INTEGRATION OF GIS AND DATA MINING TECHNOLOGY FOR LAND MANAGEMENT

Ya-Wen Lin*[1], Ming-Chi Wu[1], Yao-Cheng Kuo[1], Chi-Farn Chen[2]
[1]Research Assistant, Center for Space and Remote Sensing Research, National Central University,
300, Jhongda Rd., Jhongli, Taoyuan 32001, Taiwan; Tel: +886-3-4227151#57659
Email: asarwen@csrsr.ncu.edu.tw

[2]Professor, Center for Space and Remote Sensing Research, National Central University,
300, Jhongda Rd., Jhongli, Taoyuan 32001, Taiwan; Tel: + 886-3-4227151#57624;
E-mail: cfchen@csrsr.ncu.edu.tw

KEY WORDS: Land Management, Spatial Data Mining, Knowledge Discovery

ABSTRACT: Spatial data mining is growing in popularity as a tool to discover implicit information from massive amounts of spatial and temporal data which can help to understand complicated environmental issues. This study explores the applicability of data mining and knowledge discovery in combination with Geographic Information System (GIS) technology to find pertinent information from the land use/land cover change detection database. The main objective of this study is to utilize spatial data mining techniques to clarify the association patterns hidden within spatial attribution and land change investigation data for non-urban land to gain insight into potential causal relationships between land exploitation and spatial objects. The extracted knowledge can serve as a reference for authorities to observe the essential elements of illegal agrarian usage and lead to more effective land management policies for protecting Taiwan's land resources.

1. INTRODUCTION

Traditional Geographic Information Systems (GIS) technology alone is not designed to discover the valuable knowledge hidden behind the massive amounts of graphic and attribute data. The availability of vast sets of high-resolution spatial and spatio-temporal data has dramatically enhanced spatial analysis leading to the development of an integrated data mining and GIS technology approach that can be used to gain new knowledge and a better understanding of complex geographic phenomena (Gómez et al., 2011; Guo and Mennis, 2009; Kondaveeti et al., 2011; Zhu et al., 1998). Collaborative efforts in spatial data mining underpins the knowledge discovery technologies, providing new opportunities to reveal significant patterns and trends in geographic data arising due to the human–environment interaction (Guo and Mennis, 2009).

Limited land resources have made efficient land use management a pressing issue in Taiwan. The land management administration, the Construction and Planning Agency Ministry of the Interior (CPAMI), has been funding a "Land Use Change Detection Program" since 2002. A large-scale land use/land cover change detection system has been developed to assist with investigation of land practices from temporal satellite images. Through long-term regular monitoring of land change and the results of investigation from the relevant government agencies, a huge mass of geospatial data has accumulated. In the last ten years, that the ratio of illegal land usage to all land changes is approximately 1:4. This survey focuses on understanding the extent of this type of exploitation, the changes it has made within regions and circumstances and conditions affecting these changes. This is done by combining data mining and GIS techniques to uncover patterns and rules from the related data.

The goal is to analyze temporal and spatial land use/land cover change data as well as field investigation data, to infer a high likelihood of illegal land exploitation. Data mining in conjunction with GIS techniques are applied to gain geographic knowledge and identify meaningful patterns from multidimensional datasets.

The rest of this paper is structured as follows. We first present the methodological approach used in this study, in particular, the data collection method, operationalization of research variables, and analytical technique. This is followed by discussion of the results derived through spatial association rule mining, showing the high probability of illegal land usage derived through spatial association rule mining. In the final section the implications and conclusions are outlined.

