INTEGRATED DATA MINING FRAMEWORK FOR NATURAL RESOURCE EXPLORATION

Muhammad Shaheen[1], Muhammad Shahbaz[1], Syed Muhammad Ahsan[1], Syed Athar Masood[2]

[1]Department of Computer Science and Engineering, UET, Lahore, Pakistan
shaheentanoli@gmail.com
[2] Department of Engineering Management, EME College, NUST, Rawalpindi, Pakistan.
athermasood2000@hotmail.com

Abstract: The study is aimed at developing spatio temporal data mining framework for natural resource exploration. A framework will be said as spatio temporal if it deals with changes in space over time. The study revealed different applications of existing and newly proposed spatio temporal data mining techniques on huge databases collected from diverse sources. The work is integrated on the basis of different proposed Techniques in a unified framework which could serve as a basic skeleton for hydrocarbon prospection.
[Muhammad Shaheen], Muhammad Shahbaz, Syed Muhammad Ahsan, Syed Ather Masood. **Integrated Data Mining Framework for Natural Resource Exploration.** Life Science Journal. 2012;9(1):314-316] (ISSN:1097-8135). http://www.lifesciencesite.com. 45

Keywords: Data Mining; Classification; Hydrocarbon, Clustering, Decision Tree

INTRODUCTION

Hydrocarbon exploration takes place by following a dedicated procedure which utilizes specialized resources and resource persons. The presence of a hydrocarbon reserve beneath earth's surface needs a great deal of time, cost and expertise. The techniques used to predict hydrocarbon reserves starts from surface manifestations, proceeds through detailed magnetic, seismic and geological surveys and ends at predicting a wildcat which can either be a huge reserve or an empty reservoir in extreme considerations.

The use of data mining is not limited to hydrocarbon exploration; it has also produced some promising results in reservoir characterization, well logging, and in the calculation of various attributes such as shear-wear velocity [5]. With the exception of the expert's knowledge in geological survey, hydrocarbon exploration does not usually use mathematical models; rather, it is an interpretive science that requires a highly accurate prediction about the presence of hydrocarbons prior to digging into the Earth and mining. The limited diffusion of data mining in this area can be attributed to the lack of complete understanding, because of its emergence, of the knowledge discovery model, which spans across the steps of exploration, as well as the need to apply it to multi-natured data collected from diverse sources [7].

Hydrocarbon is found in natural gas, crude oil, bitumen, gasoline, petroleum, natural gas liquids, liquefied natural gas, and fluids. The earliest sediments of oil and gas were deposited into the Earth about 560 million years ago. The oil and gas industry has been evolving since the 19th century, when geological survey was the sole means by which to prospect oil and gas. However, the 19th and 20th century brought technological innovation which has since been used in exploration. Subsequently, correlating the prediction of known reservoirs with expected reservoirs became standard for geologists and explorers [2] and additional development in exploration techniques improved geologists' understanding of exploration. Despite this improvement, the exploration of hydrocarbons continues to be a risky endeavor which takes, on average, three to six years and costs millions of dollars [4].

The paper is organized in a systematic way. Section 2 will state the preliminaries of proposed framework. Section 3 will exhibit and discuss the final framework. Section 4 will conclude the work.

PROJECT DESCRIPTION

This project has been directed to provide a framework which utilizes intelligent data mining techniques on history data in order to reduce enormous cost and time. The project is basically divided in two parts;

1. Pre-prospection activities
2. Prospection activities

First part deals with identifying critical dimensions that are conclusive for hydrocarbon exploration and consumption. In this part, we have classified world nations w.r.t. sustainable hydrocarbon development.

