911	51003	16575	19466	16575	47977

COMMERCIAL NETWORK PROGRAMMING

In the last two sections, we have mined the survey data and calculated the customer traffic and shopping demand of each of the 20 commercial sites in the commercial network. In this section, we will design optimization models for the quantitative analysis of the commercial network programming.

Using Commercial Capacity to Determine Construction Scale

We have figured out the shopping demand of each commercial site in section 4. Now based on such information, we will design a model to determine the reasonable construction scale of each of the 20 commercial sites in the commercial network.

$$s_i = \frac{q_i}{t}. \tag{1}$$

In (1), s_i represents the reasonable construction scale of the i^{th} commercial site in the network; q_i denotes the commercial capacity of the i^{th} commercial site; t refers to the unit service capacity. This model is referenced from Song (2002), but we have improved it. In their original model, q_i is the customer traffic; while t means the traffic one unit area can service per day. As we have mentioned before, since people differ in their meanings of customer traffic, using customer traffic as the decision variables will easily lead to misunderstandings and even errors.

In our improved model, q_i represents the shopping demand of the commercial site; while t refers to the shopping demand one unit area can supply per day. Now, it comes to work out what t is.

We get the total amount of area of national wide convenience stores and the total sales volume in year 2004 in China Statistical Yearbook 2005 (NBS, 2005). Based on the 80/20 principle (Koch, 1997) and adjusted by the amount of a single purchase (CCFA, 2005), we work out that t is 3194.40 *yuan* per unit area.

Table 5. Reasonable construction scale (by square meter) for each of the 20 commercial sites.

Commer Site	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Construction Scale	1067	507	507	595	792	2003	792	595	507	507
Commer Site	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4
Construction Scale	623	492	896	492	623	1597	519	609	519	1502

Genetic Algorithm-Based Optimization Model

The model in the last section determines the construction scale by the commercial capacity. That is, it just meets one of the objectives, meeting shopping demand. In order to reach the three objectives of the design—meeting shopping demand, balanced distribution, and making profits—we design an optimization model based on genetic algorithms to calculate the optimal construction scale of each site.

Meeting shopping demand.
The construction scales in Table 5 is figured out according to the shopping demand. We let the optimal construction scale change in a certain range around the figure in Table 5. In the genetic algorithms, we set the range of construction scale as follows.

$$s_i(1-k) \le x_i \le s_i(1+k)\,. \tag{2}$$

In (2), x_i represents the optimal construction scale; s_i is the reasonable construction scale in Table 5; k is the parameter which can be set according experience value. In this paper, we assign 0.2 to k .

Balanced distribution.
In our model, the definition of balanced distribution is that the more closely the construction scales among the 20 commercial sites, the more balanced it is.

$$\min E_1 = \sum_{i=1}^{20} \sqrt{\left(x_i - \overline{x}\right)^2}\,. \tag{3}$$

In (3), x_i represents the construction scale of the i^{th} commercial site; $\overline{x}$ denotes the average construction scale of the 20 commercial sites; E_1 is the sum of the Euclidean distance between x_i and $\overline{x}$.

Making Profit.
According to principles of the evaluation of economic effectiveness, when there are several plans that create the same outcome, the one with the least investment is a best one (Zhou, 2004). The building cost of the temporary commercial network is directly in proportion to its construction scale. In order to make profit, we should choose the plan with smallest possible construction scale that meet the demand of audiences.

$$\min E_2 = \sum_{i=1}^{20} x_i\,. \tag{4}$$

In (4), x_i represents the construction scale of the i^{th} commercial site; E_2 is the sum of construction scale of the 20 commercial sites.

Programming Model.
This programming model seeks the minimum value of weighted sum of the above two objectives E_1 and E_2, subjected to the constrain of meeting the shopping demand.

$$\begin{aligned} &\min E = \alpha E_1 + \beta E_2 \\ &s.t.\ \ s_i(1-k) \le x_i \le s_i(1+k) \end{aligned}\,. \tag{5}$$

In (5), α, β are the weighted value, while $\alpha + \beta = 1$. α, β can be set by the experience value. In this paper, we let both α, β equal 0.5. x_i represents the optimal construction scale; s_i is the reasonable construction scale in Table 5; k is the parameter which we assign 0.2 in this paper.

Using Genetic Algorithm to Solve the Optimization Model

Genetic Algorithms, which are based on Darwin's theory of evolution and Mendelian genetics, are proposed by Dr. Holland (1975). Genetic algorithms are global optimization algorithms that apply principles of evolutionary biology to computer science. It uses biologically-derived techniques such as selection, crossover and mutation in the searching process (Shao, 2000; Wang, 2002).

Table 6. We set the parameters of our genetic algorithm as follows.

Individual Number	40
Chromosome Length	40
Maximum Generations	500
Generation Gap	0.9
Probability of Crossover	0.7
Probability of Mutation	0.01

Table 7. We select the genetic operators of our genetic algorithm as follows.

