An Emergence of AI in Data Mining and KDD: ANN its Strength & Weakness

Mohammad Asim khan, Sharik Ahmad

Abstract :Rapid increases in informational and technological methods have led companies to gather customer data in great databases. Data mining is the procedure involving analyzing, searching information to make it helpful for human use. Great amount of information is modeled, selected as well as explored in order to figure out comprehensible info. Artificial intelligence as well as data mining methods has been utilized in numerous domains to resolve classification, diagnosis, and association, segmentation, and prediction issues. This particular paper discuss about the Data mining as well as KDD procedure from an AI perspective. We talk about popular subjects in data mining as well as AI, which includes key AI suggestions which have been utilized in both data mining as well as KDD. The general goal of this particular paper is taking a perspective of different application and algorithms of AI in Data Mining. In this particular paper, we are going to take into consideration several historic contributions and talk about these various thoughts with an example to human intelligence. Additionally, an overall framework for the style of even more clever devices can we proposed.

Keywords: Artificial intelligence, Knowledge discovery, Neural network, Data Mining.

I. INTRODUCTION

Data mining is a method which utilizes an assortment of information analysis tools to find relationships & patterns in information which might be utilized to make legitimate predictions. The very first as well as simplest analytical stage of data mining is describing the information - summarize its statistical characteristics (such as standard deviations) and means, visually examine it using graphs and charts, and also search for likely substantial backlinks amongst variables. In the Data Mining selecting, exploring, collecting, and Process the proper information are crucial. Expertise Discovery demonstrates smart computing at its best, and is considered the most desired as well as fascinating end product of Information Technology. in order to have the ability to discover and also to extract knowledge from information is a process that lots of practitioners and researchers are endeavoring to achieve.

There's a great deal of concealed knowledge waiting to be found - this's the challenge produced by present day abundance of information. Expertise Discovery in Databases (KDD) is the procedure of determining legitimate, useful, novel, and also understandable patterns from big datasets.

Synthetic neural community ANN is yucky simplification of actual networks of neurons. The paradigm of neural network which started in the 1940's promises was being an extremely crucial tool for learning the structure function connection of human mind. Due to incomplete comprehension as well as the complexity of biological neurons different structure of synthetic neural network has been found in the literature. The target neural community is usually to mimic the human power to adjust to changing in circumstances as well as the present setting. In this paper I am going to discuss about neural networks are helpful for data mining and decision support applications.

Figure 1: Fields of Data Mining

II. DATA MINING AND KNOWLEDGE MANAGEMENT

Knowledge discovery and learning is definitely iterative process which stretches the compilation of data mining strategies right into a knowledge management framework. Higher education is going to find wider and larger uses for data mining than the counterpart of its in the business sector, since higher education institutions offer 3 that data mining comprehensive duties: scientific research which pertains to the development of knowledge, teaching that issues with the transmission of knowledge, along with institutional research which relates to the usage of knowledge for decision making. Each of the above jobs are best within the boundaries of Knowledge Management, and that moves the demand for faster and better decision creating strategies and tools. Owing to the strength of its,

Mohammad Asim khan, Research Scholar, Department of Computer Science and engineering Glocal University, Saharanpur, India. Email: asimk0068@gmail.com

Sharik Ahmad, Department of Computer science and engineering, Glocal university, Saharanpur India. Email: shariqgee@gmail.com

Data Mining is widely known as an effective Business Intelligence application for knowledge discovery. The procedure for Data Mining is a Knowledge Management activity since it requires human knowledge.

Many writers also have discussed the elements behind the dawn of data mining. For example, Therling (1995) identified 3 reasons: The simplicity of storage space and data collection, the computing power of modern processors, so the demand for quick and real time data mining. Nevertheless, one significant reason absent from these may be the increasing interest in Knowledge Management.

III. A FRAMEWORK FOR ARTIFICIAL INTELLIGENCE IN DATA MINING AND KDD

Minsky published A laptop is able to do, in a sense, only what it's told to do. But even if we don't understand precisely how to resolve a particular issue, we might system a machine to browse through several great rooms of solution attempts. Additionally, he published certainly we should utilize whatever we k immediately ahead of time to direct the trial generator. And we should also have the ability to make use of outcomes obtained in the process.