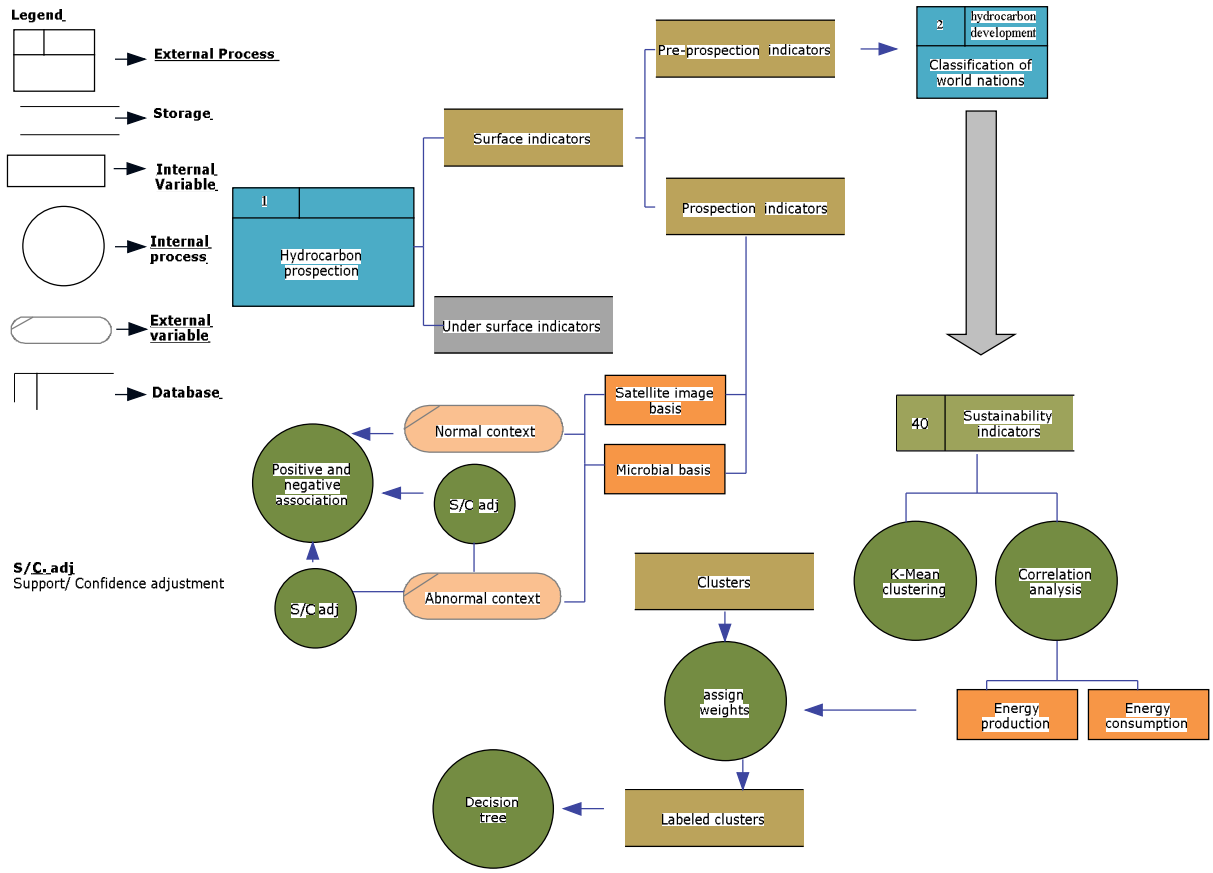

Figure 1 – Integrated Spatio Temporal Data Mining Framework for Natural Resource Exploration

We have then identified the dimensions on critical path. For this purpose, sustainability indicators are mined by proposing two innovative extensions in unsupervised classification methodology. Once the hydrocarbon development standing of a nation is identified, an optimal hydrocarbon prospection method may be proposed to fulfill consumption needs. In the proposed framework, the prospection is solely based upon data mining of current/ history data [6].

In existing scenario, hydrocarbon prospecting involves some typical earth engineering, mining and geological techniques. Data mining and its minion domains has the potential to be utilized on such data repositories. A successful application will obviously cut down the cost and time and it will also reduce the need of expertise. The data sources used in prospecting at current can be divided into 1. Surface indicators. 2. Under surface indicators. Surface indicators are initial manifestations for site drilling. In our study surface indicators are only considered to be included in spatio temporal data mining framework. In the detailed study only two data sources have been focused upon. 1. Remotely sensed satellite images. 2. Microbial data. These two data sources are changing the convention for initial decision. The dependency of initial decision of drilling a wildcat should remain on the patterns extracted from remotely sensed satellite image and microbial data.

The detailed literature survey of supervised and unsupervised classification techniques revealed that the use of association rule mining on both of these data sources will produce the best [3]. The reason is twofold;

1. The classification techniques divide datasets in multiple groups where each is tagged by a class label. Hydrocarbon prospecting is not about dividing the data into classes but to predict a value based on the fact that conditions are provided.
2. The decisions about hydrocarbon prospecting are rule based. Its equivalent modeling in computer science is IF-THEN-ELSE rules which can be obtained by using association rule mining.

