

An Approach to Determine Non-filer's of Property Tax using Clustering and Classification

Kanchan Jha
 Department of Computer
 Science and Engineering
 BUIT BU
 Bhopal, (MP)

Divakar Singh
 Department of Computer
 Science and Engineering
 BUIT BU
 Bhopal, (MP)

Diwakar Chaudhary
 Department of Information
 Technology
 BUIT BU
 Bhopal, (MP)

ABSTRACT

Data mining use everywhere successively, it is the knowledge extraction from large data. In this research, we are using clustering and classification methods to mine the data by using hybrid algorithm. We use K-MEANS algorithm from clustering and CART algorithm from decision tree. Clustering method will use for make the clusters of similar groups to extract the easily features or properties and decision tree method will use for choose to decide the optimal decision to extract the valuable information. This paper reports the development of an approach for property tax that works for the administrator as well as local persons. Using this approach administrator can determine the non-filer of property tax and improve the quality of decisions making.

Keywords

Data mining, Property tax, classification, clusters.

1. INTRODUCTION

Data mining is the important step for discover the knowledge in knowledge discovery process in data set. Data mining provide us useful pattern or model to discovering important and useful data from whole database [2].

A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A good clustering algorithm is able to identity clusters irrespective of their shapes. Other requirements of clustering algorithms are scalability, ability to deal with noisy data, insensitivity to the order of input records, etc. Data mining is a multi-step operation. It requires collect and prepare data for a data mining algorithm, mining the data, interpret results and taking appropriate action. The accessed data can be stored in one or more operational databases, a data warehouse or a flat file. In data mining the data is mined using two learning approaches i.e. supervised learning or unsupervised clustering.

Classification is an important task in data Mining. It belongs to directed learning and the main methods include decision tree, neural network and genetic algorithm. Decision tree build its excellent tree model by selecting important association features. While selection of test attribute and partition of sample sets are two parts in building trees. Different decision tree methods will approve different technologies to settle these problems. Algorithms include ID3, C4.5, CART and SPRINT etc.

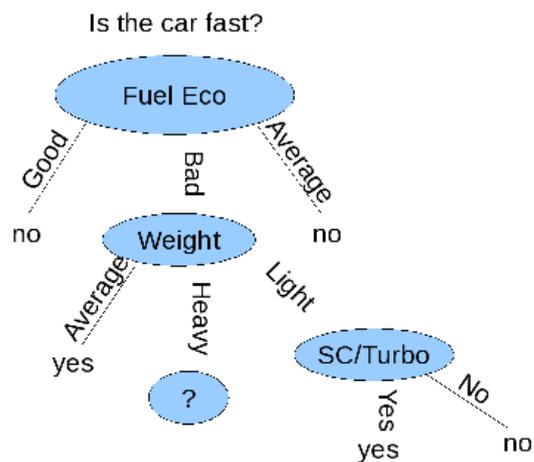


Fig 2: Decision Tree

Hybrid techniques of clustering and classification method for large dimensional dataset. Clustering analysis is an important and popular data analysis technique that is large variety of fields. Clustering and classification are the mostly used methods of data mining. Clustering can be used for define and decision tree can be applied to analyzing [5].

