

SCHOOL MAPPING ANALYSIS SYSTEM BY INTEGRATION OF SPATIAL DATA MINING FOR AURANGABAD CITY

Sudhir P. Khobragade¹, K.V.Kale²

¹Department of CS and IT, Dr. BAMU Aurangabad, India

²Department of CS and IT, Dr. BAMU Aurangabad, India

Abstract

India is a developing country with large and rapidly growing population; it overtook different challenges and result socio-economic indicators tend not to show positive movement. Education is the basic elementary for any country. School mapping is an crucial planning tool to overcome probabilities of regional variation arising from the investment policies of the public jurisdiction. In Developing Countries School mapping (SM) is a normative approach to the planning of schools which is used to create necessary conditions to achieve universal primary and secondary education. The objective of this research paper is to generate knowledge based rules by using spatial data mining to classify the category of data and generate the rules, which is useful for government authorities to get the locations of categories of schools with detail information and parents to decide best school for their children.

Keywords: School Mapping, Spatial Data Mining, Smart city, GIS, K-means Classifier, Google Map API, Clustering

-----***-----

1. INTRODUCTION

The Smart City is a new concept and mode of city build out in recent days, which is the combination of the new production of ICT application, urban transformation and the extensive utilization of network applications. Cities are gaining increasingly high significance as a means of making accessible all the services and applications permitted by ICT to citizens, companies and command authorities that are part of a city's system. It aims to grow citizens' quality of life and improve the efficiency and quality of the services provided by governing entities and businesses. The fabrication of Smart City will make the infrastructure more intelligent, the public services more appropriate, the social management more sophisticate, the ecological environment more worth liveable and the industrial structure more optimize. It will generate a long-term procedure for sustainable growth, and help the city speedily develop. Smart City is good for industrial; ICT and the city itself simultaneously expand. [3]

Clustering is one of the most important tasks in data mining process for discovering groups and identifying distributions and patterns in the underlying data. Clustering problem is about split-up a given data set into groups (clusters) such that the data points in a cluster are more alike to each other than points in different clusters. A clustering procedure could group the customers in such a way that customers with similar acquire patterns are in the same cluster. Thus, the main concern in the clustering process is to reveal the organization of patterns into "sensible" clusters, which allow us to locate similarities and differences, as well as to derive useful conclusions about them. This idea is applicable in many eras, such as life sciences, medical sciences and engineering [7].

Spatial data mining is the process of discovering interesting and formerly undisclosed, but potentially important patterns from huge spatial datasets [1]. The extensive and world-wide use of spatial datasets has led to the need for discovering spatial knowledge. The complexity of dealing with spatial datasets has contributed the traditional methods redundant, which calls for specialized techniques in spatial data mining for extraction and locating useful information [2].

2. RELATED WORK

School mapping is the science and art of producing geospatial databases with relational databases of educational, demographic, economic and social statistics for educational and schools directorates to support educational plans and decision makers.

In order to do proper planning and to execute, the educational division should not start any project before making available to all involved parties a comprehensive database of all schools, students, teachers and resources. Building that part of GIS layers with showing school locations and other notable geographic characteristics such as streets, city-zoning, badger, school-directorates, sub-districts, districts, locality area, governorates & other features would provide an crucial tool for planners. For example, school mapping can be utilized in Maharashtra to approve proposal scheme for constructing new schools or/and refurbish existing schools. Also, it is very important to build the management intake and the staff in general to be adaptation of utilizing such technology as it is an ongoing process. Furthermore, the data has to be renovating regularly [4, 5]. In this section we will study comparative analysis of different techniques using spatial data mining.

Table 1.Comperative analysis of spatial mining Techniques

Characteristics	Clustering and outlier detection	Classification
Algorithms uses	K-means, PAM, CLARA, CLARAN etc.	K-nearest neighbor classifier, Apriori SVM etc .
Machine learning	Unsupervised	Supervised
Performance	High	Low
Robust	Yes	No

3. METHODOLOGY

In methodology we study different methods that we used with dataset for the classification and generations of rules to choose the best school around for the children.

We have used dataset in the form of ARFF (Attribute Relation File Format) which contains the data in the form of numeric and nominal attributes. By using different classification algorithms we classify the datasets and generate the rules, the classification algorithms we divided into the classification using J48 classifier and Generate rules using Jrip algorithms.

3.1. K-nearest Neighbor Algorithm (KNN)

In this algorithm, an object is classified by its majority of the neighboring objects with the object being assigned to the class most common among the k-neighbors [2].

Algorithm:

Step 1: Assume k training objects, and a test object $z = (x', y')$.

Step 2: Compute distance of object z with each object in the training set.

Step 3: Select k-objects closest to the test object z.

Step 4: Assign the class of test object z as the most common class among the k-neighbor objects.

