

Multidimensional and Data Mining Analysis for Property Investment Risk Analysis

Nur Atiqah Rochin Demong, Jie Lu, and Farookh Khadeer Hussain

Abstract—Property investment in the real estate industry has a high risk due to the uncertainty factors that will affect the decisions made and high cost. Analytic hierarchy process has existed for some time in which referred to an expert's opinion to measure the uncertainty of the risk factors for the risk analysis. Therefore, different level of experts' experiences will create different opinion and lead to the conflict among the experts in the field. The objective of this paper is to propose a new technique to measure the uncertainty of the risk factors based on multidimensional data model and data mining techniques as deterministic approach. The propose technique consist of a basic framework which includes four modules: user, technology, end-user access tools and applications. The property investment risk analysis defines as a micro level analysis as the features of the property will be considered in the analysis in this paper.

Keywords—Uncertainty factors, data mining, multidimensional data model, risk analysis.

I. INTRODUCTION

INVESTMENT in the real estate industry is a high risk, high return and a long cycle, which need the real estate investors to research it carefully to mitigate the risk [1]-[3]. Risk occurs at different stages of the investment process. The risk analysis is part of the process involve when making decision on which project that have a high priority for investment with the limited budget and time given. There are many techniques available for the risk analysis. The investor should consider a technique that provides comprehensive analysis with a high level of trust and satisfaction.

There are many factors that might influence the risk analysis of property investment. The dynamic changes of risk factors such as contract and policy adjustment impress a direct influence on the success of property investment in the real estate industry. Property investment in the real estate industry normally based on the investor's goals of investment and tailored to meet with their preferences and requirements. For example, there are three main goals for property investment either for rental, residential or commercial. The investor normally fully depending on the real estate agents who are experts in the field in giving them suggestions and helping them in making decisions.

Nur Atiqah Rochin Demong is with the QCIS, School of Software, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia (e-mail: ndemong@it.uts.edu.au).

Jie Lu is with the QCIS, School of Software, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia (e-mail: Jie.Lu@uts.edu.au).

Farookh Khadeer Hussain is with the QCIS, School of Software, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia (e-mail: Farookh.Hussain@uts.edu.au).

A popular technique for indexing the risk factors rank and weight in the real estate industry is known as analytical hierarchical processing (AHP) which takes into account the expert judgment for indexing. The AHP technique which applies the concept of the hierarchical data model is used to determine the weight of every index to deal with the uncertainty of the risk analysis of the real estate investment [1], [4]-[5]. AHP is used to evaluate the risk factors by combining the subjective judgment and the objective reasoning for the risk analysis in the real estate industry [3]. The process of achieving the results involves with expert opinion and different level of experience will give different judgment and leads to misinterpretation. Moreover, by referring to expert's opinion, limited information will be given and some information might be restricted because of the organizational policy that might be beneficial to investors. The experts would answer and give suggestion based on investor's requirements.

In order to solve this issue, this paper proposes the concept of a multidimensional data model and a data mining technique as a deterministic approach to rank and weight the risk factors. The use of a multidimensional model for the property investment risk analysis helps the investor to make more reasonable and scientific decision. The proposed technique helps the investor to analyze the data more systematically and provides a comprehensive analysis through a high dimensionality of data. This paper identifies the power and benefits of multidimensional analysis and data mining technique can bring to the practice of determining the rank and weight of the uncertain risk factors. Deterministic approach uses massive available data for ranking and weighting the uncertain risk factors. The rest of the paper organized as follows. Following this introduction, property investment risk analysis, multidimensional data model and intelligent decision support system concepts were reviewed in Section II. Section III elaborates on the use of a multidimensional model for property investment risk analysis in the real estate industry. Section IV explains the experiment and the results. In Section V, the conclusion and future works.

II. LITERATURE REVIEW

A. Property Investment Risk Analysis

The risk analysis concept has always been present in business transactions especially in the real estate industry which involve a high cost and high capital [6]. Risk analysis has proven its value in reducing risks in projects. Regardless of the type of risk analysis process, the risk management has a positive effect in finding and taking actions to avoid events

that could cause negative consequences for the project and the organization [7]. Risk analysis is a vital process for property investment in the real estate industry with a low liquidity and a high cost. Risk analysis consists of three stages: risk identification, risk estimation and risk assessment [8]. Risk analysis includes identification of risks, gathering information related to risk analysis or risk assessment and determine the risk level for each choice or alternatives given and last but not least deciding and select the best option for investment. This paper aims to measure the rank and the weight of the risk factors for the property investment risk analysis based on multidimensional analysis and data mining technique as deterministic approach.