2. METHOD

The approach described in this work is based on the CRISP-DM model (Cross-Industry Standard Process for Data Mining) which was developed by DaimlerChrysler, SPSS, and NCR in 1996. Larose (2005) pointed that the

CRISP-DM is a general problem-solving strategy divided into six phases for processing data mining projects, where the sequence of next phase often depends on the outcomes connected with the previous phase. An outline of each phase is given below (Kurgan and Musilek, 2006; Larose, 2005):

1. **Business understanding phase**: Focus on understanding project objectives and requirements for deciding on the adaptive data mining analytic framework, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.
2. **Data understanding phase**: Become familiarized with the data and verify data quality problems through initial data collection and basic statistics.
3. **Data preparation phase**: Table, record and attribute selection as well as transformation and cleaning of data in order to construct final dataset for modeling tasks.
4. **Modeling phase**: Choose modeling techniques and calibrate optimal values for essential parameters. Given the specific requirements of some techniques, it is often necessary to return back to the data preparation phase.
5. **Evaluation phase**: Evaluate the model and review the steps of model construction to determine whether all business objectives are reached.
6. **Deployment phase**: Organize and present the results of data mining, but, depending on the requirements, this phase can simply generate a report or repeatedly implement a data mining process across the enterprise.

In accordance with CRISP-DM, the iterative process of this research is illustrated in Figure 1. The goal of this study is to uncover interesting and actionable association rules to aid government administration in examining regulations for non-urban development and to strengthen land management. In an effort to extract important factors within various spatio-temporal datasets related to land use/land cover changes, we first review land management policies, laws, regulations, yearly statistical reports, administration strategies, and so on. Referring to historical government publications, most unlawful non-urban incidences of exploitation are related to facilities, industry or building construction around public transportation facilities, roads or riversides that affect non-urban land use planning within regional planning rules and regulations.

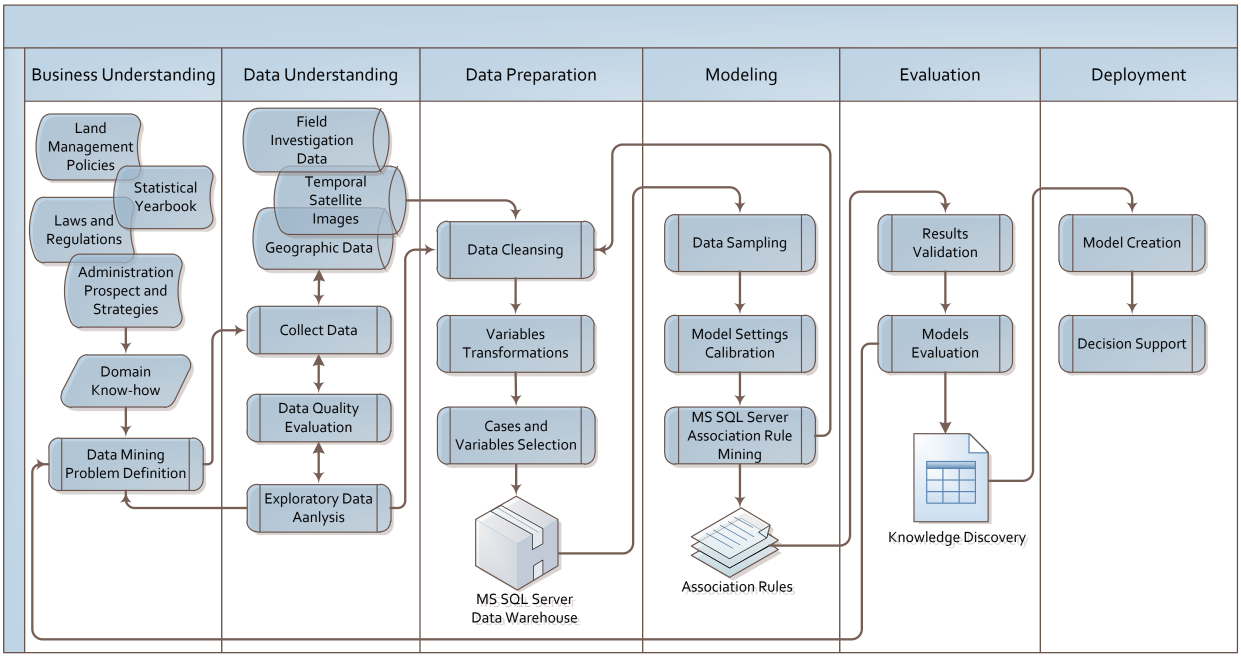

Figure 1 Research Framework

The data are provided by the Land Use Change Detection Program, and include field investigation data, temporal satellite images, geospatial information, and so on, from March 2002 to June 2011. The research approach is based on investigation results and related spatial distribution of land use/land cover change areas. Detailed data items in the analysis are given below:

1. Investigation reports.
2. Geographical features of land change areas.
3. Non-urban land-use zoning map collected from the Land Surveying and Mapping Center.
4. Average value based on 40 m grid DEM and slope for each land change area.
5. Measurement of the nearest distance from rivers, lakes, roadways, expressways, railways, buildings, industries to the central point of all land change areas.

6. Attributes of prior and subsequent satellite images near the issue date of land change areas.

To begin with the data cleansing process, while data that missing or incorrectly coded attribute values would be deleted by inspection quality checks, such as investigative records or land change areas with missing non-urban land-use zoning data are omitted. Next, use ArcGIS to analyze the geographic features regarding spatial relationships between the land change areas. Lastly, the final data are relatively transformed to a compatible format. To assure the effectiveness of spatiotemporal datasets for the data mining analyses, altogether 1,882 investigative records for 2010-2011 are extracted after the various screening procedures. Table 1 shows the survey results for change areas per chosen year.

Table 1 Land Change Investigation Statistics

Issued Date	Investigation Number	Verification of Change		
		Legal	Illegal	Illegal Rate
January, 2010	184	119	65	35%
May, 2010	173	129	44	25%
August, 2010	150	100	50	33%
November, 2010	218	134	84	39%
February, 2011	322	192	130	40%
June, 2011	835	629	206	25%
Total	1,882	1,303	579	31%

The MS SQL Server is utilized as the data warehouse for future data mining tasks. Association rule mining technology is then used to separate useful rules and patterns. Finally, the model is verified by analysts as meeting the research and to determine whether the deployment will take place.

3. RESULTS AND DISCUSSION

Spatial association rule mining is utilized to indicate spatial relationships between diverse potentially relevant properties given the geographical data (Kondaveeti et al., 2011). Therefore, this study utilizes geospatial data as a rich source of structure and pattern for data mining research. There are a number of possible factors related to land change areas that can be discovered from embedded models and information related to non-urban land exploration, as classified in the histogram as Figure 2. The classification distribution between land change areas and the weighted values of influence factors are discussed below:

1. The prior and subsequent satellite images of land change areas demonstrate that most changes are in bare lands.
2. The amount of legal field investigations shows a nearly twofold increase in illegal cases.
3. The majority of non-urban land-use zoning data indicates farming and grazing land in special agricultural or regular agricultural districts.
4. The DTM are overall below 210 m and the greatest slope is 4.5 degrees.
5. The nearest distance to building areas is usually less than 500 m, but the shortest distance to an industrial area is usually around 10,811 m.
6. In general, the closest distance along a straight line to a railway or a road is less than than 3,189 m or 242 m, respectively.
7. The greater part of land change areas are at a distance of lower than 100m from rivers; but the nearest distance to a lake is mainly spread from 139 m to 9,968 m.

Apriori algorithms are utilized in the process of spatial association rule mining conducted with the Analysis Services tool in the MS SQL Server. Figure 3 shows the highest Support value calculated by the Apriori algorithms. According to this diagram most land change areas are nearby roads on level ground. Furthermore, to satisfy the purpose of the spatial data mining project, the survey screens out 18 association rules and 20 item sets indicating illegal non-urban land development derived from large spatial datasets, individually presented in Figure 4 and Figure 5. The reliability and validity are considered, as in the diagram. All rules have a 100% Confidence level (named “Probability” in the MS SQL Server) and the general Support value of item sets surpasses 300. There is fairly general agreement that strong spatial association rules are produced by this spatial mining process.

