BIM-GIS ORIENTED INTELLIGENT KNOWLEDGE DISCOVERY

H. Kiavarz[1], M. Jadidi [1],A. Rajabifard[2], G. Sohn[1]

[1] Geomatics Engineering, Department of Earth & Space Science & Engineering, York University, Toronto, Canada - (hkiavarz, mjadidi, gsohn)@yorku.ca
[2]Departmen of Infrastructure Engineering, Melbourne School of Engineering, University of Melbourne, Melbourne, Australia – abbas.r@unimelb.edu.au

Commission VI, WG VI/4

KEY WORDS: Smart City, Knowledge discovery, BIM, GIS, Fuzzy Reasoning, Rules Extraction

ABSTRACT:

Urban and population growth results in increasing pressure on the public utilities like transport, energy, healthcare services, crime management and emergency services in the realm of smart city management. Smart management of these services increases the necessity of dealing with big data which is come from different sources with various types and formats like 3D city information, GPS, traffic, mobile, Building Information Model (BIM), environmental, social activities and IoT stream data. Therefore, an approach to mine/analysis/interpret these data and extract useful knowledge from this diverse big data sources emerges in order to extract the hidden pattern of data using computational algorithms from statistics, machine learning and information theory. However, inconsistency, duplication and repetition and misconducting with the different type of discrete and continuous data can cause erroneous decision-making. This paper focuses on providing a rules extraction and supervised-decision making methods for facilitating the fusion of BIM and 2D and 3D GIS-based information coupling with IoT stream data residing in a spatial database and 3D BIM data. The proposed methods can be used in those applications like Emergency Response, Evacuation Planning, Occupancy Mapping, and Urban Monitoring to Smart Multi-Buildings so that their input data mostly come from 2D and 3D GIS, BIM and IoT stream. This research focus on proposing the unified rules extraction and decision engine to help smart citizens and managers using BIM and GIS data to make smart decision rather than focus on applications in certain field of BIM and GIS.

1. INTRODUCTION

The population of cities is rising which is going to make the cities as megacities. The research shows there are 31 megacities in 2016 with more than 10 million inhabitants, which increase to 41 in 2030 (United Nations Department of Economic and Social Affairs Population Division, 2017). Demands of citizens in these cities will be increased tremendously such as new residential building, new commercial area, energy consuming, public transportation, and security (Suakanto et al., 2013). These days with the emerge of new technologies, the citizens share and produce a lot of information and location based data in social media, Internet and transactions, mobile, GPS in cars and cell phones, connected watches and sensors (Abbad and Bouchaib, 2017). Also, the modelling of reality is improved with the new 2D and 3D modelling to store and visualize the urban environments data in GIS modelling and building information in Building Information Modelling (BIM). These data carry significant information, we can have significant insights with exploiting data effectively, which helps us to explain, diagnose, predict and prescript certain phenomena. The most current researches in the realm of GIS and BIM shows that the integrated application of BIM and GIS engage in two types. One type contemplates on application in certain fields for certain phases and the other type proposes a unified model for comprehensive management for the whole lifecycle (Ma and Ren, 2017). However, researches of the first type usually use different integration patterns and platforms while researches of the second type still lack enough practice (Ma and Ren, 2017).

Data mining is a suitable solution to help decision makers recognize the rules from large databases (Moghaddam and Wang, 2014). A rule is semantic way of representing a pattern in the database by combination of words with If Situation Then conclusion (Iyer et al., 2018). Through this paper, we focused on mining 3D GIS, BIM data and their related metadata for a multi-building like campus area. After data mining we need a decision engine to use the extruded knowledge from data mining and leverage the prediction method to say what happened in the future. A unified and novel approach is proposed decision making as predictive method upon mentioned data. Briefly, we propose using decision tree as a rules extraction because of its performance of dealing with big data and the meaningful rules as output of decision tree. However, there are some challenges when using the traditional decision tree methods. First, the multi-building datasets usually have a large number of attributes, such as individual building structural information, building specifications, weather information, environmental attributes and network geometry between building and 3D GIS information. Traditional decision tree methods are not good at handling datasets with large number of attributes because large number of branches with the duplication and repetition of sub-trees will be generated. Secondly, dataset usually contains both discrete and continuous attributes. For example, the weather condition in the time of fire event is a discrete attribute, and the measures of the distance between tenant and building emergency exit door are continuous attributes. The conditional entropy in the traditional decision tree algorithms considers only discrete attributes. Therefore, discretization is generally applied on continuous variables before applying traditional decision tree methods. However, discretization of continuous values in the dataset increases the uncertainty of classification and reduces the accuracy of final decision class. The third problem is related to select the proper attributes to splitting a node. In traditional classification methods, an attribute is chosen solely based on information about this node, which creates redundancy in the decision tree. Last, datasets include inconsistent data. For example, two instances in the dataset with same attributes may fall into different decision classes, which make confusion in the traditional decision tree methods. To solve these issue three