Property investment risk analysis involves with micro level of analysis as detail features of the property would be analysed using a multidimensional analysis and a data mining technique. Some of the risk factors that might affect the property investment decision are as follows: location, minimum and maximum price, size of the property, number of bathroom, bedroom and car park, additional features such as a balcony, walk in wardrobe, sauna and zoning. The ultimate aim of any investor is to maximize his returns and minimize the risk [9]. Investors engage with the property market through the asset market and their determination of yields; the rates of return required to attract direct investment into property [9].

A set of factors that are specific to a local property market and affect both the risk premium and expectation of rental growth includes: the structure and performance of the underlying local economy, location, the quality level of local infrastructure and services, the balance between demand for and supply of business accommodation and local market stability and liquidity [10]. Hence, there is a need for multidimensional analysis in the property investment risk analysis which needed a high dimensionality of data and accurate information. Property investment risk analysis is depending on the type of investment which mainly divided into residential and commercial real estate. Each type has different sensitive degree to location, policy, macroeconomics, and so does the risk-withstand ability [1]. The deterministic approach which applies the concept of multidimensional data model and data mining technique working as the decision making tools and technologies to support the decision making process.

B. Decision Making Tools

Some of the examples of the decision making tools available are decision support system, data mining, Online Analytical Processing (OLAP) and group decision support system. The review of literature indicates that there is much work that need more attention in developing an intelligent decision support system (IDSS) for handling risk-based decision making in business operations or as a tool for businesses managing their business task especially when they deal with decision making processes in their daily routine as a manager. This is perhaps the most important concern for the future of an information system related to risk analysis or risk aggregation for managers who dealt with decision making processes, since it will promote vital and useful technology

that helps investors identify the risk involve when making certain decisions to meet the organization goals and objectives.

Uncertainty and complexity are becoming common facts leading to the greater recognition of systemic and holistic approaches to problem solving [11]. Managers need to integrate different types of intelligent information systems that are capable of supporting them throughout the decision-making lifecycle, which starts with structuring a problem from a given set of symptoms and ends with providing the information needed to make the decision [12]. A research on how to design, build and implement intelligent decision making support system (i-DMSS) from a more structured and software engineering/systems engineering perspective are still missing in 1980 – 2004 period [13]. The quality, speed and realization of the decision making could be increased when the right information is available to the right persons, at the right time, and in the right form [14].

C. Multidimensional Data Model

The multidimensional data model provides users with the flexibility to view data from different perspectives. This model is used in the data warehouse architecture which contains different types of end-user access tools such as data mining and online analytical processing (OLAP). OLAP is one of technologies that enable client applications to efficiently access the data multi-dimensionally. OLAP provides many benefits to analytical users, for example, an intuitive multi-dimensional data model makes it easy to select, navigate and explore the data, an analytical query language provides the power to explore complex business data relationships and pre-calculation of frequently queried data, enables very fast response time to ad-hoc queries [15].

OLAP systems are a popular business intelligence technique in the field of the enterprise information systems for business analysis and decision support [16]. OLAP not only integrates the management information systems (MIS), decision support systems (DSS) and executive information systems (EIS) functionality of the earlier generation of information systems, but goes further and introduces spreadsheet-like multi-dimensional data views and graphical presentation capabilities. The core component of an OLAP is the data warehouse, which is a decision-support database that is periodically updated by extracting, transforming, and loading data from several online transaction processing (OLTP) databases [17]. OLAP implementations typically employ a star schema (dimensional modelling), which stores data denormalized in fact tables and dimension tables. The fact table contains mappings to each dimension table, along with the actual measured data.

D. Knowledge Discovery and Data Mining

Knowledge discovery and data mining (KDD) have become areas of growing significance because of the recent increasing demand for KDD techniques, including those used in machine learning, databases, statistics, knowledge acquisition, data visualization, and high performance computing. KDD can be extremely beneficial in the field of artificial intelligence in

many areas, such as industry, commerce, government, and education. Information can be converted into knowledge about historical patterns and future trends.