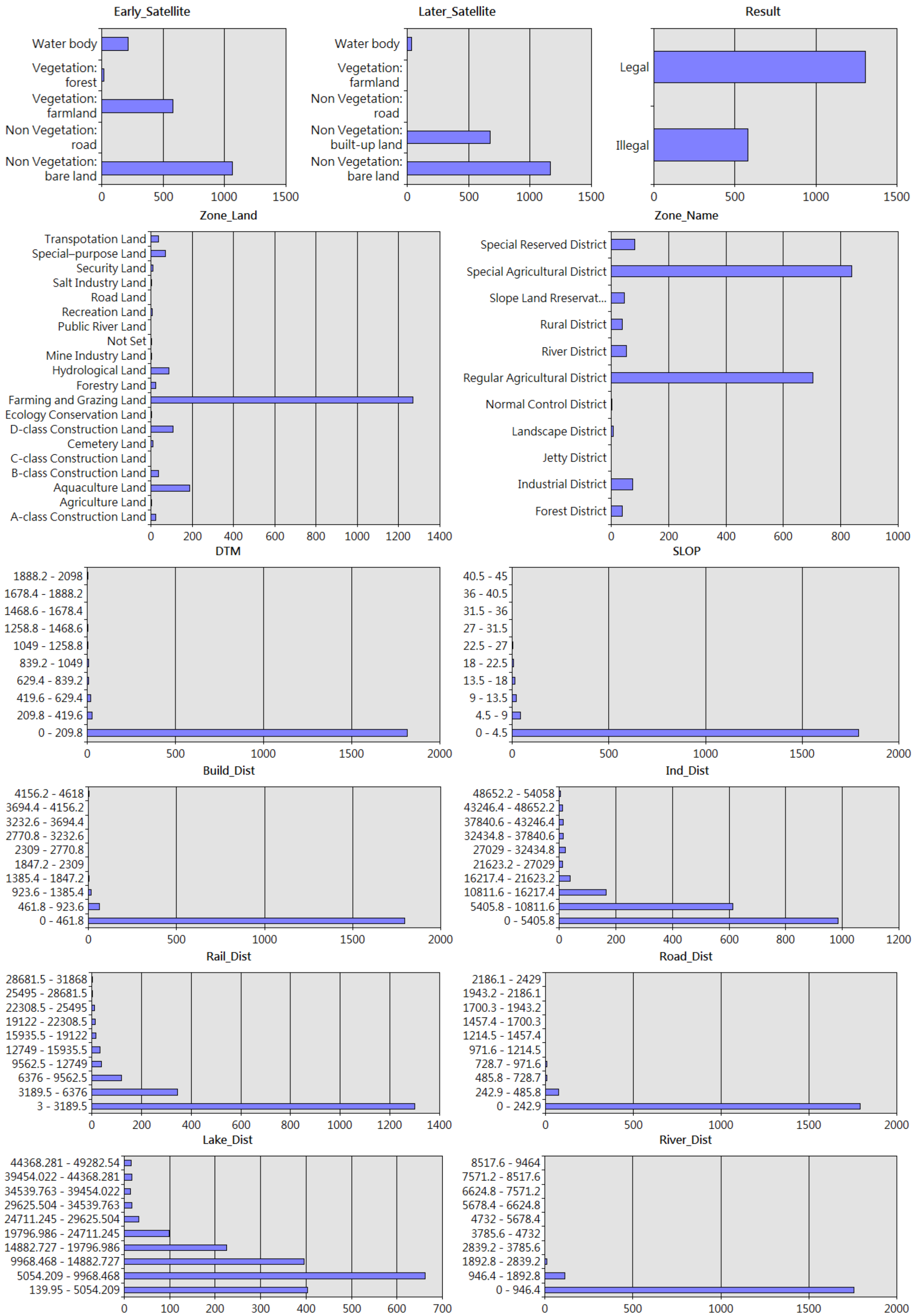

Figure 2 Classification of non-urban land change areas from spatial features and field investigations

ATTRIBUTE_NAME	ATTRIBUTE_VALUE	SUPPORT	PROBABILITY	VARIANCE	VALUETYPE
SLOP	< 1.0673950802	1586	0.842720510095643	0	5 (離散化)

