A Study on the Association Rule Mining and GIS to Seniors Living Alone in Taiwan Population Census

Chin-Jui Chang[1], Chi-Hsin Chin[2] and Chien-Hui Liu[3*]

[1]*Department of Information and Electronic Commerce Management, Transworld University, Douliou, Yunlin,Taiwan.*
[2]*Department of Marketing Management, Transworld University, Douliou, Yunlin, Taiwan.*
[3]*Department of Multimedia and Animation Design, Transworld University, Douliou, Yunlin, Taiwan.*

Authors' contributions

All three authors work as a team in Transworld University. Author CJC designed the study, performed the data mining analysis and used GIS tool to display the results. Author CHL collected the census data and wrote the first draft of the manuscript and managed literature searches. Author CHC managed the data mining analyses of the study and literature searches. All three authors read and approved the final manuscript together.

Article Information

Review Article

ABSTRACT

The elderly have been considered disadvantaged social classes, and their widening problems will have a tremendous impact and influence on society. This study aims to apply data mining techniques to investigate the demographic features of socially disadvantaged groups in Taipei and Kaohsiung by using population data collected in the 2000 census to provide reference for social welfare decision makers in understanding these groups and forming policy. The demographic features, marital features and educational attainment of the elderly lives were investigated. The marital features, educational attainment, care and life patterns of the elderly were studied. This

_____

Corresponding author: E-mail: cha0722@gmail.com;

article offers two results of creative research about seniors living alone in Taiwan. (1) Suggestions on a research structure for data mining visualization with GIS. (2) A search for the distribution of the elderly population groups in society using the census 2000 as research background.

1. INTRODUCTION

Aging is a global problem and will be one of the major challenges of the 21st century. In contrast to Europe and North America, which have undergone protracted demographic transformations over the past hundred years, and have developed various welfare measures and systems to better care for the elderly, population aging is a relatively new and rapidly-developing trend in Taiwan. According to a national statistical report issued by the Directorate General of Budget, Accounting & Statistics, Executive Yuan, due to the declining birth rate and the increasing life expectancies of citizens, Taiwan's elderly population (those 65 years of age and older) exceeded 7% in 1993, putting Taiwan among the ranks of countries with aging populations, and the "graying" of the population is becoming increasingly evident. Furthermore, according to the Council for Economic Planning and Development's population estimates [1], Taiwan's elderly population will rapidly increase to 3.6 million persons over the next two decades (by roughly 2021), and the elderly will account for over 14% of the population by that time. In contrast, the support ratio between working-age persons and the elderly will shift from the current 8 : 1 to 4.7 : 1 during the next twenty years. This rapid transformation, especially the tremendous growth of the elderly population expected over this very short period of only two or three decades, will have an extreme impact on Taiwan's social structure, family relationships, and the government's health and social welfare measures, and even the country's political and economic situation. In Taiwan's revised Senior Citizens Welfare Act (1997), Article 9 prescribes that local governments may, when needed, establish and fund various types of senior welfare organizations. Seniors' main needs consist of the four primary items of housing, care, long-term care, and entertainment/recreation. With regard to government welfare services for the elderly, Article 105 of the Constitution explicitly requires: "The elderly, very young, disabled, those unable to live independently, and victims of various disasters shall be provided with appropriate support and assistance." Furthermore, the Executive Yuan's 1994 Social Policy Framework established the government's commitment to senior welfare services in official policy. Article 17 of the Social Policy Framework prescribes that: "The government shall reinforce health and care facilities for seniors in the community, and shall establish residential care service networks in conjunction with community resources." In order to help the elderly perform career planning at an early date, and maintain a healthy living attitude, this project applied data mining techniques to investigate demographic structural correlations concerning seniors living alone based on past demographic survey data research literature [2]. In addition, spatial autocorrelation analysis is employed to study the spatial distribution and clustering of seniors living alone, enabling decision-makers to understand the demographic characteristics and spatial distribution of seniors living alone, and thereby determine appropriate policies that may benefit the public. The elderly living alone constitute a disadvantaged group from the perspective of society, and their problems will have an enormous impact on Taiwan's society in the future. This project seeks to investigate the current population of elderly living alone in Taiwan, and it is hoped that the results of the analysis can provide the government with guidance in allocating social resources.