Data mining allows users to analyse data from different dimensions or angles, categorize it, and summarize the relationships identified [18]-[22]. Data mining is used to transform knowledge from data format into some other human understandable format like rule, formula, and theorem. A data mining process is not only mining knowledge from data, but also from human [23]. Based on the knowledge discovered, the decision makers would match their requirements with measurement of risk factors generated by the decision support system to achieve their goals.

III. MULTIDIMENSIONAL ANALYSIS AND DATA MINING FRAMEWORK FOR PROPERTY INVESTMENT RISK ANALYSIS

The implications of studying multidimensional analysis and data mining for property investment risk analysis is to respond to the limitations of expert opinion and AHP that apply the hierarchical data model. By using a multidimensional analysis and a data mining, the results generated are accurate and based on structured data stored in the database or data warehouse. The main purpose of the proposed technique is to improve the accuracy of risk measurement and to avoid misinterpretation of risk factors rank and weight from expert judgments cause by a different level of expert's experience. The multidimensional analysis and data mining framework proposed includes the following four modules: (1) User module; (2) Information Technology module; (3) End-user access tools module; and (4) Business Application module. Fig. 1 shows the basic framework of multidimensional analysis and data mining framework for property investment risk analysis.

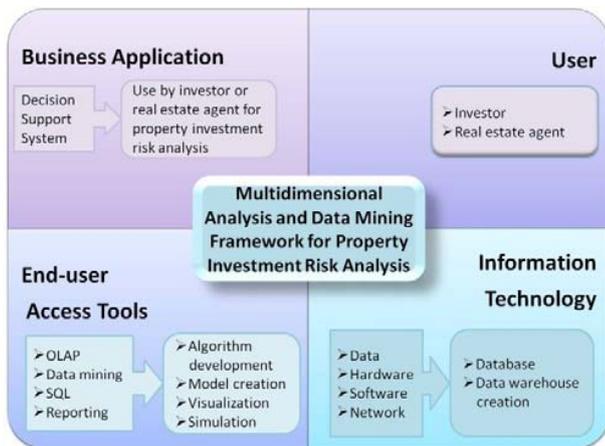


Fig. 1 The basic framework of multidimensional analysis and data mining technique for property investment risk analysis

A. Module 1: User

The user refers to the decision maker particularly the investor and real estate agent who are dealing with the property investment risk analysis. They will need to set their goals and requirements for the investment and will gain

knowledge based on the results generated by the multidimensional analysis and the hidden patterns of data based on data mining technique requested. The explicit knowledge of the decision support tools would be transferred to users as tacit knowledge for them to decide the best investment that aligned with their goals.

B. Module 2: Information Technology

This module mainly includes the data, hardware, software and network to create a database and data warehouse for the multidimensional analysis and data mining technique. The data as input gathered from different types of form such as a report, receipt, and business transaction data either from the user or customer, contractor, real estate agent, or investor as a query. The data collected for the analysis must valid and for the multidimensional analysis and data mining technique, it needs a high dimensional data. The hardware, software and network are the technologies used to keep the data, collecting, storing, processing, analyzing the data and providing information as knowledge of the decision maker. The data model applies in this paper is a multidimensional data model.

C. End-user Access Tools

The end-user access tools module refers to a set of data warehouse tools and technologies in which determine the success of data warehouse applications. End-user access tools for data warehouse application include OLAP, data mining, structured query language (SQL) and reporting. This module will become the interface for the user as in the module 1 to integrate with module 4 by using elements in the module 2. The user will develop an algorithm and create models for the property investment risk analysis. The output of the analysis from this module either displayed as visualization or simulation for the user to understand and analyze.

D. Module 3: Business Application

The business application refers to how this framework would be helpful and useful for business transactions. The real estate industry applies this technique for property investment risk analysis. The type of information system pertaining to the proposed technique is the decision support system (DSS) which helps the investor to achieve their goals. For the real estate industry, the DSS would be useful for managerial level to achieve organization targets and gaining competitive advantage. The process of multidimensional analysis and data mining technique for property investment risk analysis includes input collection as a query from a user, data processing using information technology and end-user access tools and results generated displayed to users for them to make a decision.