KNN is suited for multi-model classes as well as applications in which an object has many class labels.

3.2. JRIP Rules Classifier

JRip (RIPPER) is one of the fundamental and most accepted algorithms. Classes are inspected in increasing size and an opening set of rules for the class is generated using incremental reduced error JRip (RIPPER) proceeds by considering all the examples of a particular decision in the training data as a class, and finding a set of rules that covers all the members of that class. Then after it proceeds to the next class and does the exact, repeating this until all classes have been considered [9].

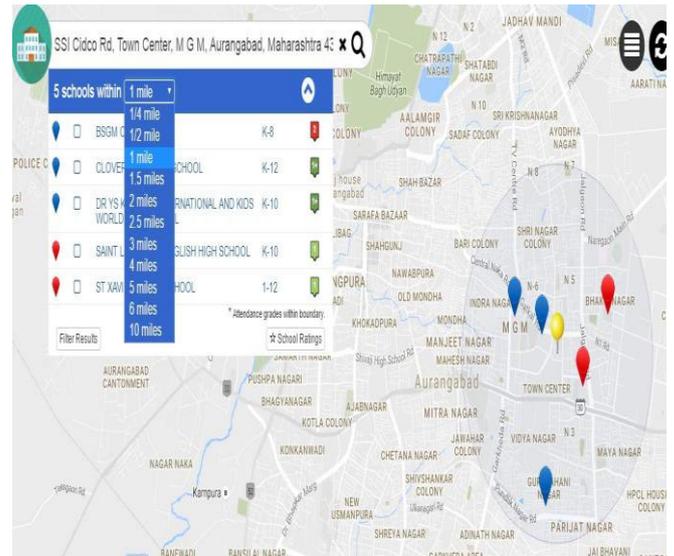


Fig 1: School mapping system (Within 1 KM Search distance)

As shown in fig 1 , we mapped the school locations using latitude and longitude onto the google maps Using API [8], database where collected and information are placed onto the maps with required information, database originally collected in the .CSV format which is in stylesheet ,later we converted requited attribute to the ARFF (Attribute relation file format) , for that we choose some important attribute like classification of the school,type of school (i.e Elementary or high school) to extract the knowledge by applying different classification and generates rules using JRip classifier , where 4 rules are generated.

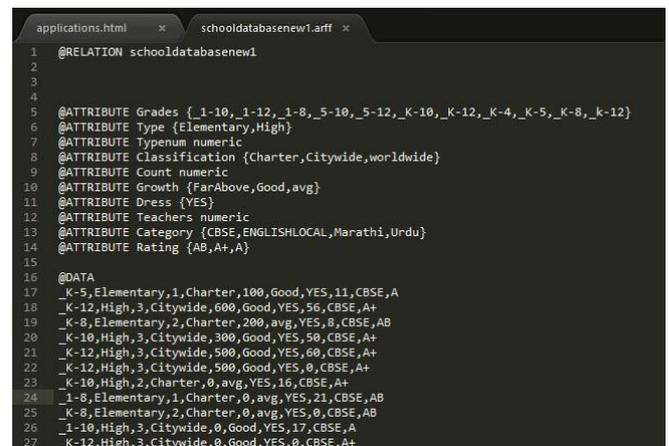


Fig 2.ARFF Of SchoolDatabase

From the dataset which is useful for the people to decide the best school in Aurangabad city.

4. EXPERIMENTAL WORK

4.1. Data Collection

The GPS Data (Latitude and Longitude) of schools are collected along with detail information of the categorized schools to develop the database. Database of school information consist of name of school, address, zip, phone,

grade, grade boundary, type of school, classification, rating of the school, growth, attainment, culture, medium etc. we collected 46 different school information and mapped them onto Google maps along with required information. Total 29 attribute table we used to deal with schools information and all are in its original form [1].

4.2 Result Analysis

The main motivation of this project is to locate all the schools such as English medium schools, CBSE, Urdu medium Schools & Marathi medium schools of Aurangabad City through Fusion Table on Google API [6]. We can analyze the results by comparative study using different classification algorithms.