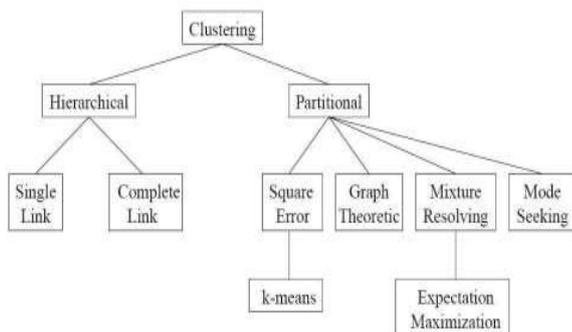


Fig 1: clustering

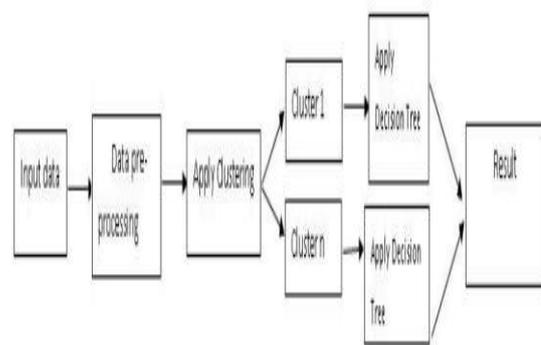


Fig 3: Hybrid model

2. LITRETURE RIVIEW

Heena Sharma, Navdeep Kaur Kaler compare the effectiveness of two stage clustering and decision tree data mining algorithms by applying them to data sets. Experiment results will show that like two stage algorithms produce the best classification efficiency. Their approach can remove the shortcoming of hybridization of algorithms and improve the results on applying them to data sets [3].

N. Sivaram, K. Ramar, use engineering data sets and the input attributes to determine through knowledge engineering in an IT industry. The process involves defining the problem, identifying relevant stake holders, and learns about current explanation to the problem [5].

Norul Hidayah Ibrahim, Aida Mustapha, Rozilah Rosli, Nurdhiya Hazwani Helmee, apply hybrid classification model refers to a combination of two data mining tasks, which are clustering and classification in effort to obtain higher accuracy result. Previously, hybrid classification models have been applied to predict patients who have higher risk in having diabetes by looking at the patients' profiles. This information is useful to help overcome a predictable diabetic patient [7].

D.Lavanya, Dr.K.Usha Rani apply a hybrid approach of CART decision tree classifier and clustering on breast cancer data sets. The capability of hybrid approach has been compared against CART with Feature Selection, Classification with Clustering and without Feature Selection in terms of efficiency [8].

Md. Hedayetul Islam Shovon Mahfuza Haque apply K-means clustering algorithm on the data so that they can group the students in three classes "High" "Medium" and "Low" according to their new grade. New grade is calculated from the previous semester grade that means external assessment and internal assessment. Then apply decision tree to make correct decision about the student which is need to take by the instructor [9].

Varun Kumar, Nisha Rathee experimented and equated the outcome of a simple classification method with the outcome of integrated clustering and classification method. As a result they found, the integration of clustering and classification techniques gives more accurate and robust result than applying either of them alone [10].

3. PROPOSED WORK

We divide database into n clusters by applying k-means cluster after divide the database we apply decision tree to determine the name who did not file the property tax.

3.1 K-means

Step 1: Randomly select 'c' cluster centers

Step2: Calculate the length between each data point and cluster centers

Step 3: Allow the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers

Step 4: Again calculate the new cluster center

Step 5: Recalculate the distance between each data point and new obtained cluster centers

Step 6: If no data point was changed then stop, otherwise repeat from step 3.

After applying cluster data is divided into two clusters then apply decision tree on them. The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner. Decision Tree Algorithm: generate a decision tree from the given training data.

3.2 Decision tree

Step 1: Create a node N

Step 2: If samples are all of the same class, C then

Step 3: Return N as a leaf node classify with the class C;

Step 4: If attribute-list is empty then

Step 5: Return N as a leaf node classify with the most common class in samples.

Step 6: Select test-attributes, the attribute among attribute-list with the highest information gain;

Step 7: Label node N with test-attribute;

Step 8: For each known value a_i of test-attribute.

Step 9: Grow a branch from node N for the position test attribute = a_i ;

Step 10: Let S_i be the set of samples for which test-attribute = a_i ;

Step 11: If S_i is empty then

Step 12: Attach a leaf classify with the most common class in samples;

Step13: Else attach the node returned by generate-decision-tree (S_i , attribute-list-attribute);

Each internal node tests an attribute, each branch corresponds to characteristic value, and each leaf node assigns a classification.