ATTRIBUTE_NAME	ATTRIBUTE_VALUE	SUPPORT	PROBABILITY	VARIANCE	VALUETYPE
DTM	< 57.5954430208	1483	0.787991498405951	0	5 (離散化)

ATTRIBUTE_NAME	ATTRIBUTE_VALUE	SUPPORT	PROBABILITY	VARIANCE	VALUETYPE
Road Dist	< 82.6699991808	1390	0.738575982996812	0	5 (離散化)

Figure 3 Top 3 Support value computed by Apriori algorithms

Probability	Important	Rules
1.000	0.388	Zone Land = Transpotation Land, Lake Dist = 19390.8704149504 - 28692.7548481536 -> Result = Illegal
1.000	0.388	Zone Land = Security Land, Road Dist = 566.1166286848 - 2103.4999996416 -> Result = Illegal
1.000	0.388	Zone Land = Security Land, Early Satellite = Vegetation: forest -> Result = Illegal
1.000	0.388	Zone Land = Not Set, Road Dist < 82.6699991808 -> Result = Illegal
1.000	0.388	Zone Land = Not Set, Rail Dist = 6748.7967985664 - 13598.9134491648 -> Result = Illegal
1.000	0.388	Zone Land = Not Set, DTM = 621.6604964864 - 1009.0125713408 -> Result = Illegal
1.000	0.388	Zone Land = Forestry Land, Road Dist = 311.8417903104 - 566.1166286848 -> Result = Illegal
1.000	0.388	Zone Land = Forestry Land, River Dist = 169 - 296 -> Result = Illegal
1.000	0.388	SLOP = 13.7061730624 - 26.7992635136, River Dist = 169 - 296 -> Result = Illegal
1.000	0.388	SLOP = 13.7061730624 - 26.7992635136, Lake Dist >= 28692.7548481536 -> Result = Illegal
1.000	0.388	SLOP = 13.7061730624 - 26.7992635136, Ind Dist < 5907.23221504 -> Result = Illegal
1.000	0.388	Road Dist = 566.1166286848 - 2103.4999996416, SLOP = 13.7061730624 - 26.7992635136 -> Result = Illegal
1.000	0.388	Road Dist = 566.1166286848 - 2103.4999996416, Ind Dist = 12344.7657955328 - 28108.6967349248 -> Result = Illegal
1.000	0.388	Lake Dist >= 28692.7548481536, Rail Dist = 6748.7967985664 - 13598.9134491648 -> Result = Illegal
1.000	0.388	Ind Dist >= 40960.6415187968, Rail Dist = 6748.7967985664 - 13598.9134491648 -> Result = Illegal
1.000	0.435	DTM = 621.6604964864 - 1009.0125713408, Rail Dist = 6748.7967985664 - 13598.9134491648 -> Result = Illegal
1.000	0.388	DTM = 621.6604964864 - 1009.0125713408, Lake Dist >= 28692.7548481536 -> Result = Illegal
1.000	0.388	DTM = 621.6604964864 - 1009.0125713408, Early Satellite = Non Vegetation: bare land -> Result = Illegal

Figure 4 Association rules for illegal non-urban land development

Support	Item Size	Itemset
493	2	Result = Illegal, Zone Land = Farming and Grazing Land
480	2	Result = Illegal, SLOP < 1.0673950802
461	2	Result = Illegal, Road Dist < 82.6699991808
436	2	Result = Illegal, DTM < 57.5954430208
415	3	Result = Illegal, Zone Land = Farming and Grazing Land, SLOP < 1.0673950802
402	3	Result = Illegal, DTM < 57.5954430208, SLOP < 1.0673950802
396	3	Result = Illegal, Zone Land = Farming and Grazing Land, Road Dist < 82.6699991808
394	3	Result = Illegal, Road Dist < 82.6699991808, SLOP < 1.0673950802
374	3	Result = Illegal, Zone Land = Farming and Grazing Land, DTM < 57.5954430208
365	2	Result = Illegal, Early Satellite = Non Vegetation: bare land
361	2	Result = Illegal, Ind Dist < 5907.23221504
356	3	Result = Illegal, Road Dist < 82.6699991808, DTM < 57.5954430208
353	2	Result = Illegal, Zone Name = Special Agricultural District
330	2	Result = Illegal, Rail Dist < 2451.586179072
326	3	Result = Illegal, Zone Name = Special Agricultural District, Zone Land = Farming and Grazing ...
325	3	Result = Illegal, Early Satellite = Non Vegetation: bare land, SLOP < 1.0673950802
315	3	Result = Illegal, Early Satellite = Non Vegetation: bare land, Zone Land = Farming and Grazing...
315	3	Result = Illegal, Ind Dist < 5907.23221504, Zone Land = Farming and Grazing Land
314	3	Result = Illegal, Ind Dist < 5907.23221504, SLOP < 1.0673950802
309	3	Result = Illegal, Early Satellite = Non Vegetation: bare land, Road Dist < 82.6699991808

