IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

Data Mining and Pattern Recognition Techniques Using GIS/RS

Mr. N.K. Gupta[1], Sachin Kumar[2], Anshu Verma[3]

Assistant Professor, CSE, SSET, SHIATS, Allahabad, U.P., India[1]

Student, M.Tech-CSE, SSET, SHIATS, Allahabad, U.P., India[2,3]

Abstract: Data Mining techniques is one of the major research domain and in the recent past it is useful in extracting the implicit and very helpful/handy information to increase the knowledge of the human beings, as it increases the knowledge base of the system to some certain levels about the particular database system and its subsets which further increases the chances of the new inventions as the data analyser analyses the extracted information. Artificial Intelligence is one of the invented techniques by the studied of the data mining techniques. Data Mining extracts the implicit information from the database which analysed manually using statistical techniques of the mathematics. Due to the technology advancement, semi-automated data mining came into existence but it was not enough to support as the storage led to the analysis demands which leads to the failure of the semi-automated data mining. After sometimes, fully automated data mining were developed which supports the analysis demands by the users on the large storage of the data. In this paper, data mining and pattern recognition techniques/methods are shown here to search the lost living beings and non-living beings and also exploits the unseen and untold facts about the earth. The use of the GIS/RS technologies are used to show the definite and exact locations of the objects found in the unknown and non-visited areas.

Keywords: Data mining, Pattern Recognition Techniques, Decision Trees, GIS, RS imagery, knowledge base system, LIDAR.

INTRODUCTION

In this paper we are discussing the data mining and the pattern recognition techniques for the benefits to the human beings and to the environment as well.

1.The data mining techniques will be used with GIS that will lead in extracting unknown/covered vital contents of the earth(in the earth's atmosphere and below the earth's surface).

The use of the Artificial Intelligence here is to create the automatic moving device(ROBOTS) with GPS(Global Positioning System) enabled which will be programmed to detect the objects and notify some different activities other than the usual in the earth. The ROBOT will be helpful in going through the atmosphere and below the surface of the earth where its difficult for the humans to reach at an instant and survive there for the longer period of the time.

Since LIDAR can detect the atmosphere activities and also the beneath of the earth surface but still it is not very useful for the purpose below the earth surface. As LIDAR cannot detect the activities to the distant below the surface as it is limited in the ranging upto

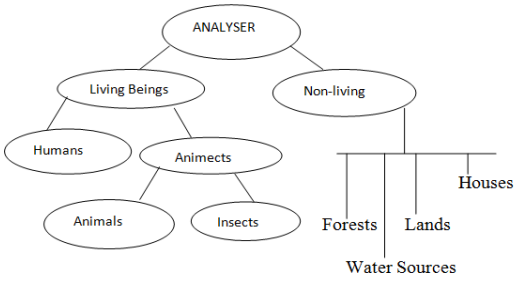
S No.		Earth's Atmospheres	Below the Earth's Crust
1	Nutrients		
2	Metals/Non-metals		
3	Gases		
4	Genes		
5	Etc		

TABLE 1 will show how the data mining using GIS with GPS inbuilt device/s in the earth atmosphere and below the earth crust extracts the important implicit in formations which can be turn as the mystery solved because there are still believe that not all the truths and facts are known by the world of the Scientists.

2. Use of the decision tree in the Pattern Recognition Techniques:

As the pattern recognition techniques used for finding the lost living beings/non-living beings from the records, it will also discover the new people, new cultures and new cultivations which is not present in the records available and that leads to the discovery of the increase in the populations when it will be recorded and enable the scientists/researchers to know the exact human effects whether goods or bad on the earth.

The decision tree will be used in the pattern recognition methods to categorised the unseen and unknown living/non-living beings in the predefined database. The decision tree analyses the data comes from the LIDAR/RS imagery and the insert it into the database system created. The decision tree can be shown as:



Decision Tree

DOI 10.17148/IJARCCE.2016.5315

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

The GIS will help in telling the position of the lost objects and the decision tree will match that objects from the database and that makes this task easier and less time taken.

The use of the GIS in this method can also be stated as that with the help of the GPS enabled device, the unseen/untold objects/resources of the earth from the earth's itself. By telling the geographical information/position which makes easy for the scientists or the research teams to reach those places.

