Combination of Association and Neural Networks Techniques: A Review

Ishan Juneja
Department of Production & Industrial Engineering
PEC University of Technology, Chandigarh

Abstract— Data Mining and its techniques have been in huge demand from past decades due to sudden rise of automation and robotics involvement in industrial applications. Association of events to form a useful pattern, sequencing of items in a trending form or usage of Artificial Neural Networks to extract hidden relations or anomalies etc. are the various data mining activities done by researchers. This review paper explains the various techniques and designing of models formed using association and Neural Network Techniques to form useful patterns or trends between input and output variables.

Key words: Association Rule Mining, Artificial Neural Networks, Data Mining, Auto Associative Networks

I. INTRODUCTION

Data mining has been a rapidly growing field of research and practical applications during the last two decades. From a somewhat niche academic area at the intersection of machine learning and statistics it has developed into an established scientific discipline and a highly valued branch of the computing industry. This is reflected by data mining becoming an essential part of computer science education as well as the increasing overall awareness of the term “data mining” among the general (not just computing-related) academic and business audience.

Data mining can be viewed as an analytic process that uses one or more available datasets from the same domain to create one or more models for the domain, i.e., models that can be used to answer queries not just about instances from the data used for model creation, but also about any other instances from the same domain. For e.g., if some attributes are generally available (observable) and some attributes are only available on a limited dataset (hidden), then models can often be viewed as delivering predictions of hidden attributes wherever their true values are unavailable. The unavailable attribute values to be predicted usually represent properties or quantities that are hard and costly to determine, or (more typically) that become known later than are needed. The latter justifies the term “prediction” used when referring to a model’s output. The attribute to be predicted is referred to as the target attribute, and the observable attributes that can be used for prediction are referred to as the input attributes.

A. Data Mining Tasks

The most common types of predictive models – or queries they can be used to answer – correspond to the following three major data mining tasks:

1) Classification.

Predicting a discrete target attribute (representing the assignment of instances to a fixed set of possible classes). This could be distinguishing between good and poor customers or products, legitimate and fraudulent credit card transactions or other events, assigning failure types and recommended repair actions to faulty technical devices, etc.

2) Regression.

Predicting a numeric target attribute which represents some quantity of interest. This could be an outcome or a parameter of an industrial process, an amount of money earned or spent, a cost or gain due to a business decision, etc.

3) Clustering.

Predicting the assignment of instances to a set of similarity-based clusters. Clusters are not predetermined, but discovered as part of the modeling process, to achieve possibly high intra-cluster similarity and possibly low inter-cluster similarity.

Data mining techniques have their roots in two fields: machine learning and statistics. With the former traditionally addressing the issue of acquiring knowledge or skill from supplied training information and the latter the issue of describing the data as well as identifying and approximating relationships occurring therein, they both have contributed modeling algorithms. They have also become increasingly closely related, which makes it difficult and unnecessary to put hard separating boundaries between them.

B. R programming

R is an increasingly popular programming language and environment for data analysis, sometimes referred to as the “lingua franca” of this domain, with a huge set of contributed packages available from the CRAN repository, providing implementations of various analytic algorithms and utility functions.

R and its libraries implement a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others. R is easily extensible through functions and extensions, and the R community is noted for its active contributions in terms of packages. Many of R's standard functions are written in R itself, which makes it easy for users to follow the algorithmic choices made. For computationally intensive tasks, C, C++, and Fortran code can be linked and called at run time. Advanced users can write C, C++, Java, .NET or Python code to manipulate R objects directly. R is highly extensible through the use of user-submitted packages for specific functions or specific areas of study. Due to its S heritage, R has stronger object-oriented programming facilities than most statistical computing languages. Extending R is also eased by its lexical scoping rules.

Another strength of R is static graphics, which can produce publication-quality graphs, including mathematical symbols. Dynamic and interactive graphics are available through additional packages.

II. ASSOCIATION RULE MINING

Association is mostly used for decision making with the measures such as support and confidence. Association is used to find patterns in large data. Association helps in business to make a decision in marketing and other fields. Decision making is most important in association rule mining. Association rule mining is to discover available association rules that fulfill the predefined least possible support and confidence from a supposed database.

Two types of association rule mining are used in large databases. One is positive association rule mining. The second one is negative association rule mining. Mining negative association rules acting as vigorous role in decision making consequence. Association rule mining seeks to discover associations among transactions encoded in a database. It can be used to improve decision making in a wide variety of applications such as: medical diagnosis, GIS, relational database, large database and distributed database etc. These databases are reviewed. This study discusses about how the association rule mining used in different databases.