2. MATERIALS AND METHODS

The content of this chapter integrates the three areas of population census data, spatial data and geographical information systems, and data mining, and conducts an exploratory analysis of relevant major cases in the literature from the last few years [3]. Fig. 1 shows a flowchart for the literature review process. Data mining relies on exceptional data processing and operating capabilities to perform multi-dimensional searches of large bodies of data, and extract knowledge latent in the data [4]. By uncovering useful information needed by users hidden in databases, data mining can provide companies and users information needed in decision-making, helping companies to take advantage of opportunities. Unlike algorithmic methods, data mining is not used to prove or disprove hypotheses, but rather to find hidden patterns in data by means of induction and to generate hypotheses. Users can use the intelligent data

analysis provided by data mining to gain a more thorough understanding of the content of data, enabling them to effectively resolve various problems they may encounter. Starting in the 1990s, data mining has been widely and successfully applied in market surveys, marketing analysis, business decision-making analysis, manufacturing project control, bio-information research, etc. [5]. Use of the data mining classification method to establish a data classification model has two chief goals: (1) To analyze factors affecting data classification: Classification models established using the classification method can be used to extract classification rules, and these rules constitute factors affecting data classification [6]. (2) Predicting data type: When the data used to establish a classification model is sufficient in quantity, the representativeness of the classification model will have statistical significance [7].

This study used census data to develop a spatial analysis model employing a data mining mechanism. This study aims to apply data mining techniques to investigate the demographic features of socially disadvantaged groups in Taiwan by using population and household data collected in the 2000 census to provide reference for social welfare decision makers in understanding these groups and forming policy. The chief goal of the study was to employ a data mining analytical model and geographical information system (GIS) spatial analytical model with census data [8]. This study's contributions include two aspects: (1) Interdisciplinary research: This study used data mining techniques to analyze and explain the linkage between social and demographic characteristics and used GIS technology to explain the spatial distribution of social and demographic characteristics, which will help to strengthen analytical techniques used in the social sciences. (2) Innovation in research methods: The study used GIS to visualize spatial distributions and display the results of the data mining analysis, which enabled data mining techniques originally without spatial implications to yield meaningful spatial distribution results [9]. The study consisted of two stages; the first stage comprised: (1) collection and establishment of a demographic census database; (2) in-depth analysis of data mining applications technology; and (3) selection of Taipei and Kaohsiung cities as the research areas. The second stage comprised: (1) using data mining to establish an analytical model, and (2) visualization of knowledge rules derived using data mining on

GIS images, followed by the use of spatial autocorrelation to analyze the spatial distribution and degree of clustering [10]. The research process as shown below was developed based on seven steps proposed for knowledge discovery in databases (KDD) [11]. The research steps employed in this study were as follows:

1. Confirmation of research questions: To investigate the background of the elderly population living alone, establish the research scope of the elderly population living alone, and thereby form a research concept concerning the elderly living alone.
2. Development of research motivation and goals: To use data mining to find relevant constituent factors concerning the elderly living alone.
3. Review of the literature: To collect literature concerning the application of data mining technology and spatial analysis technology in censuses, and perform reading, investigation, analysis, and summarization.
4. Development of a research framework: To establish an overall framework for this study, census research categories, and census research subjects based on the results of the literature review and the correlations between research questions.
5. Data mining techniques: The first type is association rule mining proposed by Agrawal and Srikant [12] for finding interesting associations and correlations among sales items in huge volumes of sales data. Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true [13,14].
6. Collection of census data: The Fourth Bureau of the Directorate General of Budget, Accounting and Statistics provided census data for use by this project.
7. Establishment of an analytical model: To establish an analytical model generating knowledge primarily using association rule mining techniques and secondarily using GIS spatial analysis methods [15].
8. Research results and discussion: To assess the various results of analysis and explain the meaning of spatial distribution findings.
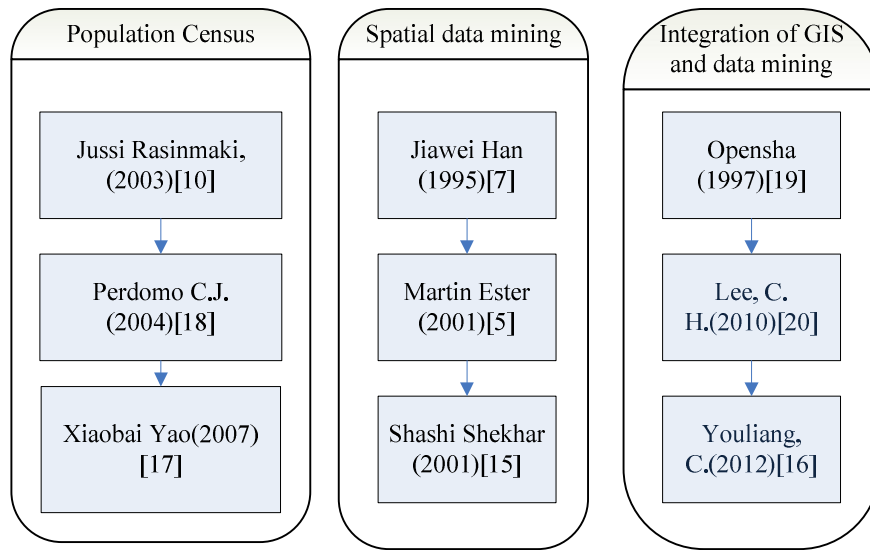
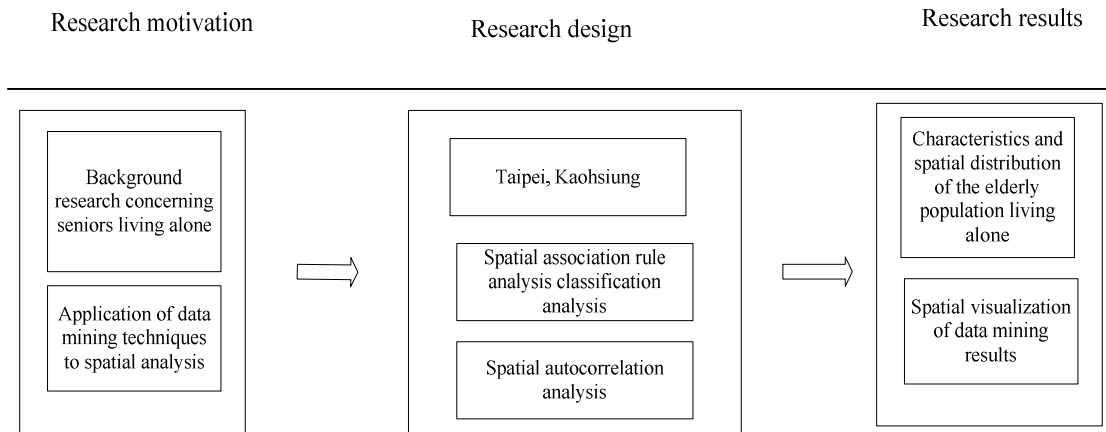**Fig. 1. Flowchart of literature review process**

Fig. 2. The research procedure for spatial distribution of data mining results

Fig. 2 above displays research procedures. The motivation for this study included a wish to determine the background of seniors living alone and apply data mining techniques to spatial analysis. The research design included a spatial association rule analysis and spatial autocorrelation analysis [16]. The study's results included the demographic characteristics and spatial distribution characteristics of the elderly population living alone.