solutions are proposed. First, a fuzzy feature selection method is proposed to apply on historical dataset. It selects and reduces the attributes from the environmental; geometry and 3D spatial features and BIM to improve the performance of decision tree and help handle large number of features. Secondly, a fuzzy decision tree is proposed to handle continuous values in the dataset Thirdly, to avoid choosing split attribute solely based on the information of one node, in fuzzy decision tree, the object set is replaced by a node and, by this replacement the proper node is selected by considering the information in an object set. Beyond the fuzzy decision tree, we have applied the inference process of fuzzy reasoning system to use fuzzy membership functions input and the rules that are extracted from fuzzy decision tree.

The rest of the paper is organized as follows: The proposed methodology is presented in Section 2. Section 3 describes the expected results result and Section 4 concludes the paper and future study.

2.   METODOLOGY

In this section, we present the overall view of suggested methodology to cover the conceptual model of methodology. The methodology introduced the supervised decision engine for different application of BIM-GIS in a multi-building environment.

2.1   Data Storage Model

To have a descriptive, diagnostic and predictive model for smart managing of a multi-building like a campus, a big data platform is necessary. In the framework of smart city, the big data can be stored in a data lake and conventional database. The proposed method for data storage is to store the historical data that are structured data in database and use data lake as enterprise repository.

2.1.1   Data Lake for a Multi Smart Building Environment

The concept of a data lake become prominent in smart city framework as a popular technologically to build the next generation repository of new big data challenges (Fang, 2015). Cities with the big data of 3D GIS, BIM and sensor based data are seeking to create data lakes because they manage and use data with huge volume, variety, and a velocity rarely seen in the past (Fang, 2015). The Lake data can include unstructured data like any reports, building specifications, any documents or email related to buildings or the environment around the building (Khine and Wang, 2018). Also, it can be included semi-structured data such as BIM in IFC format, 3D GIS in CityGML format and sensors stream data in different types of XML CSV files or log files. These types of data can be transferred to structured data for tasks like machine learning, analytics and visualizations. Figure 1 illustrates the infographic of data lake.
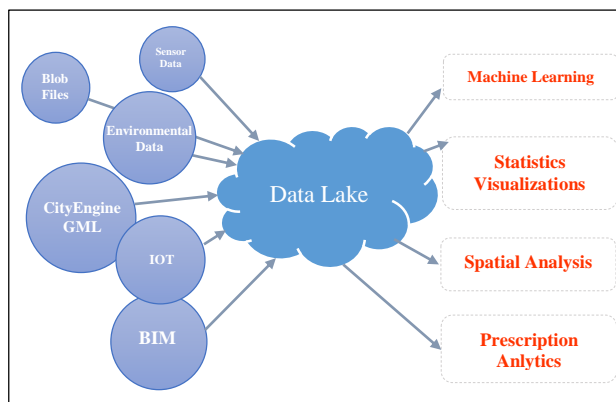

Figure 1. Data Lake Infographic

2.1.2   Database for Structured Data

The historical data as structured data is stored in a relational database such as multi-building historical fire drill data which contains both discrete and continuous attributes. We call this dataset as an information table as the rows in this tables demonstrates the different scenarios in historical data and columns shows the important criteria (features) of a specific application. One column as a decision feature shows the decision class or label of each row The information table is divided into two part as training part which be involved in rule extraction algorithm as the proposed algorithm is leaned based on training part of information table and testing part which is input data for designed algorithm to measure the accuracy of prediction and decision making. The Table 1 shows a sample of information table in the database.

Object	Age	Distance to Exit (m)	Floor	Age of Building	Disability	Decision
O_1	32	1	2	10	NO	Use Exit Stairs
O_2	52	1	4	15	Yes	Saviour Help
O_3	42	3	2	20	NO	Use Exit Stairs
O_4	50	2.5	2	20	NO	Use Exit Stairs
O_5	11	1	1	25	NO	Use Exit Stairs
O_6	30	2	1	12	NO	Saviour Help
O_7	50	1	37	40	NO	Saviour Help

Table 1. The information table of fire evacuation plan

In the Table 1, the columns of age, Distance to Exit door, Floor number, Age of Building and Disability are the features (attributes) and the Decision column is the decision class or label of each row.