IV. EXPERIMENTS AND RESULTS

Two results of the data exploration analysis and two data mining techniques namely association and clustering techniques of data mining were chosen for experiment and results. The sample data used for the experiments were gathered from Australian Property Monitor domain database

and stored in the develop prototype system as shown in Table I. About 138 attributes were chosen and stored in the prototype system developed for the analysis. For data mining techniques, the Microsoft SQL Server 2008 integrated with Microsoft SQL Data Mining Add-ins were used to analyze the data and provide the output. The data shown in Table I refers to the price history for selected property for the analysis.

TABLE I
SOME OF THE SAMPLE DATA USED FOR THE ANALYSIS

Date	Price	Type of Sale	Property Type	Sale Result
1984-05-14	152000	Private Treaty	House	Sold
1988-08-09	2250000	Auction	Land	Highest Bid
1988-11-17	220000	Auction	House	Sold
1989-02-02	270000	Private Treaty	Townhouse	Sold
1989-04-24	255000	Private Treaty	House	Sold
1991-09-02	120000	Private Treaty	House	Sold
1993-04-17	300000	Private Treaty	Townhouse	Sold
1993-07-01	268000	Private Treaty	House	Sold
1993-10-07	176000	Private Treaty	House	Sold
1993-12-02	1240000	Private Treaty	House	Sold
1993-12-10	300000	Private Treaty	Cottage	Sold
1994-03-26	383000	Auction	Cottage	Sold
1994-04-29	420000	Private Treaty	House	Sold
1994-11-12	290000	Private Treaty	House	Sold (LA)
1995-06-20	450000	Private Treaty	House	Sold
1996-03-27	2650000	Auction	House	Highest Bid
1996-05-17	295000	Private Treaty	House	Sold
1996-12-21	310000	Auction	House	Sold
1997-02-22	125000	Auction	Unit	Highest Bid
.
.
.

of property available in the market conquered by private treaty. This output shows that the price of property is negotiable between sellers and buyers in selling the property.

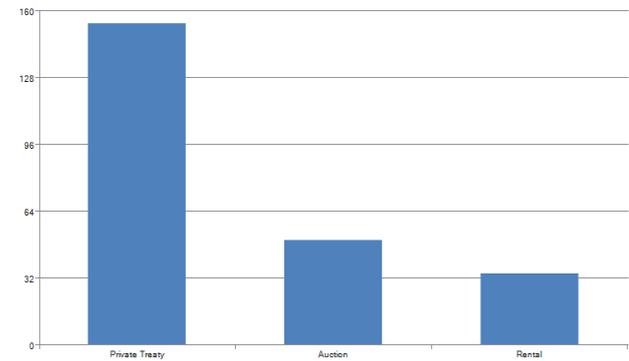


Fig. 3 Graph showing the type of sale of property available in the market for the selected sample data

B. Association Rule Mining

Association rule mining finds interesting associations and correlation relationship between a large set of data items. Association rule shows attribute value conditions that occur frequently together in a given dataset [24]. As shown in Fig. 4, the price of property was associated with other attributes such type of sale, sale result, year of selling and property type based on selected sample data.

A. Multidimensional Analysis

For multidimensional analysis, it provides historical value of house price or rental rate for property in different suburbs based on sample data extracted. Based on some sample data shown in Table I, 238 rows of data selected for the data exploration. Fig. 2 and Fig. 3 show the results based on sample data collected.

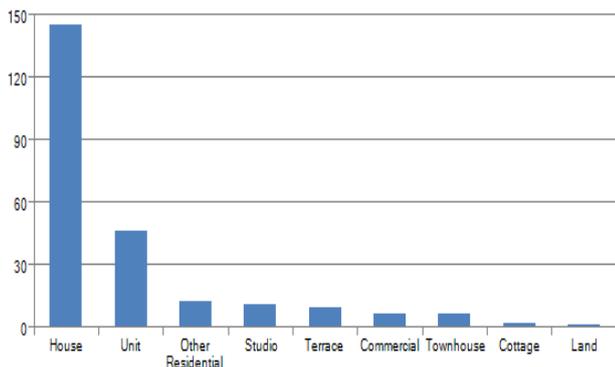


Fig. 2 Graph showing the type of property available in the market for the selected sample data