4.2.1 J48 Classifier

Table2: Attribute wise result of data

Sr. No	Instance(Independent Attribute)	Correctly Classified(Out of 46)	Incorrectly Classified (out of 46)
1	Grades	34 (73.913 %)	12 (26.087 %)
2	Type	43 (93.4783 %)	3 (6.5217 %)
3	Classification	42 (91.3043 %)	4 (8.6957 %)
4	Growth(Reputation)	42 (91.3043 %)	4 (8.6957 %)
5	Category	33 (71.7391 %)	13 (28.2609 %)
6	Rating	36 (78.2609 %)	10 (21.7391 %)

From table 2 we classified the attribute wise results for correctly and incorrectly classified attributes by applying J48 on each attribute and classified the results. By using decision tree we classified the results and by using Jrip classifier we generated the rules which is helpful for school committees, parents for decision making to choose appropriate school in the city. Data mining technique generates the knowledge on the dataset so that it predict the best attributes and apply the rules, rules are text based results using Jrip classifier we displayed it in figure 3. For data mining, we have used Weka, an java based open source software. We have done the classification, clustering for further investigation(6 attributes, 46 instances). We can observe Table 2 gave good classification for the some attributes like type, classification, growth. So, we can guess the best school with the analysed data in detail. As per from the growth we conclude the best and reputed school around. For example we consider attribute Rating and observe the result. The confusion matrix is as follows:

=== Confusion Matrix ===

```
a b c <-- classified as
22 0 0 | a = AB
1 9 1 | b = A+
6 2 5 | c = A
```

The classified matrix will show the classified results using J48 Classifier on attribute rating, ratings are declared in terms of AB,A,A+ to categorise the school category.

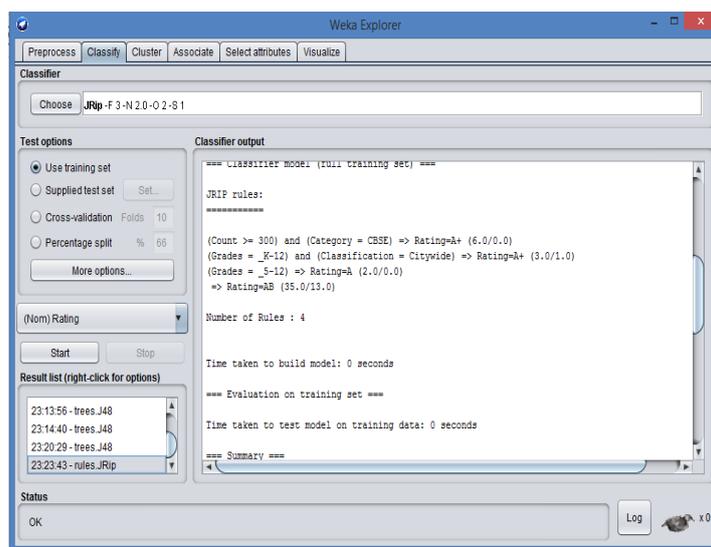


Fig3. Result (weka) Classified JRip Rule on Attribute Rating

The associated results are generated for attribute rating, rules are based on if and else conditions. => Portion declares the associated result of the condition on each condition. 4 rules are generated for current attribute.

Cluster is used to collect the instances together to get the best clustered results from the selected attribute, in fig4, we showed results for attribute classification which chooses best instance of attribute and generates the clusters. K-means algorithm applied to cluster the result for attribute classification. Incorrectly classified and different clusters are from are shown in fig4.

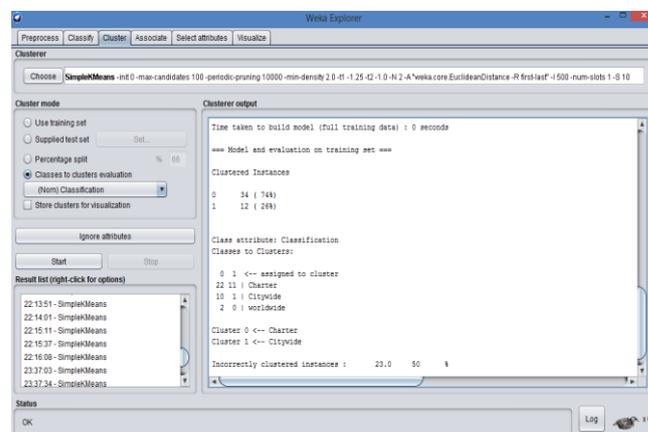


Fig4. Result (weka) clustered instance result for the attribute classification

4.2.2. Comparative Results of Classification

Algorithm Used

In paper we used J48 classifier for classification of ARFF result set, as J48 best suits for classifying dataset as compare to Naive Bayes, Random tree and ZeroR classifiers. From

Table 2 we analyses the result of J48 classifier and in Table3 we compare 3 classifiers and compared with Table 2 , found that J48 gives best results in terms of best selection of school from dataset.