The search method will become easy and can be done on time as the decision tree will be used for the sorting of the objects into the categories as shown in the figure to consume less time and costs indulged in the searching methods. This search methods using GIS and RS imagery will help the government/organisations to know that whether the rescue team is needed from the organisations or the research team/s should reach the place that is detected in the database system. When the objects found is a non-living objects and is found to be suitable for some research then the experts analysed it and decide which research team should be headed to the founded regions for the extraction of the data/informations. This will save the time in the decision making as the images from the RS imagery will make it clear that the from which research field objects is associated with.

3.Use of the Biometric in Pattern Recognition Techniques: The Biometric facility in the pattern recognition method will be used to identify the human beings as the known and the unknown sources to the Government and the Organisations. The Biometric will identify the human beings on the basis of the fingerprints and the facial, way of walking, voice recognising/speech pattern recognition. When the person matches predefined criteria through the Biometric, it will be defined as the known source else it will declared that person as the Intruders/unknown source and that will be notifies to the Head of the Organisation.

Thus use of Biometric in the pattern recognition method is very helpful as it provides the security and safety to the nation or organisation by verifying the person after validating the checks its needs to be done on the entrance of the organisation. This will lead to the unwanted entrance of the intruders and the unnecessary risk from them.

The Table 2 will show how the Biometric verifies the Person by validating the checks. Let two persons p1 and p2 at the entrance of the organisation.

TABLE 2

Sr No.	Checks	P1	P2
1	FINGERPRINTS	YES	NO
2	FACIAL	YES	NO
3	VOICE/SPEECH PATTERN	YES	NO
4	WAY OF WALKING	YES	NO
5	RETINA SCAN	YES	NO

Since the p1 validates all the checks in the Biometric, it will be defined as the known source and will be provided the entrance in the organisation whereas the p2 will be arrested as it tries to enters the organisation with the non-validations. It is necessary for any person to validates all the checks else not be allowed to enter.

LIDAR:
LIDAR can be defined as Light Detection And Ranging. Our approach is based on data mining principles to take advantage on intelligent techniques (attribute selection and C4.5 algorithm decision tree) to classify quickly and efficiently without the need for manipulating multie spectral images.

DATA DESCRIPTION

This study is based on LIDAR data provided by REDIAM (Consejeria de Medio Ambiente de la Junta de Andalucia, Red de Informacion Ambiental de Andalucia, n.d.) that belongs to the Regional Ministry of Andalusia. Data were acquire from coastal zones in the provinces of Huelva and C´adiz, as can be seen in Figure 1, between the 23th and 25th of September in 2007 and it was operated at a flight altitude of 1200 m with low angles(< 11 grades) and with a point density of 2 returns/m2. The pulses were geo-referenced and validated. The accuracy report indicates an accuracy of 0.5 m. in x-y position and an accuracy of 0.15 m. In z position. In addition, the rest of variables in standard LAS were provided: intensity, angle,... Together with LIDAR data, aerial photography were collected in the same flight. The aerial photography was used to assist in the selection of training and test sets.

The study zone locates in the south of the province of Huelva in the mouth of rivers Tinto and Odiel next to Atlantic Ocean(UTM30; 150960E 4124465N). Close to the city of Huelva, a mix of land covers can be found in which industrial zones, roads and railways, port facilities and natural zones stand out. Vegetation can be divided in three classes. One of them is the scarce trees of genus eucalyptus forming high vegetation class. Middle vegetation class is formed by different kinds of Mediterranean shrub that surround roads and urban zones mostly. Dry grass and bare Figure 1: Study site. It locates in Huelva city, between the mouths of the rivers Tinto and Odiel. Andalusia (Spain). earth is classified as low vegetation. In addition, the primitive land formed by marshlands near the river is another important class for land covers in this ecosystem. LIDAR data can mainly be exploded depending of three main features: density, intensity and height of the points. A brief study of the different answers by each type of land cover in every characteristic can be useful to figure out the main differences among every class. Water LIDAR does not usually reflect on water. That means plots classified as water will have low density. In addition, the few returns that reflect on water will have a low intensity because a great part of its energy is lost when it tries to go through the water surface. At last, height difference will not be very high because river usually have soft slopes near its

mouth. Marsh Marshlands are transition zones between watered terrain and vegetation and urban terrains. They are formed by low shrubs and grass. They are characterized by low heights and a medium/high distribution of intensities.