3. RESULTS

The results of association rule mining showed that the distribution of seniors living alone in Taipei and Kaohsiung is correlated with the distribution of medical organizations. The distribution of these medical organizations can provide a useful reference for other areas. For instance, by connecting this distribution information with the government's medical resource allocation planning, medical resources can be allocated to where they are most needed, reducing social costs. When ranked in terms of the number of medical organizations, from the fewest to the most, the districts of Taipei City are listed as follows: Nangang, Beitou, Wenshan, Neihu, Datong, Wanhua, Songshan, Zhongzheng, Xinyi, Shilin, Zhongshan, and Da-an. Here the peripheral districts of Beitou, Wenshan, and Datong are all ranked as level 5. This result suggests that districts located in peripheral areas

are may lack medical resources compared to other districts. The fact that the hilly, peripheral districts Nangang, Beitou, and Wenshan have fewer per capita medical resources than other districts indicates that medical resources do not favor hilly areas (such as Wenshan, Nangang, Neihu, Beitou, Shilin). In contrast, while also in a peripheral location, the Datong District, which has relatively level terrain and was developed earlier, has more abundant medical resources than the hilly areas. Shilin was also developed at an early date, and has enjoyed strong economic development and has comparatively sufficient medical resources. In terms of medical resources per square kilometer, the foregoing districts have relatively low levels of medical resources per unit area, there is little likelihood that residents of other districts will seek care in these districts, and the medical choices of local residents are comparatively few, especially for those persons living in hilly areas. The fact that the peripheral, hilly, and relatively undeveloped districts of Wenshan, Nangang, and Beitou are lacking in medical resources indicates that the level of medical care is influenced by such factors as location, topography, development history, and current state of economic development.

Medical resources are unevenly distributed among the districts of Taipei City. Most such resources are located in core areas, and are relatively scarce in peripheral areas. This suggests that districts with a high level of economic development and good accessibility enjoy funded medical resources. The residents of areas where clusters of medical organizations exist enjoy a wider range of healthcare options, but the residents of most peripheral districts are not so fortunate. This situation happens to parallel the commercial principle of pursuing the greatest possible profit. With regard to the relationship between the distribution of seniors living alone and that of medical care organizations, apart from census data, input data also included location information for medical care organizations in Taipei and Kaohsiung. Medical organizations are distributed as shown in the attached diagrams. We designed the following terms to facilitate analysis of spatial association rules: P1: administrative district, P2: marital status, P3: level of education, Q1: relationship between administrative district and medical organizations (Close_to). Close_to = High: indicates that there is a relatively high number of medical organizations in an administrative district containing seniors living alone, and implies that there is a is short actual distance to a medical organization; Med: indicates that there is a moderate number of medical organizations in an administrative district containing seniors living alone, and implies that there is a is medium actual distance to a medical organization; Low: indicates that there are relatively few nearby medical organizations, and implies that there is a long actual distance to a medical organization. The derived spatial association rules are shown in Table 1. which shows the numbers of medical organizations in districts with different densities of seniors living alone.

(administrative district) ^ (marital status, level of education of seniors living alone)→ (spatial relationship with medical care organizations), (S%; C%); the following rules were obtained:

Rule 1, P1(administrative district) ^ P2(marital status) ^ P3(level of education)→ Q1(X: administrative district, spatial relationship with medical care organizations), (S%; C%).