Selection Operator	Roulette Wheel Selection
Crossover Operator	Single-Point Crossover
Mutation Operator	Simple Mutation

We run the program for 500 iterations. The optimal solution is obtained at the 482nd iteration, while the value of the objective function is $\min f(x) = 8810.57$.

Table 8. The optimal construction scales of the 20 commercial sites.

Commercial Site	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Optimal Construction Scale	854	406	407	476	633	1603	634	476	406	406
Commercial Site	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4
Optimal Construction Scale	499	394	717	394	499	1278	415	488	415	1202

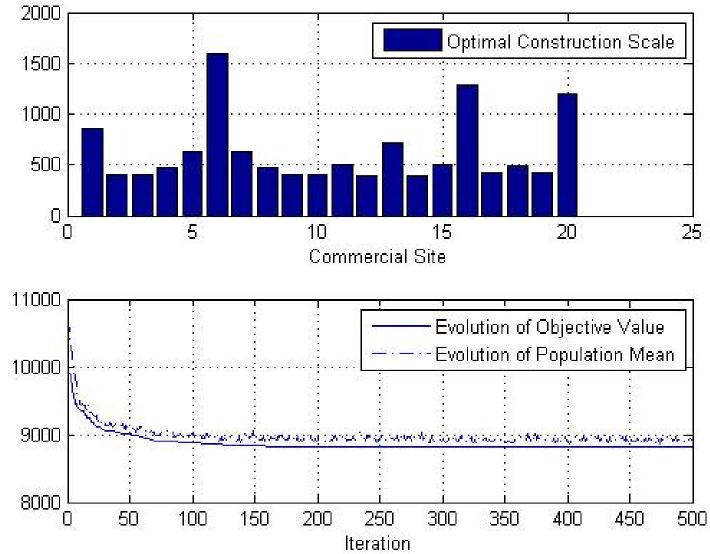

Figure 3. The Optimal Solutions and the Solving Process.

Results Analysis.
The method in 5.1 takes into account just one objective, meeting the shopping demand. The Genetic Algorithm-Based approach in 5.2 simultaneously considers three objectives: meeting shopping demand, balanced distribution and making profit.

As for the method in 5.1, the balance measure $E1 = 6275.75$, while the profit measure $E2 = 15745.33$.

As for the approach in 5.2, the balance measure $E1^{'} = 5018.69$, while the profit measure $E2^{'} = 12602.45$.

From the comparison we can conclude that results obtained by the genetic algorithms-based approach in 5.2 are more balanced and profitable.

Linear Programming Model for Supermarket Type Selection

Based on the analysis of the economies of scale in retailing industry (Zhuang, 2000; Ma, 2002), we establish the principle that when there are two types of size of supermarket to be selected, we should first consider the bigger on for it is more cost-effective. Only when the remained construction scale is not large enough for a big one, the small one is chosen.

$$\begin{aligned} &\min N_i = m_i + n_i \\ &s.t.\ m_i A_1 + n_i A_2 = x_i \\ &i = 1, 2, \cdots, 20 \\ &m_i, n_i \text{ is integer} \end{aligned} \quad . \qquad (6)$$

In (6), x_i is the optimal construction scale of the i^{th} commercial site; A_1 is the size of the big supermarket; A_2 is the size of the small supermarket; m_i represents the number of big supermarket in the i^{th} commercial site; n_i represents the number of small supermarket in the i^{th} commercial site. In this paper, we assign 120 square meters and 50 square meters to A_1 and A_2 respectively based on the analysis by Liu (2005).

We use Ling software package (LINDO, 1999) to solve this linear programming model. When using Lingo, we loosen the integer constrain of n_i . After the solutions are obtained, we rounded n_i into integer.

Table 9. The number of the two different types of mini supermarket in the 20 commercial sites.

Commercial Site	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Big Supermarket	7	3	3	3	5	13	5	3	3	3
Small Supermarket	0	1	1	2	1	1	1	2	1	1
Commercial Site	B1	B2	B3	B4	B5	B6	C1	C2	C3	C4
Big Supermarket	4	3	5	3	4	10	3	4	3	10
Small Supermarket	0	1	2	1	0	2	1	0	1	0

CONCLUSION

This paper is dedicated to develop optimization models for the temporary commercial network programming of the 2008 Beijing Olympics. By mining the survey data, we obtain insightful information about the audience; based on the extracted information, we develop quantitative methods for the analysis about the commercial network. The results we obtained are valuable for the reference for the decision-making of the commercial networks around the main stadiums.

Acknowledgments. We are profoundly grateful to colleagues, students at Xi’an Jiaotong Univ. and South China Univ. of Tech., who constantly challenge and test us, helping to make our work more complete. Special thanks goes to Xiang-zhen Kong, Man-fa Liang, Yi Hong, and Rong-bin Ruan for their valuable opinions.

REFERENCES

Ai, L., Sun, J., and Liu, Y. (2004). The Application of Data Mining in the Distribution of Miniature Commerce Net Sites, *Chinese Journal of Engineering Mathematics*. 21(7):64-70.