Grass and bare earth They are interior zones with very scarce vegetation or very low vegetation which produces few returns. It has the biggest intensities because of its high reflectivity in comparison with the rest of the land covers. Its height distribution is low but higher than marshland's.

Middle vegetation It is formed by bushes with medium height and they are mainly located between roads, trees,... They have a medium level of double and triple returns for every pulse. Intensities are in a medium level depending if they beat trunk or leaves. Their heights are over 1 m.

High vegetation High vegetation are mostly trees and big bushes with similar heights as trees. They have the biggest number of returns per pulse and their averaged height is high.

Roads and railways This class is formed by the infrastructure made to transport people or materials. It is characterized by low heights and high intensities. In addition, most of pulses produce just one return because of the absence of obstacles.

Urban zones The most complex class because of its variety. Intensities vary from minimum to maximum. The same can be applied for heights. This is possible because in this class we can find buildings, rubbish dumps, dock facilities and they are very different from each other.

CONCLUSION

The data mining and the pattern recognition techniques/methods with using GIS and Remote Sensing Imagery will help in the benefits of the humans and to the nature. It can also help a country in increasing the developments. The GIS/RS also helps in getting the data about the natural hazards like earthquakes/floods/cyclones which certainly saves the life of the people of the country living in there. The data mining and the pattern recognition method together used as the revolution in the today's world about the researches in exploring and extracting the unseen/unknown objects from the earth's atmosphere and beneath the crust. The pattern recognition with GIS helps in providing the securities and safety to the nation and the individual organisations. The future relies on this technique using GIS on the large impacts.

REFERENCES

[1]. Yongjian Fu " data mining: task, techniques and application".
[2]. Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar" Data Mining Techniques & Distinct Applications: A Literature Review" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9, November- 2012.
[3]. Adachi, H., Kikuchi, M., & Watanabe, Y. (2006). Electric switch machine failure detection using data-mining technique. Quarterly Report of RTRI (Railway Technical Research Institute) (Japan), 47(4), 182–186.
[4]. Ahn, H., Ahn, J. J., Oh, K. J., & Kim, D. H. (2011). Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. Expert Systems with Applications, 38(5), 5005–5012.
[5]. Akerkar, R. A., & Sajja Priti Srinivas (2009). Knowledge-based systems. Sudbury, MA,USA: Jones & Bartlett Publishers.
[6]. Al-Hamami, A. H., Al-Hamami, M. A., & Hasheem, S. H. (2006). Applying data mining techniques in intrusion detection system on web and analysis of web usage.
[7]. Information Technology Journal, 5(1), 57–63.Andronie, M., & Andronie, M. (2009). Data mining techniques used in metallurgic industry. Metalurgia International, 14(12), 17–22.
[8]. Assous, F., & Chaskalovic, J. (2010). Methodes de data mining pour l'analysed'approximations numeriques: Le cas de solutions asymptotiques des equations de Vlasov–Maxwell= data mining techniques for numerical approximations analysis: A test case of asymptotic solutions to the Vlasov–Maxwell equations.Comptes Rendus.Mécanique, 338(6), 305–310.
[9]. Assous, F., & Chaskalovic, J. (2011). Data mining techniques for scientific computing:Application to asymptotic paraxial approximations to model ultrarelativisticparticles. Journal of Computational Physics, 230(12), 4811–4827.
[10]. Bae, J. K., & Kim, J. (2011). Product development with data mining techniques: Acase on design of digital camera. Expert Systems with Applications, 38(8),9274–9280. Bae, S. H., Kim, J., & Lim, H. (2009).
[11]. A study on constructing the prediction system using data mining techniques to find medium-voltage customers causing distribution line faults. Transactions of the Korean Institute of Electrical Engineers,58(12), 23–54.
[12]. Chandrakala, D., Sumathi, S., & Saraswathi, D. (2010). Blur identification with image restoration based on application of data mining techniques. International Journal of Imaging, 4(10 A), 99–122.
[13]. Neelamadhab Padhy ,Rasmita Panigrahi "Survey of data mining application and Feature scope " ,International Journal of Computer Science, Engineering and Information Technology (IJCSEIT),Vol.2, No.3, June 2012.
[14]. Antonarakis, A., Richards, K. and Brasington, J., 2008. Objectbased land covers classification using airborne lidar. Remote Sensing of Environment 112, pp. 2988–2998.
[15]. Arroyo, L. A., Pascual, C. and Manzanera, J. A., 2008. Fire models and methods to map fuel types: The role of remote sensing. Forest Ecology and Management 256, pp. 1239–1252..
[16]. Brzank, A., Heipke, C., Goepfert, J. and Segel, U., 2008. Aspects of genearating precise digital terrain models in the wadden sea form lidar-water classification and structure line extraction. ISPRS journal of Photogrammetry & Remote Sensing 63, pp. 510–528. Canty, M. J., 2008.
[17]. Boosting a fast neural network for supervised land cover classification. Computers & Geoscience. Chust, C., Galparsoro, I., Borja, A., Franco, J. and Uriarte, A.,2008.
[18]. Coastal and estuarine habitat mapping, using lidar height and intensity and multi-espectral imagery. Estuarine, Coastal and Shelf Science.
[19]. Consejeria de Medio Ambiente de la Junta de Andalucia, Red de Informacion Ambiental de Andalucia, n.d.Dorigo, W. A., Zurita-Milla, R., de Wit, A. J. W., Brazile, J.,Singh, R. And Schaepman, M. E., 2007.
[20]. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International journal of Applied Earth Observation and Geoinformation 9, pp. 165–193.
[21]. Gamanya, R., Maeyer, P. D. and Dapper, M. D., 2009. Objectoriented change detection for the city of harare, zimbabwe. Remote Sensing of Environment 36, pp. 571–588.
[22]. Goetz, S., Steinberg, D., Dubayah, R. and Blair, B., 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperature forest, usa. Remote Sensing of Environment 108, pp. 254–263.
[23]. Hofle, B. and Pfeifer, N., 2007. Correction of laser scanning intensity data: Data and model-driven approaches. ISPRS journal of Photogrammetry & Remote Sensing. Holmes, G., Donkin, A. and Witten, I., 1994.
[24]. Weka: A machine learning workbench. In: Proc Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia.