Based on Fig. 2, the type of property available in the market was dominated by house followed by a unit, other residential, studio and terrace. This will give an idea or knowledge for investors to choose the best property type for investment which meets with their goals. Similar to Fig. 3, the type of sale

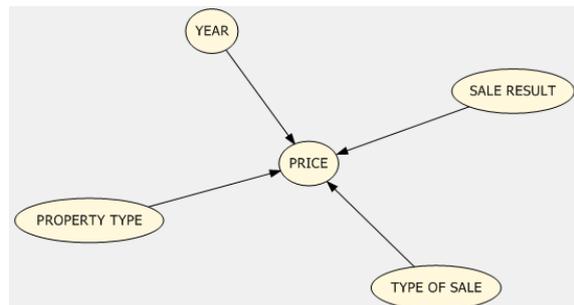


Fig. 4 The dependency network of property prices for the selected sample data

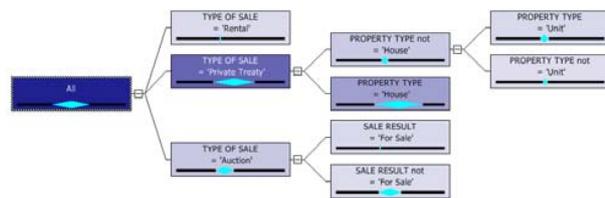


Fig. 5 The decision tree of the factors contributing to prices of property for the selected sample data

Fig. 5 shows the decision tree of the factors contributing to prices of property for the selected sample data. This information would give knowledge to investors in measuring the risk factors that will affect their decision making.

C. Clustering Technique

The clustering technique group data object into clusters. Clustering technique is good for overview data which contains many groups and unusual similarity measures are needed. Fig. 6 depicts the number of clusters generated using clustering techniques based on selected sample data. The result shows that cluster 1 dominated the populations with 167 cases followed by cluster 4, cluster 2, cluster 6, cluster 5, cluster 7 and cluster 3 respectively as shown in Fig. 7.

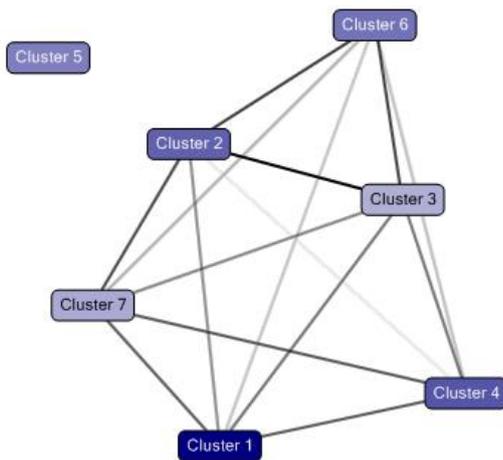


Fig. 6 Number of clusters generated using a clustering technique based on selected sample data