Most of historical data includes inconstant data. The inconsistency in the dataset confuses decision engine method and predictions can cause erroneous decision-making. Row 1 and row 4 in the Table 1 shows inconstancy so that these rows have fairly same values in the features but represent two different decision classes. Feature selection is a technique for selecting the attribute of feature set as an important component of both supervised and unsupervised classification (Janecek et al., 2008). Fuzzy Roughest Feature Selection (FRFS) is proposed in the next section to deal with inconsistence data and select those features to reduce inconstancy in the database.

2.2   Fuzzy Rough Set Feature Selection for BIM-GIS Historical Inconsistent Data

The main objective of feature selection in this study is three-folds: preparing the BIM-GIS applications information table with the minimum relevancy and redundancy, providing faster and efficient rules extraction decision tree generation, and improving the prediction performance of the prediction via predictor method. High dimensionality and inconstancy in BIM-GIS multi-building datasets generate the decision tree with a large number of nodes, redundancy and low performance. These problems restrict the usability of decision trees in the rule generation and classification. This method prompted our research into the use of fuzzy rough sets for feature selection.

As this research focuses on multi-proposed smart application in the realm of BIM and GIS so the smart application has to deal with big dataset, which includes inconsistency. Data inconsistency in the dataset confuses the decision-making methods. When a prediction method, which is the combination of fuzzy decision tree and fuzzy madman method (Nayak and Devulapalli, 2016) in this study, is studied, wrong predictions can sometimes be observed when inconsistent data are present.

For example, two instances in the dataset with same attribute values may fall into different decision classes, which may result in the wrong decision. Incorrect predictions can cause erroneous decision-making. GIS and BIM datasets usually have a large number of attributes, such as building structural, mechanical and electrical geometry and specifications, temporal information, environmental attributes and weather information. Not every attribute contributes to a decision-making. Traditional rules extraction methods are not efficient at handling datasets with large numbers of attributes, because many branches are generated with duplication and repetition of sub-trees. FRFS reduce inconsistency in the dataset by selecting the proper attributes with the high dependency to the decision class. Also, by fuzzification of values, it can deal with both continues and descript values.

FRFS identify the dependencies between conditional attributes. Intuitively, a set of attributes DF depends totally on a set of conditional attributes $P \in CF$, if all attribute values from DF are uniquely arranged by the values of attributes from P. If there exists a functional dependency between values of DF and P, then DF depends totally on P. In fuzzy rough set theory, dependency is defined in the following way (Moghaddam, 2015):

$$\gamma_p(DF) = \sum_{X \in U} \mu_{POS_P(DF)}(X) \qquad (1)$$

The dependency value is between 0 and 1. The higher value of dependency for a conditional feature indicates the more significant of the feature. If the significance is 0, then the conditional feature is dispensable. The next section describes the method of feature selection based on dependency values.

2.3 Fuzzy Decision Tree for Multi-Building Rule Extraction

BIM-GIS dataset contain both discrete and continuous attributes. For example, the weather condition at the time of fire is a discrete attribute; and, the measures of building structure, the length of stairs, the age of buildings, the speed of wind, the distance between buildings, the distance between building and major roads, are usually continuous attributes. The traditional rules extraction methods consider only discrete attributes. Therefore, discretization is generally applied to continuous variables before applying traditional methods. However, discretization of continuous values in a dataset increases the uncertainty of prediction and reduces the accuracy of the final decision-making and recommendation to smart citizens. The fuzzy decision trees help to apply the real value of continuous values in the datasets. To overcome the over-fitting problem in decision tree methods, the FDT chooses an attribute-value in favour of all nodes at the same level when splitting a node. It can handle both discrete and continuous attributes.
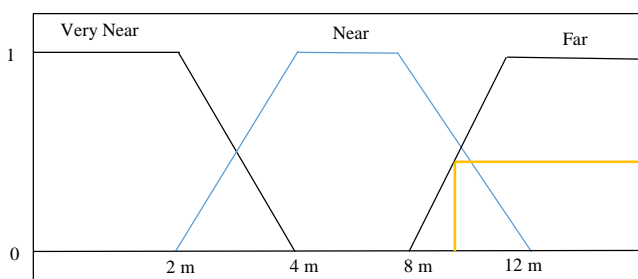

Figure 2. Fuzzification of the input value

Then, the fuzzy entropy is calculated for each record in the dataset. The records in the dataset encompass 2D value like

distance between buildings, 3D data value like the height of building structure or elevation of tenant location and environmental information like weather condition. Next, according to the calculated fuzzy and generality and redundancy criteria, the fuzzy decision tree is constructed. The last step expresses the decision method of the final classification which is done by training and testing data by fuzzy rules based system (Moghaddam, 2015).