[25]. Hudak, A. T., Crookston, N. L., Evans, J. S., Halls, D. E. And Falkowski, M. J., 2008. Nearest neighbor imputation of speciesslevel, plot-scale forest structure attributes from lidar data. Remote Sensing of Environment 112, pp. 2232–2245. Hughes, M., Schmidt, J. and Almond, P. C., 2009.
[26]. Automatic landform stratification and environmental correlation for modelling loess landscapes in north otago, south island, Newzealand. Geoderma 149, pp. 92–100.
[27]. Jensen, J. L. R., Humes, K. S., Vierling, L. A. and Hudak, A. T., 2008. Discrete return lidar–based prediction of leaf area index in two conifer forests. Remote Sensing of Environment 112,pp. 2988–2998.
[28]. Koetz, B., Morsdorf, F., van der Linden, S., Curt, T. and Allgower, B., 2008. Multi-source land cover classification for forest fire management based on imaging spectrometry and lidar data.Forest Ecology and Management 256, pp. 263–271.
[29]. Magnussen, S., McRoberts, R. E. and Tomppo, E. O., 2009.Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. Remote Sensing of Environment 113, pp. 476–488. McColl, C. and Aggett, G., 2007.
[30]. Land-use forecasting and hydrologic model integration for improved land-use decision support.Journal of Environmental Management 84, pp. 497–512.
[31]. Pascual, C., Garcia-Abril, A., Garcia-Montero, L., Martin-Fernandez, S. and Cohen, W., 2008. Object-based semiautomatic approach for forest structure characterization using lidar data in heterogeneus pinus sylvestris stands. Forest Ecology and Management 255, pp. 3677–3685.
[32]. Quinlan, J. R., 1996. Improved use of continuous attributes in c4.5. Journal of Artificial Intelligence Research 4, pp. 77–90. Schneider, J., Grosse, G. and Wagner, D., 2009. Land cover classification of tundra environments in the arctic lena delta based on landsat 7 etm+ data and its application for upscaling of methane emissions. Remote Sensing of Environment 113, pp. 380–391.
[33]. Schubert, J. E., Sanders, B. F., Smith, M. J. and Wright, N. G.,2008. Unstructured mesh generation and landcover-based resistance for hydrodynamic modeling of urban flooding. Advances in Water Resources 31, pp. 1603–1621.
[34]. Sithole, G. and G.Vosselman, 2003. Comparison of filtering algorithms. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34, pp. 71–78.
[35]. Tooke, T. R., Coops, N. C., Goodwin, N. and Voogt, J. A., 2008.Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications. Remote Sensing of Environment.
[36]. Witten, H. and Frank, E., 2005. Data mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers. H.M. Abbas and M.M. Fahmy, ªNeural Networks for Maximum Likelihood Clustering,º Signal Processing, vol. 36, no. 1, pp. 111-126, 1994.
[37]. H. Akaike, ªA New Look at Statistical Model Identification,º IEEE Trans. Automatic Control, vol. 19, pp. 716-723, 1974.
[38]. S. Amari, T.P. Chen, and A. Cichocki, ªStability Analysis of Learning Algorithms for Blind Source Separation,º Neural Networks,vol. 10, no. 8, pp. 1,345-1,351, 1997.