Table 1. Analysis of association rules between seniors living alone in Taipei City and medical care organizations

| Item	Degree of support	Degree of confidence	Second term (proximity)	Former term 1 (administrative district)	Former term 2 (level of education)	Former term 3 (marital status)
3507	7.58	29.0	High	Da-an	University	Divorced or separated
3194	6.91	41.0	Med	Songshan	High school	Divorced or separated
1934	4.18	96.0	High	Xinyi	Junior high	Spouse deceased
1881	4.07	60.0	Med	Shilin	Elementary school	Spouse deceased
1463	3.16	53.0	Low	Nangang	Elementary school	Spouse deceased

5
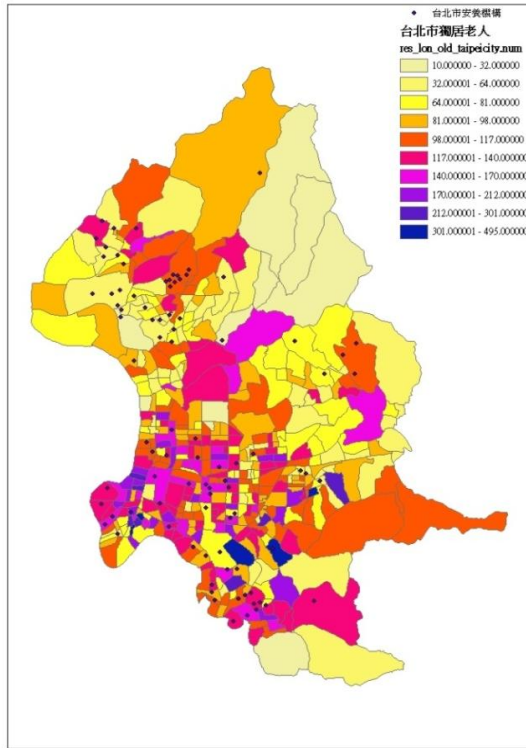
Fig. 3. Distribution of seniors living alone and medical organizations in Taipei

The correlation rules may contain two or more dimensions (dimension) or predicate (predicate), for example: Age (X, "40 ... 45") ^ income (X, "7 l ... 80000) → purchase (X, overseas funds). The above example will contain 'age, income, purchase' three predicates, if each predicate in the rules were only appears once, we will call interdimensional connected rules, shall be guilty of the above-mentioned examples. The results of analysis of association rules for Kaohsiung seniors living alone indicated: Rule 1, for age

greater than 74 years, educational level of illiterate, marital status of spouse deceased, and no income from employment, there were 9,497 data items, the degree of support was 48.94%, and the degree of confidence was 31%. Rule 2: for age less than 71 years, educational level of junior high school, marital status of divorced or separated, and no income from employment, there were 9,108 data items, the degree of support was 46.94%, and the degree of confidence was 31%. Most elderly persons living alone in Kaohsiung are approximately 70 years of age, have no income from employment, and have an elementary school educational level of or are illiterate. After calculating the degree of support for Association Rule 1 in each administrative district, we found that such areas as Fengming Li, Lungfeng Li, Fengsen Li, Fengxing Li, Fengyuan Li, and Haiyuan Li in the Xiaogang District in the southern part of Kaohsiung City, and Weibei Li, Fushan Li, and Caigong Li in the Zuoying District all have relatively high populations complying with Association Rule 1.

The results of the spatial association rule analysis for Kaohsiung seniors living alone (Table 3) were as follows: Seniors living alone in Kaohsiung's Lingya District have access to a large number of nearby medical organizations, which indicates that seniors living alone can obtain medical resources where there are many medical organizations. In contrast, seniors living alone in the Zuoying and Nanzi districts have access to a moderate number of medical organizations, and there are great numbers of seniors living alone in these two districts. Finally, seniors living alone in the Qijin and Xiaogang districts have access to relatively few nearby medical organizations, which suggests that there

Table 2. Results of association rule analysis for seniors living alone in Kaohsiung

Item	Degree of support	Degree of confidence	Latter term	Former term 1	Former term 2	Former term 3
9497	48.94	31.0	Illiterate	Age >74	Spouse deceased	No job
9108	46.94	31.0	Junior high school	Age <71	Divorced or separated	No job
8221	42.37	12.0	Self-educated	Age >71	Divorced or separated	No job
7429	38.29	49.0	Elementary school	Age >65	Divorced or separated	No job
5568	28.7	11.0	High school	Age >65	Divorced or separated	No job