We ought to make ontology for Information and data science (DIS) and also make use of a method that undertakes cycles, as shown in Figure 2:
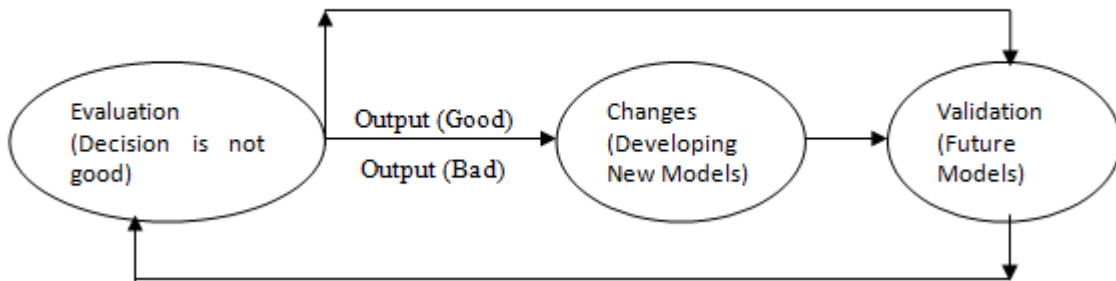

Figure 2: Improved cycle for the models.

The above mentioned ideas are usually applied from an ontology shown in Figure three. The answering machine might differ from machine to machine, it might become a classifier as, for example: k nearest neighbors (KNN), a fuzzy k means clustering technique utilized as a histogram or a classifier based set of classifiers, along with some other remedies. Nevertheless, it must be in a position to put in a brand new couple classifier/feature, each classifier actually being specialized with just one feature. This will aid in generalization as well as simplification of the jobs (decisions, mixture of outcomes, etc.).
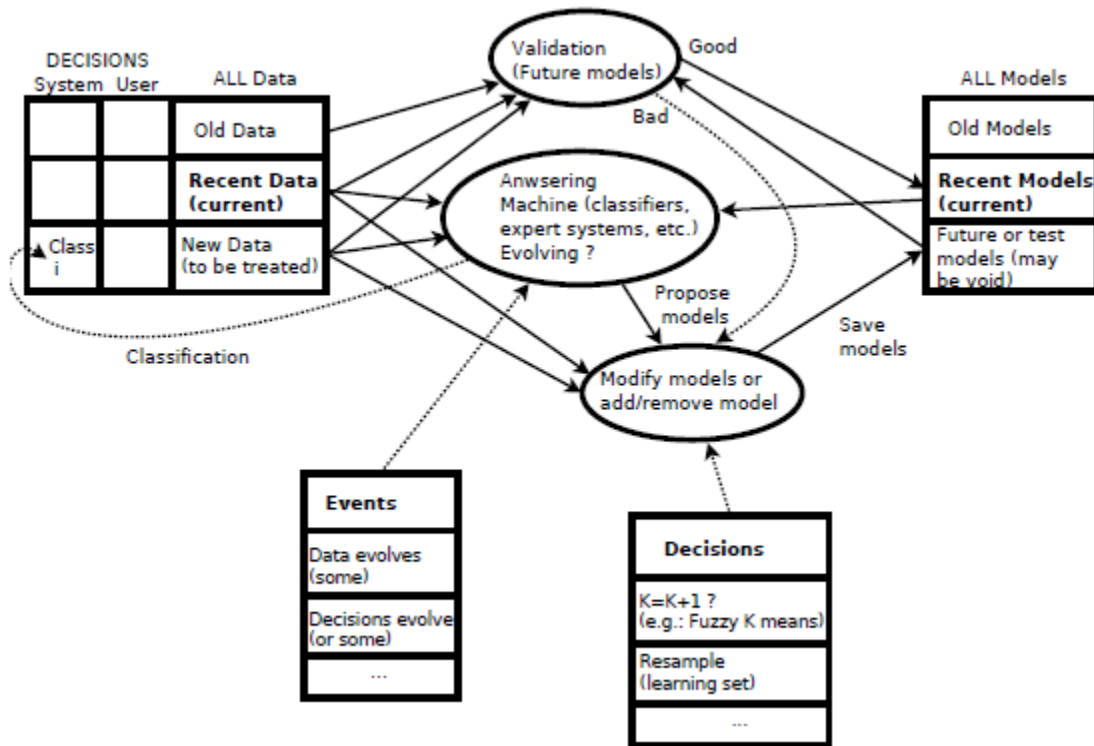

Figure 3: an ontology for a combining approach involving artificial intelligence

- Events (some) :
  - Entropy measure of results shows low entropy (poorness of decisions),
  - Local accuracy measures of last pattern classified
  - ROC curves.
- Decisions (some):
  - Change models by: change point

of view, mainly by Random Subspaces (statistical paradigms).

- o Random Forests (decision trees paradigm).
- o Change references: mainly by Bagging and Boosting
- o Add or remove a class.
- o Stop iterations on a time limit and/or loop limit or other criteria.
- o Add or remove a feature, if needed.