$$Fuzzy\ Entropy(a = v) = -\sum_{i=1}^{c} \frac{\sum_{j}^{N} \mu_{ij}}{S} log_2 \frac{\sum_{j}^{N} \mu_{ij}}{S} \qquad (2)$$

where μ_{ij} is the membership value of the jth object to the ith class. This equation is defined based on the concept of applying the value of the membership function (i.e. membership value) of each feature of structured data, semi structured and unstructured data which are transferred to database by pre-processing procedure in the database rather than using the discrete values.

After creating the fuzzy decision tree, the rules can be extracted from it. The rules that are extracted from fuzzy decision tree have description structure based on IF-THEN phrases called linguistic rules. Each element in rules can come from 2D, 3D, building specification or environmental information. The approach to extract the rules is to follow a path through the tree to one of the leaves. These rule statements start from the root of the tree to a leaf and establish conditions, in terms of specifying the final class. These rules are generated from FDC whose antecedents and consequents are compound of fuzzy statements, related by the fuzzy implication and the compositional rule concepts (Roisenberg et al., 2009). If a set of conditions is satisfied, then set of consequents can be driven. The consequent is an output decision classes. It can be denoted as:

R_j: IF $(a_{j1}\ IS\ v_{j1})$ AND ... AND $(a_{jm}\ IS\ v_{jn})$ THEN $(b\ IS\ C_j)$ with j=1 to L which is the number of extracted rules, and with a_{j1} to a_{jm} and b being the input and output variables, with v_{j1} to v_{jn} and C_j being the involved antecedents and consequent labels, respectively.

The extracted rules from FDT are used to determine the final decision of checking data. Next section describes how the fuzzy reasoning method as a decision engine is applied on extracted rules to make a final decision for different BIM-GIS applications.

2.4 Fuzzy Reasoning as a Decision Engine in 3D Spaces

Since in a fuzzy reasoning, the decisions are based on considering all of the rules, the rules must be combined in some manner in order to make a decision. The fuzzy reasoning is proposed to be applied for decision engine or prediction model to determine the final decision class BIM-GIS applications based on the extracted rules on the testing data. This procedure includes five steps. At the fuzzification step, the crisp data is converted into a value of the corresponding fuzzy membership degree. This step is applied for all conditional and decision attributes of rules. A value between 0 and 1 is assigned for each attribute in the generated rules. After fuzzifying the antecedent part of rules, we obtain the fuzzy membership value for each attribute of rules. The next step is the rule evaluation, which is to join the different attribute-value of the rule by a disjunction function to obtain one membership value for each rule. The third step is the implication. As each rules includes different attribute-value parts so a fuzzy implication operator is applied on the membership function to obtain a new concluded fuzzy set for each rule to determine the effect of this rules on decision membership function. The output of this step is a fuzzy set of decision attribute. The fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set by aggregation

process. Perhaps the most popular defuzzification method is the centre of gravity, which returns the centre of area under the curve. There are five built-in methods supported: centroid, bisector, middle of maximum (the average of the maximum value of the output set), largest of maximum, and smallest of maximum (Pradhan, 2013).

It is called aggregation step in which the fuzzy operator "OR" is used to aggregate all values in the output membership function which are extracted from Step 3. The last step is defuzzification. To determine a certain severity collision for each event, the output needs to be a number (crisp value) and not a fuzzy set. This crisp number is obtained in a process known as the defuzzification. The defuzzification considers the centre of gravity which denotes a point representing the centre of gravity of the aggregated fuzzy set A, on the interval [a, b] which can be calculated using Equation 3

$$z_{COG} = \frac{\int_a^b z.\mu_A(z)dz}{\int_a^b \mu_A(z)dz} \qquad (3)$$

where z_{COG} is the crisp output, $\mu_A(Z)$ is the aggregated membership function, and z is the output variable.

3. EXPECTED RESULTS AND DISCUSSION

The proposed fuzzy decision tree will find the most suitable nodes defined by an attribute-value pair that is selected by considering the continuous and discrete values in the database. Moreover, using the fuzzy data input and fuzzy entropy impacts the performance of the learning by involving the discrete and continuous values in the database efficiently. Besides, the fuzzy rough set selection is applied on vehicle collision dataset to select those conditional features with the minimum correlation and improve the performance of constructed decision tree. As the dataset includes a series of inconsistent objects (uncertainty and vagueness), the FRFS feature selection algorithm should deal with the uncertainty and reduce the vagueness in rule extraction method and result in accurate rules with the minimum correlations. In this paper, the Fuzzy rough set Feature Selection is used to improve the efficiency of proposed method. This leads to more accurate results in comparison with the existing classification method like ID3 (Iterative Dichotomiser 3), C4.5, and CART (Classification And Regression Tree) (Cherfi et al., 2018).