Variables	States	Population (All)	Cluster 1	Cluster 4	Cluster 2	Cluster 6	Cluster 5	Cluster 7	Cluster 3
Size		167	42	28	26	23	21	14	13
PRICE	Mean	822,037.00	881,838.66	0	303,029.31	651,491.03	527.38	4,904,574.75	320,974.19
PRICE	Deviation	1,752,979.48	444,300.58		244,351.38	838,226.30	261.58	3,805,623.71	454,909.05
PROPERTY TYPE	House	103	62%	100%	37%	76%	24%	86%	58%
PROPERTY TYPE	Unit	31	31%	1%	20%	5%	19%	14%	26%
PROPERTY TYPE	Other Residential	11	0%	0%	38%	0%	0%	0%	9%
PROPERTY TYPE	Studio	7	0%	0%	4%	0%	29%	0%	0%
PROPERTY TYPE	Terrace	6	0%	0%	0%	1%	19%	0%	8%
PROPERTY TYPE	Townhouse	5	7%	0%	1%	8%	0%	0%	0%
PROPERTY TYPE	Cottage	2	0%	0%	0%	10%	0%	0%	0%
PROPERTY TYPE	Commercial	2	0%	0%	0%	0%	10%	0%	0%
PROPERTY TYPE	...								
SALE RESULT	For Sale	109	93%	86%	78%	5%	0%	90%	79%
SALE RESULT	Sold	26	5%	5%	21%	77%	0%	2%	7%
SALE RESULT	For Rent	21	0%	0%	0%	0%	100%	0%	0%
SALE RESULT	Highest Bid	6	0%	0%	0%	11%	0%	8%	11%
SALE RESULT	Passed In	2	0%	5%	0%	2%	0%	0%	3%
SALE RESULT	Withdrawn	1	0%	5%	0%	0%	0%	0%	0%
SALE RESULT	Sold Prior	1	0%	0%	0%	1%	0%	0%	1%
SALE RESULT	Sold (IA)	1	0%	0%	0%	4%	0%	0%	0%
SALE RESULT	...								
TYPE OF SALE	Private Treaty	111	98%	43%	96%	52%	0%	92%	57%
TYPE OF SALE	Auction	35	2%	57%	5%	48%	0%	8%	43%
TYPE OF SALE	Rental	21	0%	0%	0%	0%	100%	0%	0%
YEAR	Mean	2,006.00	2,009.94	2,010.21	2,002.85	1,997.45	2,008.52	2,007.73	2,002.72
YEAR	Deviation	5.31	0.58	0.67	3.12	6.22	1.69	2.91	3.71

Fig. 7 Cluster profiles based on Fig. 6

Based on the experiment and the results shown, the time dimension in the multidimensional analysis is compulsory for the investment risk analysis especially for predictive technique for forecasting. The investor or real estate agent should utilize all related information extracted as knowledge. This could be a guideline for them to measure the risk factor for the property investment risk analysis that meets with their requirements. The use of the multidimensional analysis and data mining technique helps to measure the risk factors for property investment risk analysis more useful, accurate and reliable. The evaluation of the proposed technique was supported by the experiments and the results shown which discover hidden patterns and would be useful for the analysis and thus helps the investors in making critical business decision.

V. CONCLUSION AND FUTURE WORKS

Property investment risk analysis is a vital process that needs more attention by individuals as investors and real estate agents. The use of the multidimensional analysis and data mining techniques to measure the risk factors of property investment through a decision support system helps the investors to avoid or reduce risk. The investors can discover more knowledge about the features of available property for investment by applying this technique. Reflecting on the limitations and disadvantage of expert opinion to measure the risk factors for investment, the use of the multidimensional analysis and the data mining technique through deterministic approach helps the investors to achieve the best result that meets with their requirements. The historical data kept using multidimensional data model helps to provide forecasting analysis and aggregate data in providing foundations for the investment decision. The proposed technique is helpful in measuring the risk factor of the property investment in the real estate industry. Even though the cost is higher to implement the multidimensional analysis and data mining technique, the results generated would be more accurate and useful for investment risk analysis. A prototype decision support system that applies the concept of multidimensional analysis and data mining technique for property investment risk analysis will be developed for future works.