Table 3. Results of association rules about Seniors living alone close to medical care organizations in Kaohsiung

Item	Degree of support	Degree of confidence	Latter term (spatial relationship)	Former term 1 (administrative district)	Former term 2 (educational level)	Former term 3 (marital status)
2641	13.61	90.0	High	Lingya	Elementary school	Spouse deceased
1317	6.79	83.0	Med	Zuoying	Elementary school	Unmarried
715	3.68	75.0	Med	Nanzi	Elementary school	Spouse deceased
609	3.14	57.0	Low	Xiaogang	Illiterate	Spouse deceased
565	2.91	100.0	Low	Qijin	Elementary school	Spouse deceased
319	1.64	100.0	Low	Qijin	Illiterate	Spouse deceased

is a lack of medical care organizations in these two districts. Fig. 4 shows the distribution of seniors living alone and medical organizations in Kaohsiung.
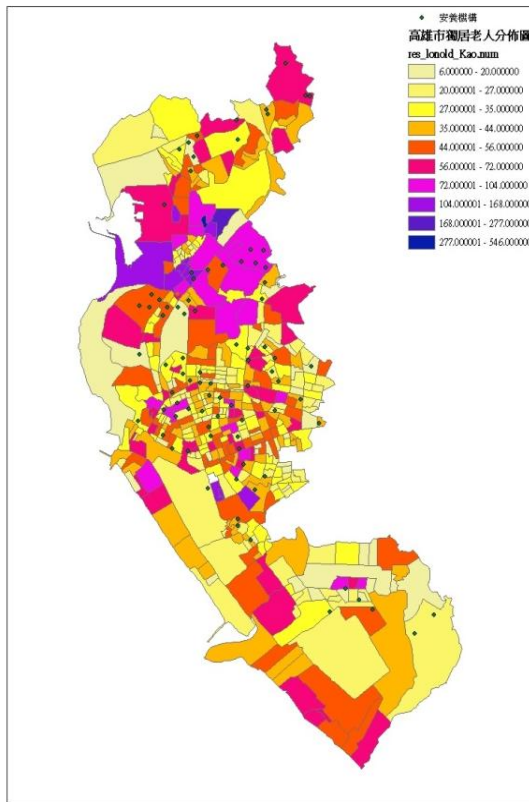

Fig. 4. Distribution of seniors living alone and medical care organizations in Kaohsiung

The clustering of seniors living alone in the Zuoying District can be explained by the fact that this district has historically contained a naval port, armed forces agencies and arsenals, and villages for military dependents; the sale of some military base land, renovation of old military dependents' villages, and disuse of some land once occupied by military dependents has induced many relatively affluent families to move away in search of a better living environments, leaving many less affluent persons and families to make do with the poor environment. As a consequence, many less-well-off seniors living alone, such as old soldiers living alone in villages for military dependents, face deteriorating living conditions. Many soldiers came to Taiwan with the government after the mainland was overrun by Communist forces, and usually left most of their family members in China. As a result, these old soldiers typically led solitary lives in Taiwan, and the lack of family members to care for them often left them with eccentric personalities and poor social skills in old age. Indifferent to their neighbors, the old soldiers have gradually become socially isolated in many cases.