4. RESULT AND ANALYSIS

OwnerName	Address	Colony
Ramjeet Yadav	H.no-no./i-259, Durga Nagar, Kf Durga Nagar Khand-3	
Neetu Yadav	H.no-no./i-260, Durga Nagar, Kf Durga Nagar Khand-3	
Sanjay Dagour	H.no-no./i-261, Durga Nagar, Kf Durga Nagar Khand-3	
Dilip Niranjn	H.no-no./i-264, Durga Nagar, Kf Durga Nagar Khand-3	
Tota Ram Niranjn	H.no-no./i-265, Durga Nagar, Kf Durga Nagar Khand-3	
Ram Kishan Khuswaha	H.no-48./i-266, Durga Nagar, Kf Durga Nagar Khand-3	
Ram Kali Bai	H.no-c/163./i-267, Durga Nagar Durga Nagar Khand-3	
Jabbu Lal Lohat	H.no-no./i-294, Durga Nagar, Kf Durga Nagar Khand-3	
Balbir Lohat	H.no-no./i-295, Durga Nagar, Kf Durga Nagar Khand-3	
Buty Lohat	H.no-no./i-296, Durga Nagar, Kf Durga Nagar Khand-3	
Gulab Valmeki	H.no-b/58./i-297, Durga Nagar, Durga Nagar Khand-3	
Ramesh Lahat	H.no-no./i-297, Durga Nagar, Kf Durga Nagar Khand-3	
Sunita	H.no-d/110./i-299, Durga Nagar Durga Nagar Khand-3	
Raju Paril	H.no-d/104./i-300, Durga Nagar Durga Nagar Khand-3	
Jetendra Saha	H.no-no./i-301, Durga Nagar, Kf Durga Nagar Khand-3	
Vinod Saha	H.no-no./i-303, Durga Nagar, Kf Durga Nagar Khand-3	
Manoj Saha	H.no-no./i-304, Durga Nagar, Kf Durga Nagar Khand-3	
Ram	H.no-b/48./i-305, Durga Nagar, Durga Nagar Khand-3	
Bashanwanti Prasad Saket	H.no-no./i-306, Durga Nagar, Kf Durga Nagar Khand-3	
Vimla Thakur	H.no-no./i-262, Durga Nagar, Kf Durga Nagar Khand-3	
Jai Sanker Prasad	H.no-no./i-263, Durga Nagar, Kf Durga Nagar Khand-3	
Gopilal Malviya	H.no-c/59./i-268, Durga Nagar, Durga Nagar Khand-3	
Govind Malviya	H.no-no/i-268, Durga Nagar, Kh Durga Nagar Khand-3	
Rajni Bai	H.no-c/58./i-269, Durga Nagar, Durga Nagar Khand-3	
Kamal Rathore	H.no-no./i-270, Durga Nagar, Kf Durga Nagar Khand-3	
Mona Saket	H.no-no./i-307, Durga Nagar, Kf Durga Nagar Khand-3	
Jaganat Prasad Saket	H.no-no./i-308, Durga Nagar, Kf Durga Nagar Khand-3	

Fig 4: Dataset

In this work we use property data, there lots of field in this data but we use only main field which is required for property tax. We use person's name, nature of property use, year of construction, water connection is available or not, property tax is paying by them or not.

	1	2	3	4	5	6
1	4.5794e+05	1.2121	20.9495	22.1515	1910	1.3737
2	5.2164e+06	1.3297	21.0073	22.0110	1.9377e+03	1.4908

Fig 5: Clusters' Centroids

Table 1: Cluster 1

Ramjeet Yadav	21	22	1910	1
Neetu Yadav	21	22	1910	2
Sanjay Dagour	21	22	1910	1
Dilip Niranjana	21	22	1910	1
Tota Ram Niranjana	21	22	1910	1
Ram Kishan Khuswaha	21	22	1910	1
Ram Kali Bai	21	22	1910	1

Table 2: Cluster 2

Rajesh Bilkhan	21	22	1997	1
Refused	21	22	1997	1
Sarvesh Kumar	21	22	1997	1
Ratan Malviya	21	22	1997	1
Ajay Rathore	21	22	1997	1
Panchwati Bai Thakur	21	22	1997	1
Chhedi Lal Gupta	21	22	1997	1

Table 1 and 2 define the clusters which are dividing on the basis of construction year of building. After divide data into cluster classification apply on both clusters.