Evaluation is an info problem, because the answering machine must inform us the models are evolving. We'll most likely require various systems, potentially with hierarchies as well as recursive structures (idea shared with Minsky).

This particular technique that we suggest is akin to what Minsky had in your mind (seen following the original exploration of this particular paper), as he described an intelligent method able of:

- "Search" ("or "trial" an "error"),
- "Pattern recognition", nowadays, called (often) categorization or clustering
- "Learning",
- "Planning"
- "Induction"

IV. THREE FUNDAMENTAL AI TECHNIQUES IN DATA MINING

AI is a wider area compared to machine learning. AI methods are knowledge processing methods. Inference, knowledge acquisition, and knowledge representation including influence and search, are 3 fundamental strategies in AI.

Knowledge representation. Data mining seeks to find out interesting patterns from huge volumes of data. These patterns are able to take different types, like association rules, classification rules, as well choice trees, moreover consequently, knowledge representation grows into a problem of attention of data mining.

Knowledge acquisition. The discovery process shares different strategies and algorithms (Topics two as well as six) with machine learning for exactly the same goal of expertise acquisition from information or even discovering from examples.

Knowledge inference. The patterns determined from information have being confirmed with numerous programs so deduction of mining benefits is a vital method of data mining uses.

Thus, knowledge representation, knowledge acquisition as well as understanding inference, the 3 basic strategies in AI are applicable to data mining.

V. KEY METHODS SHARED IN AI AND DATA MINING

AI investigation is about the concepts as well as look of logical agents [Norvig and Russell 2003], and data mining methods are generally good examples of such logical representatives. Most AI investigate parts (such as reason, preparation, natural language processing, game playing as well as robotics) have concentrated on the improvement of heuristic and symbolic solutions to resolve complicated issues effectively. These techniques also have found considerable use of data mining.

Symbolic computation. A lot of data mining algorithms contend with symbolic values. As a situation of fact, since a lot of data mining algorithms have been created to mainly deal with symbolic values, discretization of constant characteristics is an important and popular subject of data mining for numerous years, so that those algorithms could be given to deal with both real-valued and symbolic characteristics.

Heuristic search. As in AI, numerous data mining troubles are NP hard, like constructing the perfect decision tree from a certain information set, and clustering a certain number of information objects into an optimum number of groups. Thus, heuristic search, divide and conquer, and also understanding acquisition from several sources are already typical strategies in equally data mining as well as machine learning.

For instance, Ross Quinlan's info gain as well as gain ratio techniques for choice tree building, which relies on an unethical search with divide and conquer, is created to both and also, that are most likely the most favored books in AI along with data mining respectively. Decision tree construction is able to make use of both real-valued and symbolic attributes.

Evolutional algorithms and neural networks (including hereditary algorithms) can also be discussed in numerous AI and data mining references.

VI. STRENGTHS AND WEAKNESSES OF NEURAL NETWORK MODELS

Neural networks display strong freedom in the face of the problems presented by the real life. Noisy feedback or destruction of devices causes elegant degradation of functionality. Based on the dynamics of the sturdiness and also the application of the internal information patterns you are able to typically expect a system to train very well. This is applicable to issues in which the relationships might be non-linear or dynamic quite.

Neural networks are common approximates, and work best when the device you're making use of them to item features a higher tolerance to error. The way they work perfectly for recording associations or maybe finding regularities inside a set of patterns, the place that the amount, quantity of variety or variables of the information is quite fantastic, the interactions between variables are vaguely understood & the friendships are hard to summarize properly with standard tactics. It uniquely helpful for big quantity of instance information can be obtained and it's tough to establish a parametric style for the data. There are likely sound patterns in the information which are small and profoundly concealed. Iterative use of the information is necessary to detect pattern. ANN getting sound features nevertheless it's a few limitations It discover by countless passes with the instruction established therefore the learning time of NN is normally extended. ANN can't expose the knowledge of its as symbolic rules. Accessible domain knowledge is quite hard to be integrated to a NN.

VII. HOW ANN USEFUL IN DATA MINING?

Data mining techniques are mainly based on statistics, in addition to machine learning as the patterns might be inferred from various data types. Techniques employed in data mining, for example machine learning, and belong on the area of artificial intelligence.