Prospecting based on remotely sensed satellite image and microbial data is done by using association rule mining. In existing applications, association rule mining is used to mine positive rules

from databases only. Based on existing technique for positive and negative association rule mining, the rules are extracted from remotely sensed image whereas only positive association rules are taken from microbial data. Context is state of object, entity and environment and is not considered in mining association rules whereas it enormous effect on accuracy of association rules. For association rule mining the general Apriori algorithm [1] is being used all on spatial, non-spatial and spatio temporal data. We have proposed a new context based positive and negative association rule mining algorithm which is applied to remotely sensed and microbial data to predict a hydrocarbon reserve on the basis of associations among objects in satellite images and relationships among microbial indicators.

PROPOSED FRAMEWORK

In the figure 1, the complete framework for hydrocarbon exploration project is given. First, it is established that hydrocarbon exploration activity can be done with the help of surface and sub surface indicators. Sub surface indicators are not considered in this activity because of higher cost and manpower involved in it. Surface indicators are utilized both for pre-prospection and prospection tasks. In pre-prospection phase world nations are classified into five clusters which are labeled then in order to rank hydrocarbon development of that particular nation. In order to label clusters the sustainability indicators are correlated with energy consumption and production. These labeled clusters are then mapped on to a decision tree for supporting decision makers in identifying weak dimensions of hydrocarbon development. Once the decision tree is drawn, the prospection may be preceded. Prospection indicators can be elicited from various sources including geological attributes, microbial attributes and well logs. In the proposed framework only two data sources are considered in prospection phase from which maximum required data can be elicited. Context based association rule mining is applied both on remotely sensed and microbial data in different contextual situations. The support-confidence adjustments are being made according to the algorithm proposed in section 4.5.

The methodologies proposed for the above framework are used for hydrocarbon prospection but can also be used in other applications. The technique for labeling clusters is generic in nature and can be applied upon any dataset clusters. Similarly context variables and its integration in positive/negative association rule mining can be utilized on datasets which have the potential to reflect different patterns in different contextual situations [6].

CONCLUSION

In this paper, an integrated spatio temporal data mining framework is developed as a decision support system for hydrocarbon prospecting. The work is composed of diverse flavors because of following two reasons; The conventional process of hydrocarbon prospecting use data from diverse sources to make a prediction about hydrocarbon reserve. The proposed framework is aimed to utilize analytical techniques of data mining in the conventional hydrocarbon exploration framework. Since the data is from diverse sources, the intelligent techniques can better be utilized if the application of technique is considered in data source perspective. That is why; different techniques are utilized in all three phases. The practiced algorithms of data mining in its crude form might not be that much suitable to be applied for predicting type of energy development and preparing hydrocarbon prospecting plan. For this, new algorithms and extensions in existing algorithms are proposed. Since these algorithms are from diverse classes hence the framework for hydrocarbon exploration seems to be diverse which infact leads to single unified goal i.e. hydrocarbon prospecting.

ACKNOWLEDGEMENTS

The project is undertaken at Department of Computer Science & Engineering, University of Engineering & Technology Lahore, Pakistan. The data is collected from industrial sector of Pakistan.

REFERENCES

1. Agarwal, R., Srikant, R., 1994. Fast Algorithms for Mining Association Rules., Proceedings of the 20th VLDB Conference, Chile., Pp. 487-499.
2. Bott, R.D., 2004 .Evolution of Canada's Oil and Gas Industry. Canadian Center for Energy Information.
3. Han, J., Kamber, M., 2001. Data Mining: Concepts and Techniques. Morgan Kaufmann, New York. 550 pp. 284-291.
4. Jahn, F., Cook,M., Graham, M., 2003. Hydrocarbon Exploration and Production. Elsevier Science B.V. 9 - 15.
5. Justice, J.H., Hawkins, D.J., Wong, G., 1985 . Multidimensional Attribute Analysis and Pattern Recognition for Seismic Interpretation. Pattern Recognition. 18, 391-407.
6. Shaheen, M., Shahbaz, M., Guergachi, A., Rehman, Z., 2011. Mining Sustainability Indicators to classify Hydrocarbon Development., Elsevier Knowledge-based Systems., Vol. 24, No.8, pp. 1159-1168.
7. Shaheen, M., Shahbaz, M., Guergachi, A., Rehman, Z., 2010. Data Mining Applications in Hydrocarbon Exploration., Springer Artificial Intelligence Review., Vol. 35, No.1, pp. 1 – 18.

1/12/2012