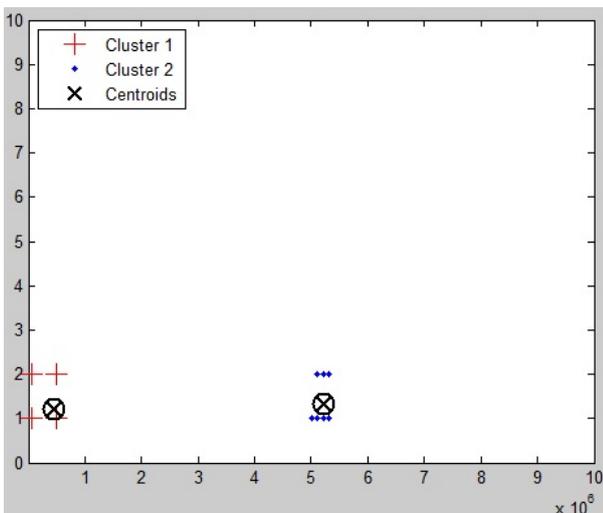


Fig 6: Cluster

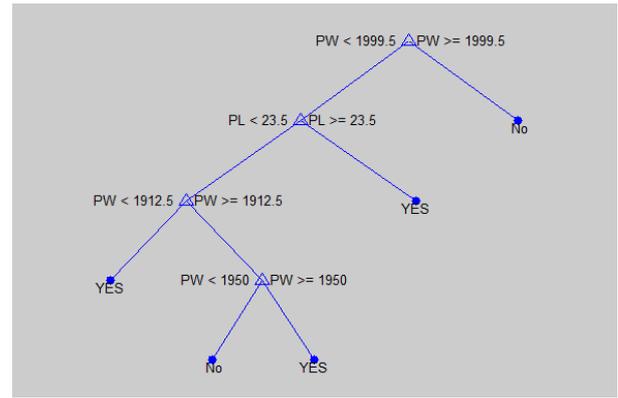


Fig 7: Decision Tree

In figure 9 decision tree represented a set that who provide the property tax or not for this we use property tax is paying or not as a class and rest of thing as attribute as nature of property use, year of construction, water connection is available or not.

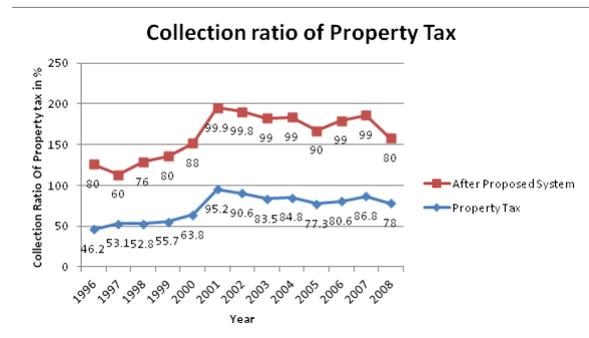


Fig 8: Graph representing the comparison of year wise Tax benefit with after applied proposed system

In figure 8 a comparison is presented between actual data and after applying proposed system data. Actual data set of property tax collected between 1996-2008, the lower line represents the actual property tax collected, and upper line represents the data that could be happened when proposed system applied.

Table 3: User Id

User ID	Name	Address
P0053112	Neetu Yadav	H. no.i-260, Durga nagar, Khand -3, Bhopal
P0053118	Manoj Saha	H. no.i-305, Durga nagar, Khand -3, Bhopal
P0053119	Ram	H. no.a-51, 19 Durga nagar, Khand -3, Bhopal
P00531120	Munna Lal Saket	H. no.i-306, Durga nagar, Khand -3, Bhopal
P00531138	Vinay Kumar	H. no.i-351, Durga nagar, Khand -3, Bhopal

In table 3 user complete detail is define. User id is unique for each filer's, username define the name of person, and contact address contains the full address of user.

5. CONCLUSION AND FUTURE WORK

In proposed work we prepare a system which determines the non-filer of property tax and give the solution how to recover from that, for this we use property data as database. After collecting the data we did preprocessing to get the valid and pure data, when data is preprocessed we apply clustering algorithm and get the n number of cluster, when data is divided in different cluster then apply classification on each data and identified the non-payer of property tax. As shown in figure 8 when proposed system is applied collection should increase in comparison of the actual data. Our propose system works as follows:

1. It divides the data into number of cluster with the help of cluster algorithm.
2. Then apply classification on clusters to find non-filers.
3. When we find non-filer then we complain them through mail or mobile by sending them massage.