Figure 5 Item sets for illegal non-urban land development

Although various interesting findings may be detected from association rules, domain knowledge is needed to filter out trivial rules (Guo and Mennis, 2009). From these association rules, it is important that the land change areas without arranging land-use zoning attributes easily lead to unlawful incidents, in particular, illegal districts around roads but far from lakes and railways. Forested land is an exception where the illegitimate land use is near the rivers. As lawless exploitations may depend on the organization of non-urban land-use zoning; certainly, this is especially important for the relevant decision-making by land administration.

4. CONCLUSION

Spatial data mining is the notion of discovering previously unknown, interesting and highly probable useful patterns from abundant spatiotemporal datasets. Applying association rule mining to the land use/land cover change data can reveal additional important spatial relationships and help determine the relevance and importance of knowledge.

This study aims to separate latent correlations between geographic features and human behavior to protect against unlawful land development. The hope is that land administrators can use the analysis model to develop better land management strategies. This study incorporates GIS and data mining techniques to classify spatial closeness based on the land use/land cover changes and then process supervised spatial association rules to reveal potential reasons for illegal land exploitation. The results obtained in this research complement several implications for land strategies. For administrative governments, it is necessary to enhance control in high proportion of illegal areas which are level grounds in neighborhood of roads. Moreover, the land without setting land-use zoning attributes will also need to be conducted as soon as possible; otherwise, the resources of these districts will be destroyed with illegitimacy. To sum up, sound association rules are recommended in this survey for the reference of non-urban land-use zoning planning and instruction for land management.

Not surprising, there are some limitations needed to note again before assessing the findings of this study. Because analysis datasets is selected by domain experts, it is possible bias existed on knowledge and experience from different researchers. Similarly, rules are restricted to the quality and quantity of data, thus many interesting association rules are uncovered, along with many uninteresting rules. Further studies should test and refine these rules, and notions of strong and weak model effects. Ideally, researchers are required to acquire more perspective domain knowledge and sufficient database to bring the results more fruitful.

REFERENCES

Gómez, L., Kuijpers, B., Moelans, B. and Vaisman, A., 2011. A State-of-the-Art in Spatio-Temporal Data Warehousing, OLAP and Mining. Integrations of Data Warehousing. In: Data Mining and Database Technologies: Innovative Approaches, Information Science Reference, pp. 200-209.

Guo, D. and Mennis, J., 2009. Spatial data mining and geographic knowledge discovery—An introduction. In: Computers, Environment and Urban Systems, 33, pp. 403-408

Kondaveeti, A., Runger, G., Huan Liu and Rowe, J., 2011. Extracting Geographic Knowledge from Sensor Intervention Data Using Spatial Association Rules. In: Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on Digital Object Identifier, pp. 127 – 130.

Kurgan, L. and Musilek, P., 2006. A Survey of Knowledge Discovery and Data Mining Process Models. In: The Knowledge Engineering Review, 21:1, pp. 1-24.

Larose, D., 2009. Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley & Sons, Inc.

Zhu, X., Healey, R.G., and Aspinall, R.J., 1998. Knowledge-Based Systems Approach to Design of Spatial Decision Support Systems of Environment Management. Environmental Management, Cambridge University Press, 22 (1), pp. 35-48.