Table 3: Attribute wise classifier result of data

Sr. No	Instance(Independent Attribute)	Naive Bayes		Random Tree		ZeroR	
		Correctly Classified (Out of 46)	Incorrectly Classified (out of 46)	Correctly Classified (Out of 46)	Incorrectly Classified (out of 46)	Correctly Classified (Out of 46)	Incorrectly Classified (out of 46)
1	Grades	19 (41.3043 %)	27 (58.6957 %)	21 (45.6522 %)	25 (54.3478 %)	12 (26.087 %)	34 (73.913 %)
2	Type	39 (84.7826 %)	7 (15.2174 %)	44 (95.6522 %)	2 (4.3478 %)	34 (73.913 %)	12 (26.087 %)
3	Classification	39 (84.7826 %)	7 (15.2174 %)	38 (82.6087 %)	8 (17.3913 %)	33 (71.7391 %)	13 (28.2609 %)
4	Growth(Reputation)	37 (80.4348 %)	9 (19.5652 %)	28 (60.8696 %)	18 (39.1304 %)	31 (67.3913 %)	15 (32.6087 %)
5	Category	20 (43.4783 %)	26 (56.5217 %)	18 (39.1304 %)	28 (60.8696 %)	18 (39.1304 %)	28 (60.8696 %)
6	Rating	33 (71.7391 %)	13 (28.2609 %)	26 (56.5217 %)	20 (43.4783 %)	22 (47.8261 %)	24 (52.1739 %)

from Table 3 Naïve Bayes also perform well in case of classification , growth and type as equal compare with J48 but the rules generated after the classification fails in generating with required attributes like grades when compare with good schools around , consequently rating will goes down and results doesn't show what we expect from the system. A Naive Bayes [10] classifier assumes that the presence of a specific feature in a class is independent to the presence of any other feature. The random trees classifier captures the input feature vector, classifies it with each tree in the forest, and outputs the class label that received the majority of "votes". In the case of a regression, response is the average of the responses over all fully trees in the forest [11]. ZeroR classifier predicts the important category. Although there is no predictability power in ZeroR classifier, it is used for determining a baseline performance as a benchmark for other classification [12].

5. CONCLUSION

Mapping of different categories of schools is helpful for user/parents to choose the right school for his children providing different filters which are new idea for the smart city development.

This paper gives an idea of spatial data mining to extract the hidden rules from the classification of attributes and clustered them to unique groups of same instance, so that we can predict the cases and decide for the appropriate school in Aurangabad city.

ACKNOWLEDGMENT

The authors would like to acknowledge and thanks to University Grants Commission (UGC), India for granting UGC SAP (II) DRS Phase-I & Phase-II F. No. 3-42/2009 & 4-15/2015/DRS-II and FIST DST for Laboratory facility to Department of Computer Science and Information Technology.

REFERENCES

- [1]. N. Sumathi, R. Geetha and Dr. S. Sathiyabama, "spatial data mining-techniques trends and its applications", Journal of Computer Applications, Vol.1, No.4, Oct – Dec 2008.
- [2]. Rushabh Shah, Arib Patel, Lynette D'Mello, "Spatial data mining techniques", International Journal of Emerging Technology and Advanced Engineering, volume 5, issues 10, 2015.

- [3]. Thomas Blaschke, Geoffrey J. Hay, Qihao Weng and Bernd Resch, "Collective Sensing: Integrating Geospatial Technologies to Understand Urban Systems" 3(8), 1743-1776, 2011.
- [4]. Swapnil R Rajput, Mohd Sohel Deshmukh and Karbhari V Kale. Article: Cross-platform Smartphone Emergency Reporting Application in Urban Areas using GIS Location based and Google Web Services. *International Journal of Computer Applications* 130 (12):27-33, November 2015. Published by Foundation of Computer Science (FCS), NY, USA.
- [5]. Sudhir P. Khobragade, K. V. Kale, "School Mapping System using GIS for Aurangabad City", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4, Issue 10, October 2016.
- [6]. Kamalpreet Kaur Jassar and Kanwalvir Singh Dhindsa. Article: Comparative Study of Spatial Data Mining Techniques. *International Journal of Computer Applications* 112(14):19-22, February 2015.
- [7]. Maria Halkidi, Yannis Batistakis and Michalis Vazirgiannis, "On Clustering Validation Techniques", *Journal of Intelligent Information Systems*, Vol.17, pp. 107–145, 2001.
- [8]. Swapnil R.Rajput and K.V.Kale, "A Review Paper on GIS Web-System to Support Emergency Situations in Urban Areas and Provide Services", *International Journal of Computer Sciences and Engineering*, Volume-03, Issue-05, Page No (345-350), May -2015, E-ISSN: 2347-2693.
- [9]. Anil Rajput , Ramesh Prasad Aharwal, Meghna Dubey, S.P. Saxena, Manmohan Raghuvanshi," J48 and JRIP Rules for E-Governance Data", *International Journal of Computer Science and Security (IJCSS)*, Volume (5) : Issue (2) : 2011.
- [10]. <https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/> [Accessed: March 22, 2017].
- [11]. http://docs.opencv.org/2.4/modules/ml/doc/random_trees.html [Accessed: March 22, 2017].
- [12]. <http://www.saedsayad.com/zeror.htm> [Accessed: March 22, 2017].