VIII. RESULT AND DISCUSSION

Data mining is an AI powered application that could learn info that is helpful in just a database which can subsequently be utilized to enhance measures. Information mining driven by AI algorithms is additionally proving to become a "great tool" for visiting brand new avenues to immediately analyze, visualize as well as uncover patterns in information. Data mining or even knowledge discovery is starting to be more essential as far more plus more corporate data is now being computerized. Intelligent applications, like genetic algorithms and neural networks are perfect for locating unknown information and trends from the huge quantities of computer information. Artificial Intelligence (AI) is able to assist with transform all of these information into organized, functional platforms. Neural networks are helpful for data mining and decision support programs, People are very good at generalizing from practical experience. Computers do well at following explicit directions again and again. Neural networks bridge this gap by modeling, on a laptop, the neural actions of human being brains.

IX. CONCLUSION

Today, an intelligent device will be able to do conceptual comparisons (as opposed to just numerical comparisons), for example: identify a berry that is sweeter compared to others. It must have the ability of revolutions, like to resolve "this is a brand new group or maybe cluster", for example: this's an innovative illness type (unknown or unseen before). They ought to be in a position to modify the knowledge base of theirs, for instance able to modify the models of theirs or even add news models. Inside NLP (Turing machine), they ought to be ready to control various fields and also alter or even add new fields (as people are able of).The area of certain apps aims to acquire certain awareness. The domain experts by thinking about the user"s requirements along with other context parameters guide the product. The smart interfaces as well as smart agents around some degree help make the program generic but have limitations. The domain professionals play role that is crucial in the various phases of data mining. The choices at various phases are affected by the elements as area as well as information specifics, goal of the data mining, as well as the context parameters. ANN is helpful particularly when there's no a priori knowledge regarding the examined data. The Initiatives in artificial intelligence bringing about important brand new applications and services.

REFERENCES

1. K. Hilvska, P. Knocz. "Application of Artificial Intelligence and Data Mining Techniques to Financial Markets", ACTA, vol. 6, no. 1, 2012.
2. G. Kumar. Artificial neural networks and its applications, 2007, Available: http://www.iasri.res.in/ebook/EBADAT/5Modeling%20and%20Forecasting%20Techniques%20in%20Agriculture/5-ANN_GKJHA_2007.pdf Accessed date 05 March 2015
3. M. V. Joseph, "Significance of data warehousing and data mining in business applications", International Journal of Soft Computing and Engineering (IJSCE), New York vol. 3, no. 1, pp. 329-333, Mar. 2013. .
4. R. Abdullah, M. G. Hassan. "Advanced Manufacturing Technology: The Perceived Impact of Producer's Value". The 3rd International Conference on Technology and Operations Management. Bandung Indonesia, 4-6 July, 2012, pp. 23-30
5. R. Pfeifer, F. Idia. "Embodied Artificial Intelligence: Trends and Challenges. Available: http://people.csail.mit.edu/iida/papers/PfeiferIidaEAIDags.pdf Accessed date 05 March 2015
6. S. D. Mohaghegh. "Artificial Intelligence and Data Mining: Enabling Technology for Smart Fields". vol. 5, no. 3. 2009
7. Shi, C., Verhagen, M. &Pustejovsky, J. (2014). A Conceptual Framework of Online Natural Language Processing Pipeline Application, Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT, Dublin, Ireland. 53–59. http://www.aclweb.org/anthology/W14-5206
8. Şentürk, A. (2006). VeriMadenciliği: KavramveTeknikler, 2. Bursa: Ekin Yay.
9. Tan, A.H., Yu, P.S. (2004), Guest Editorial: Text and Web Mining, Applied Intelligence 18, 239-241.
10. Wang, L., Wang, G. & Alexander, C.A. (2015). Big Data and Visualization: Methods, Challenges and Technology Progress, Digital Technologies, 1(1), 33-38. http://pubs.sciepub.com/dt/1/1/7

AUTHOR PROFILE

Myself **Mohammad Asim** khan currently pursuing Ph.D from Glocal university, saharanpur, India. I have completed my Master in Technology from Manav Bharti University, solan, Himachal Pardesh. My Thesis in M.tech" Automatic Document Classification: **Designed to allow users working with files and Documents to rely on the system to classify and store them with a title manual Attention.**

Dr. Shariq Ahmad awarded his Ph.D in BigData Analytics/Data Mining/Sentiment Analysis with research topic **Sentiment Analysis of mobile phones using revised Data Dictionary and Polarity** from the Glocal university, Saharanpur in 2019. Dr. Ahmad has completed his master of computer Application (MCA) from Jamia Millia Islamia (central university) delhi in 2013. Dr. Ahmad has also qualified UGC NET. Moreover he secured first position in B.sc.(H) Mathematics and received Gold medal from jamia millia Islamia.