4. CONCLUSION

Taiwan is small and has an abundance of relatively accessible medical resources. Although seniors living alone are not necessarily highly dependent on medical care organizations, the government should refer to the placement programs implemented in developed countries

when drafting placement arrangements for the elderly, and should pay special attention to health care for seniors. The government's efforts in this regard include the "Six-year National Construction Plan" proposed by the Council for Economic Planning and Development, Executive Yuan in 1991; this plan includes such items as encouragement for family care of the elderly, residential services for the elderly, residential nursing, and senior day care. In the case of social welfare agencies, the Ministry of the Interior (Department of Social Affairs), provincial and municipal social welfare bureaus, and city and county governments are all striving to implement residential services, daytime care services, and low-cost lunches for the elderly, and improve other specific care measures aimed at seniors. Many community care programs have been initiated, and private care organizations have gotten involved. As a result, community services for the elderly are growing more diversified. The results of the analysis in this study can serve as a useful reference for relevant medical resource allocation efforts.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Chin-jui Chang, Shiahn-wern Shyue. A study on the application of data mining to disadvantaged social classes in Taiwan Population Census. Expert Systems with Applications. 2010;38(4):48-59.

2. Chin-jui Chang, Shiahn-wern Shyue. Spatial and temporal data mining in census of population and housing. BAI 2006 International Conference on Business and Information at Pan Pacific Hotel, Singapore; 2006.

3. Jyotshua Solanki, Yusuf Mulge. Different techniques used in data mining in agriculture. 2015;256-267.

4. Koperski K, Han J. Discovery of spatial association rules in geographic information database. In Proc. 4th Int. Symp Large Spatial Database (SSD'95). 1995;47-66.

5. Martin Ester, Hans-Peter Kriegel, Jorg Sander. Algorithms and applications for spatial data mining, In Geographic Data Mining and Knowledge Discovery Edited By Harvey J. Miller, Jiawei Han, London and New York. 2001;161-187.

6. Bather El-Geresy, Christopher Jones. Models and queries in a spatio-temporal GIS, In GIS and geocomputation, Edited by Peter Atkinson and David Martin. 2000;27-39.

7. Koperski K, Han J. Discovery of spatial association rules in geographic information database. Proc. 4th Int. Symp Large Spatial Database (SSD'95). 1995;47-66.

8. Ester M, Hans-Peter Kriegel, Jörg Sander. Algorithms and applications for spatial data mining, Edited by Miller, J. H. and Han, J. Geographic Data Mining and Knowledge Discovery, London: Taylor & Francis. 2001;158-187.

9. John F. Roddick, Max J. Egenhofer. Spatial, Temporal and spatio-temporal database. Hot Issues and Directions for PhD Research. 2004;33(2).

10. Jussi Rasinmaki. Modelling spatio-temporal environmental data, Environmental Modelling & Software. 2003;877–886.

11. Han J, Micheline Kamber. Data mining concepts and techniques. Fraser University, Moran Kaufmann Publishers. 2001;45-61.

12. Agrawal R, Srikant R. Fast algorithm for mining association rules in large database. In Research Report RJ 9839, IBM Almaden research Center, San Jose, CA; 1994.

13. Quinlan J.R., Induction of association rules. Machine Learning. 1986;1:81-106.

14. Quinlan JR. APRIORI Programs for Machine Learning, San Mateo, CA: Morgan Kaufmann.; 1993.

15. Shekhar S, Sanjay Chawla. Introduction to spatial data mining, spatial databases: A tour. Prentice Hall; 2003.

16. Youliang C, Wang Z, Zhicheng C. Implementation of object-oriented GIS data model with topological relations between spatial objects. Int Journal of Advanced Computer Science. 2012;2(9):334-338.

17. Yao Xiaobai. Where are public transit needed – Examining potential demand for public transit for commuting trips. Computers, Environment and Urban Systems. 2007;31:535–550.

18. Perdomo CJ, Vilalta Y. The local context and the spatial diffusion of multiparty competition in Urban Mexico 1994–2000, Political Geography. 2004;23:403–423.

19. Openshaw S, Openshaw C. Artificial intelligence in Geography, John Wiley and Sons LTD; 1997.
20. Lee CH, Lee SH, Kim JW. A personalized clothing recommender system based on the algorithm for mining association rules. Journal of Korea Society for Simulation. 2010;19(4):59-66.