[39]. J.A. Anderson, ªLogistic Discrimination,º Handbook of Statistics. P. R. Krishnaiah and L.N. Kanal, eds., vol. 2, pp. 169-191, Amsterdam: North Holland, 1982.
[40]. J. Anderson, A. Pellionisz, and E. Rosenfeld, Neurocomputing 2: Directions for Research. Cambridge Mass.: MIT Press, 1990. A. Antos, L. Devroye, and L. Gyorfi, ªLower Bounds for Bayes Error Estimation,º IEEE Trans. Pattern Analysis and MachineIntelligence, vol. 21, no. 7, pp. 643-645, July 1999.
[41]. H. Avi-Itzhak and T. Diep, ªArbitrarily Tight Upper and Lower Bounds on the Bayesian Probability of Error,º IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 1, pp. 89-91, Jan. 1996.
[42]. E. Backer, Computer-Assisted Reasoning in Cluster Analysis. Prentice Hall, 1995.
[43]. R. Bajcsy and S. Kovacic, ªMultiresolution Elastic Matching,º Computer Vision Graphics Image Processing, vol. 46, pp. 1-21, 1989.
[44]. A. Barron, J. Rissanen, and B. Yu, ªThe Minimum Description Length Principle in Coding and Modeling,º IEEE Trans. Information Theory, vol. 44, no. 6, pp. 2,743-2,760, Oct. 1998.
[45]. A. Bell and T. Sejnowski, ªAn Information-Maximization Approach to Blind Separation,º Neural Computation, vol. 7, pp. 1,004-1,034, 1995.
[46]. Y. Bengio, ªMarkovian Models for Sequential Data,º Neural Computing Surveys, vol. 2, pp. 129-162, 1999. http://www.icsi.berkeley.edu/~jagota/NCS.
[47]. K.P. Bennett, ªSemi-Supervised Support Vector Machines,º Proc. Neural Information Processing Systems, Denver, 1998. J. Bernardo and A. Smith, Bayesian Theory. John Wiley & Sons, 1994.
[48]. J.C. Bedeck, Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press, 1981. Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data. J.C. Bezdek and S.K. Pal, eds., IEEE CS Press,1992.
[49]. S.K. Bhatia and J.S. Deogun, ªConceptual Clustering in Information Retrieval,º IEEE Trans. Systems, Man, and Cybernetics, vol. 28,no. 3, pp. 427-436, 1998.
[50]. C.M. Bishop, Neural Networks for Pattern Recognition. Oxford: Clarendon Press, 1995.

BIOGRAPHIES

Mr.N.K.Gupta is working in the department of CSE of SHIATS from last 10 years. He has completed his UG & PG from ALLAHABAD UNIVERSITY. He is pursuing PhD from SHIATS. He is expertise in RDBMS & DATA MINING/OBJECT ORIENTED TECHNOLOGIES field. He has guided more than 40 PhD students & 50 M.Tech and published more than 20 NATIONAL & INTERNATIONAL REPUTATED JOURNALS.

Sachin Kumar is pursuing the M.Tech in Computer Science & Engineering from Shepherd School of Engineering & Technology in the university of Sam Higginbottom Institute of Agriculture Technology & Sciences, Allahabad. His research areas interests are Data Mining using GIS/RS, Pattern Recognition Techniques using Data Mining Tools & Techniques, Artificial Intelligence, etc.

Anshu Verma is pursuing the M.Tech in Computer Science & Engineering from Shepherd School of Engineering & Technology in the university of Sam Higginbottom Institute of Agriculture Technology & Sciences, Allahabad. Her research areas interests are Data Mining, Pattern Recognition Techniques using Data Mining Tools & Techniques, Artificial Intelligence, etc.