4. CONCLUSION AND FUTIRE WORK

In this research, we propose a fuzzy rule extraction combining with a fuzzy decision-making model (predict model) in 3D spaces. In other words, a fuzzy feature selection, an intelligent rule extraction, and the role of data mining in the development of a smart geo-spatial dashboard in the term of smart city were investigated Although it is not possible to deal with all the requirements and barriers, this paper insist on providing a general view of a smart decision engine for a smart geo-spatial smart dashboard and its role in multi-building area. In the future we are going to drive this research to a smart multi building dashboard for smart citizens (tenants) and smart management with involve of 3D GIS data, BIM and sensor stream data. This dashboard will have four major types of analytics like descriptive analytics to answer the question of what happened based on historical data, diagnostic analytics to determine why happened, predictive analytics to predict the phenomena in a period of time and diagnostic analytics to help smart citizens and managers to what action to take based on multi-building structured, semi structured and unstructured data.

REFERENCES

Abbad, H., Bouchaib, R., 2017. Towards a Big Data Analytics Framework for Smart Cities. Proceedings of the Mediterranean Symposium on Smart City Application 17:1--17:5. https://doi.org/10.1145/3175628.3175647

Cherfi, A., Nouira, K., Ferchichi, A., 2018. Very Fast C4.5 Decision Tree Algorithm. Applied Artificial Intelligence 32, 119–137. https://doi.org/10.1080/08839514.2018.1447479

Fang, H., 2015. Managing Data Lakes in Big Data Era 820–824.

Iyer, A.R., Prasad, D.K., Quek, C.H., 2018. PIE-RSPOP: A brain-inspired pseudo-incremental ensemble rough set pseudo-outer product fuzzy neural network. Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2017.11.027

Janecek, A., Gansterer, W., Demel, M., Ecker, G., 2008. On the Relationship Between Feature Selection and Classification Accuracy, in: Saeys, Y., Liu, H., Inza, I., Wehenkel, L., de Pee, Y. Van (Eds.), Proceedings of the Workshop on New Challenges for Feature Selection in Data Mining and Knowledge Discovery at ECML/PKDD 2008, Proceedings of Machine Learning Research. PMLR, Antwerp, Belgium, pp. 90–105.

Khine, P.P., Wang, Z.S., 2018. Data lake: a new ideology in big data era. ITM Web of Conferences 17, 03025. https://doi.org/10.1051/itmconf/20181703025

Ma, Z., Ren, Y., 2017. Integrated Application of BIM and GIS: An Overview, in: Procedia Engineering. https://doi.org/10.1016/j.proeng.2017.08.064

Moghaddam, H.K., 2015. Vehicle Accident Severity Rule Mining Using Fuzzy Granular Decision Tree. University of Calgary. https://doi.org/http://dx.doi.org/10.5072/PRISM/28636

Moghaddam, H.K., Wang, X., 2014. Vehicle Accident Severity Rules Mining Using Fuzzy Granular Decision Tree 280–287.

Nayak, P., Devulapalli, A., 2016. A Fuzzy Logic-Based Clustering Algorithm for WSN to Extend the Network Lifetime. IEEE Sensors Journal. https://doi.org/10.1109/JSEN.2015.2472970

Pradhan, B., 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers and Geosciences. https://doi.org/10.1016/j.cageo.2012.08.023

Roisenberg, M., Schoeninger, C., da Silva, R.R., 2009. A hybrid fuzzy-probabilistic system for risk analysis in petroleum exploration prospects. Expert Systems with Applications 36, 6282–6294. https://doi.org/10.1016/j.eswa.2008.07.060

Suakanto, S., Supangkat, S.H., Suhardi, Saragih, R., 2013. Smart city dashboard for integrating various data of sensor networks, in: International Conference on ICT for Smart Society. IEEE, pp. 1–5. https://doi.org/10.1109/ICTSS.2013.6588063

United Nations Department of Economic and Social Affairs Population Division, 2017. World Population Prospects The 2017 Revision Key Findings and Advance Tables, World Population Prospects The 2017. https://doi.org/10.1017/CBO9781107415324.004