REFERENCES

- [1] X. L. Ye, "Risk Analysis in the Process of Real Estate Enterprise Project Investment," *2011 Fourth International Joint Conference on Computational Sciences and Optimization*, 2011, pp. 731–736.
- [2] L. W. Liu, E. Zhao and Y. Liu "Research into the Risk Analysis and Decision-Making of Real Estate Projects," *International Conference on Wireless Communications, Networking and Mobile Computing*, 2007, pp. 4610–4613.
- [3] S.C. Li, and Y. M. Yang, "The Research on Real Estate Project Risk Evaluation Based on Monte Carlo Simulation and the Theory of Variable Weight," *International Conference on Risk Management & Engineering Management*, 2008, pp. 449–454.
- [4] Y. Li and J. Suo, "Model on Risk Evaluation of Real Estate Investment," *6th International Conference on Fuzzy Systems and Knowledge Discovery*, 2009, vol. 3, pp. 138–140.
- [5] M. X. Yu, and L. Zhang, "Value Creation Based Cost Management of Real Estate Project," *4th International Conference on Wireless Communications, Networking and Mobile Computing*, 2008, pp. 1–4.
- [6] C. Wu, Y. Guo, and D. Wang, "Study on capital risk assessment model of Real Estate enterprises based on support vector machines and fuzzy integral," *Control and Decision Conference*, 2008, pp. 2317–2320.
- [7] R. Olsson, "In search of opportunity management: Is the risk management process enough?," *International Journal of Project Management*, 2007, vol. 25, no. 8, pp. 745–752.
- [8] J. Yu, and H. Xuan, "Study of a Practical Method for Real Estate Investment Risk Decision Making," *2010 International Conference on Management and Service Science*, 2010, pp. 1–4.
- [9] G. Y. Wang, J. Hu, Q. H. Zhang, X. Q. Liu, and J. Q. Zhou, "Granular computing based data mining in the views of rough set and fuzzy set," *IEEE International Conference on Granular Computing*, 2008, pp. 67, 26–28.
- [10] J. Henneberry, "Regional Variations in Office Property Investment and Development," *Regions Magazine*, 2010, vol. 278, No. 1, pp. 6–10.
- [11] C. C. Huang, T. L. Tseng, M. Z. Li, and R.R. Gung, "Models of multi-dimensional analysis for qualitative data and its application," *European Journal of Operational Research*, 2006, vol. 174, no. 2, pp. 983–1008.
- [12] S. Tong, and S. Z. Ji, "Demonstration of Book Issue Chain Decision Support and Multi-dimensional Data Model in E-commerce Era," *WRI World Congress on Software Engineering*, 2009, vol. 2, pp. 68–71.

- [13] E.A.M. Caron, and H.A.M. Daniels, "Explanation of exceptional values in multi-dimensional business databases," *European Journal of Operational Research*, vol. 188, no. 3, pp. 884–897, 2008.
- [14] M. Mora, G. Forgionne, J. N. D. Gupta, L. Garrido, F. Cervantes, and O. Gelman, "A strategic descriptive review of the intelligent decision-making support systems research: the 1980–2004 Period," *Book Chapters in Intelligent Decision-making Support Systems Foundations, Applications and Challenges*, Part III, 2007, pp. 441–462.
- [15] S. L. Rui, Y. W. Qi, H. H. Jian, G. Li, and L. Liang, "An intelligent decision support system applied to the investment of real estate," *Proceedings of The IEEE International Conference on Industrial Technology, 1996. (ICIT '96)*, 2007, pp. 801–805.
- [16] N. Casaigne, and L. Lorimier, "A challenging future for i-DMSS," *Book Chapters in Intelligent Decision-making Support Systems Foundations, Applications and Challenges*, Part III, 2007, pp. 401–422.
- [17] N. Karacapidilis, "An overview of future challenges of decision support technologies," *Industrial management and information systems lab, MEAD, University of Patras, Greece*, 2006, pp. 387–99.
- [18] L. Qi, "Advancing Knowledge Discovery and Data Mining," *First International Workshop on Knowledge Discovery and Data Mining*, 2008, pp. 3–5.
- [19] M. Lobur, Y. Stekh, A. Kernysky, and F. M. E. Sardieh, "Some trends in Knowledge Discovery and Data Mining," *International Conference on Perspective Technologies and Methods in MEMS Design*, 2008, pp. 95–97.
- [20] M.T. Kecahdi, and I. K. Savvas, "Cooperative Knowledge Discovery & Data Mining CKDD," *2010 19th IEEE International Workshop on Enabling Technologies: Infrastructures for Collaborative Enterprises*, 2010, pp. 96–97.
- [21] T. Horeis, and B. Sick, "Collaborative Knowledge Discovery & Data Mining: From Knowledge to Experience," *IEEE Symposium on Computational Intelligence and Data Mining*, 2007, pp. 421–428.
- [22] C. Roberts, S. Rowley, and J. Henneberry, "The impact of landscape quality on property investment decisions," *Journal of Property Investment & Finance*, vol. 30, No. 1, pp. 69–82, 2012.
- [23] X. L. Ye, "Risk Analysis in the Process of Real Estate Enterprise Project Investment," *2011 Fourth International Joint Conference on Computational Sciences and Optimization*, pp. 731–736, Apr 2011.
- [24] B. H. He, and G. F. Song, "Knowledge Management and Data Mining for Supply Chain Risk Management," *International Conference on Management and Service Science*, pp.1–4, Sept. 2009.