In future work any one can prepare a complete system for client and administration for transaction, complain, to get information about property tax, for guide new user etc.

6. REFERENCES

- [1] Dhiman Richa, Vashisht Sheveta, Sharma Kapil "A CLUSTER ANALYSIS AND DECISION TREE HYBRID APPROACH IN DATA MINING TO DESCRIBING TAX AUDIT" International Journal of Computers & Technology Volume 4, No. 1, Jan-Feb, 2013 ISSN 2277-3061.
- [2] Kumar Ela, Solanki Arun "A Combined Mining Approach and Application in Tax Administration." International Journal of Engineering and Technology Vol.2(2), 2010.
- [3] Sharma Heena, Kaler Navdeep Kaur, "Data Mining with Improved and Efficient Mechanism in Clustering Analysis and Decision Tree as a Hybrid Approach" International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-2, Issue-5, April 2013, ISSN: 2278-3075.
- [4] Bolelli Levent, Ertekin Seyda, Zhou Ding, C. Lee Giles, "K-SVMMeans: A Hybrid Clustering Algorithm for Multi-Type Interrelated Datasets." IEEE/WIC/ACM International Conference on Web Intelligence 2007.
- [5] Sivaram N., Ramar K., "Applicability of Clustering and Classification Algorithms for Recruitment Data Mining" International Journal of Computer Applications, Volume 4, No.5, July 2010, (0975 – 8887).
- [6] Enguehard René, Fowler Benjamin, Hoeber Orland, Devillers Rodolphe, Banzhaf Wolfgang "Integrating human knowledge within a hybrid clustering-classification scheme for detecting patterns within large movement data sets". AGILE 2012 – Avignon, April 24-27, 2012.
- [7] Lavanya D., Rani K.Usha, "A Hybrid Approach to Improve Classification with Cascading of Data Mining Tasks" International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 2, Issue 1, January 2013.
- [8] Ibrahim Norul Hidayah, Mustapha Aida, Rosli Rozilah, Helmee Nurdhiya Hazwani, "A Hybrid Model of Hierarchical Clustering and Decision Tree for Rule-based Classification of Diabetic Patients" Norul Hidayah Ibrahim et.al / International Journal of Engineering and Technology (IJET) Vol 5 No 5 Oct-Nov 2013.
- [9] Islam Shovon Md. Hedayetul, Haque Mahfuza, "An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.3, No. 8, 2012.
- [10] Kumar Varun, Rathee Nisha, "Knowledge discovery from database Using an integration of clustering and classification" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.3, March 2011.
- [11] Namdeo V, Singh A, Singh D, Jain RC "Result Analysis Using Classification Techniques" 2010 International Journal of Computer Applications (0975-8887) 1 (22).
- [12] Bascacov A., Cernazanu C., Marcu M. "Using Data Mining for Mobile Communication Clustering and Characterization" 8th IEEE International Symposium on Applied Computational Intelligence and Informatics, May 23–25, 2013.
- [13] Prabhu P., Anbazhagan N., 'Improving the performance of k-means clustering for high dimensional dataset', International Journal of Computer Science and Engineering, Vol 3. No.6, June 2011, pp. 2317-2322.
- [14] Balaji S. and Srivatsa S.K." Decision Tree induction based classification for mining Life Insurance Data bases" International Journal of Computer Science and Information Technology & Security (IJSITS), Vol. 2, No.3, June 2012, ISSN: 2249-9555, pp-699-703.
- [15] Bhardwaj Rupali, Vatta Sonia "Implementation of ID3 Algorithm" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 6, June 2013, ISSN: 2277 128X.
- [16] Agrawal Gaurav L., Gupta Hitesh "Optimization of C4.5 Decision Tree Algorithm for Data Mining Application" International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 3, March 2013, ISSN 2250-2459.
- [17] Bandyopadhyay Simanti "Property Taxation in Indian Cities: A Comparison of Delhi and Bangalore".
- [18] Nallathiga Ramkrishna "Analyzing the Finance of Urban Local Bodies in India: A Cross Sectional study".
- [19] Mosier Juliana, Ponomarenko Irina, Sekhon Dolly, Werner Katrina "Mining Data To Increase State Tax Revenues in California".