Chen, Y., Ruan, R., and Luo, J. (2004). Designs for Business Sections of Beijing Olympic Games, in *Proceedings of the Fourth Graduate Symposium of SCUT on Computer Science,* South China Univ. of Tech. Press, Guangzhou.

China Chain Store & Franchise Association (CCFA) (2005). *Statistics Report for the First Half of 2005*. http://www.ccfa.org.cn/research/2005/2005_1-6.pdf.

Cowan, N. (2000). The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity, *Behavioral and Brain Sciences*. 24:87-185.

CUMCM Committe (2004a). *CUMCM-2006: Contest Problem A*. http://mcm.edu.cn/ENG/PROBLEMS/2004ABCD.PDF.

CUMCM Committe (2004b). *CUMCM-2006: Database for Contest Problem A*. http://mcm.edu.cn/mcm04/A2004datae.rar.

Han, J., and Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Higher Education Press, Beijing.

Holland, J. (1975). *Adaptation in Nature and Artificial Systems.* University of Michigan Press, Ann Arbor.

Hu, C. (2005). The design model of temporary supermarket sites of Olympic Games, *Journal of Xinyu College*. 10(2):55-58.

Koch, R. (1997). *The 80/20 Principle*. Nicholas Brealey Publishing Ltd., London.

Li, G., Zhou, S., and Wei, S. (2004). Optimum Design for Networks of Beijing Olympic Temporary Supermarkets, *Chinese Journal of Engineering Mathematics*. 21(7):49-56.

LINDO Systems Inc. (1999). *LINGO User's Manual*. Chicago, Illinois.

Liu, Y. (2005). Convenience Stores—Strong Business in th 21st Certury, *Co-Operative Economy & Science*. (13):31.

Ma, S., and Min, Z. (2002). An Analysis on Competitiveness of Retail Enterprises, *China Business and Market*. 16(6):47-50.

Meng, D. (2004). Design Temporary Contact Points in Commercial Network About the Main Stadium During the 2008 Beijing Olympic Games, *Chinese Journal of Engineering Mathematics*. 21(7):71-75.

Meng, J., Tian, X., and Gao, X. (2004). The Research of the Programming Model for the Provisional Supermarket Site, *Chinese Journal of Engineering Mathematics*. 21(7):36-42.

Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity of Processing Information, *The Psychological Review*. 63(2):81-97.

National Bureau of Statistics of China (NBS) (2005). *China Statistical Yearbook 2005*. China Statistics Press, Beijing.

Schaffer, C., and Green, P. (1998). Cluster-based market segmentation: Some further comparisons of alternative approaches, *Journal of Market Research Society*. 40(2):155-163.

Shao, L., Zhang, J., and Wei, C. (2000). *Introduction to Artificial Intelligence*. Publishing House of Electronics Industry, Beijing.

Song, S., and Cao, C. (2002). The Study of Commercial Capacity and the Construction Scale, *Huazhong Architecture*. 20:47-49.

SPSS Inc. (2003). Clementine 8.0 User's Guide. Chicago, Illinois.

Sun, Y., and Wang, Z. (1996). The Application of Genetic Algorithm in Optimization, *Control and Decision*. 11(4):425-431.

Tang, X. (2002). Summarize of Data Mining (DM) and Knowledge Discovery Database (KDD), *Computer Development & Applications*. 15(4):31-33.

Tang, X., Feng, Y., Xiao, J., and Pei, B. (2005). The Optimal Design of Temporary Mini Supermarket Network in the Olympics, *College Mathematics*. 21(4):7-12.

Wang, T., and Li, B. (2005). The Optimal Programming Model of Temporary Mini Supermarket of Beijing 2008 Olympic Games, *Journal of Leshan Teachers College*. 20(12):29-32.

Wang, X., and Cao, L. (2002). *Genetic algorithm—Theory, Application and Actualization*. Xi'an Jiaotong Univ. Press, Xi'an.

Xiang, X., and Jiao, L. (2003). Clustering Method in the Field of Data Mining, *Journal of Circuits and Systems*. 8(1):59-67.

Xue, F., Fu, Q., and Xie, J. (2004). A Profit-Oriented Microeconomic Integer Programming Model for Olympics' Mini Supermarket Allocation, *Chinese Journal of Engineering Mathematics*. 21(7):57-63,108.

Yi, K., and Zhang, C. (2006). Optimied Model of Site Selecting for Temporary Supermarket in Olympic Games, *Mathematics in Practice and Theory*. 36(1):204-214.

Yin, R., et al. (1999). *Data Structure--Using Object-Oriented Method and C++ Language*. Tsinghua University Press, Beijin.

Zhang, R., Li, J., and Qiu, T. (2004). Reasonable Design of MS Site, *Chinese Journal of Engineering Mathematics*. 21(7):29-35.

Zhou, Y., Li, G., and Cai, G. (2004). *Technology Economics 2nd ed*. Publishing House of South China Univ. of Tech., Guangzhou.

Zhuang, G., and Feng, G. (2000). An Empirical Analysis of the Economics of Scale in Chinese Retailing Enterprises, *Finallce & Trade Economics*. (9):44-48.