Figure 1 refers to the association rules data mining architecture.

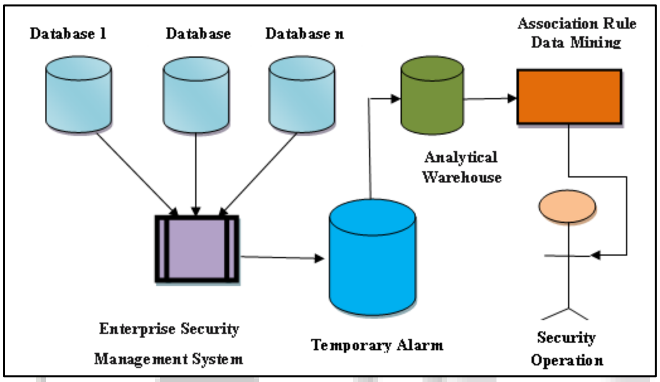


Fig. 1: The Association Rules Data Mining Architecture

The alarms arrive at the Security Operations Center (SOC), there SOC Analyst to analysis data from association rule data mining; they are stored in the short term in a database on the monitoring database. From this temporary database extracted the set of all alarms generated in a single day for all databases and loaded them into a analytical data warehouse. It is on this warehouse that executed the data mining algorithms with the goal of producing new checking rules for installation in the ESM (Enterprise Security Management System). That checks the data from different databases.

Association rule mining performing vital role in the essential area data mining. It poses many exciting problems for the enlargement of well-organized and effective techniques. Applications of association rule mining are Large and Distributed database – Businesses, e.g. logistics, marketing and Government – almost all branches e.g. defense, public safety, Spatial database – GIS, Relational database – Industries, Medical database- Medical diagnosis, Hospital, Medical shops, scan centers… Future work is to find out the better support and confidence of different algorithms with association rule mining.

A. Limitations of Association Rule Mining

In the association rule mining area, most of the research efforts went in the first place to improving the algorithmic performance, and in the second place into reducing the output set by allowing the possibility to express constraints on the desired results. Over the past decade a variety of algorithms that address these issues through the refinement of search strategies, pruning techniques and data structures have been developed. While most algorithms focus on the explicit discovery of all rules that satisfy minimal support and confidence constraints for a given dataset, increasing consideration is being given to specialized algorithms that attempt to improve processing time or facilitate user interpretation by reducing the result set size and by incorporating domain knowledge. There are also other specific problems related to the application of association rule mining from e-learning data. When trying to solve these problems, one should consider the purpose of the association models and the data they come from. Nowadays, normally, data mining tools are designed more for power and flexibility than for simplicity. Most of the current data mining tools are too complex for educators to use and their features go well beyond the scope of what an educator might require. As a result, the courses administrator is more likely to apply data mining techniques in order to produce reports for instructors who then use these reports to make decisions about how to improve the student's learning and the online courses. However, it is most desirable that teachers participate directly in the iterative mining process in order to obtain more valuable rules. But normally, teachers only use the feedback provided by the obtained rules to make decisions about modification to improve the course, detect activities or students with problems, etc.

Some of the main drawbacks of association rule algorithms in e-learning are: the used algorithms have too many parameters for somebody non-expert in data mining and the obtained rules are far too many, most of them non-interesting and with low comprehensibility. In the following subsections, we will tackle these problems. 3.1 Finding the appropriate parameter settings of the mining algorithm Association rule mining algorithms need to be configured before to be executed. So, the user has to give appropriate values for the parameters in advance (often leading to too many or too few rules) in order to obtain a good number of rules. A comparative study between the main algorithms that are currently used to discover association rules can be found in: Apriori, FP-Growth, Magnum Opus, Closet. Most of these algorithms require the user to set two thresholds, the minimal support and the minimal confidence, and find all the rules that exceed the thresholds specified by the user. Therefore, the user must possess a certain amount of expertise in order to find the right settings for support and confidence to obtain the best rules. One possible solution to this problem can be to use a parameter-free algorithm or with less parameters. For example, the Weka package implements an Apriori type algorithm that solves this problem partially. This algorithm reduces iteratively the minimum support, by a factor delta support (Δs) introduced by the user, until a minimum support is reached or a required number of rules (NR) has been generated. Drawbacks and solutions of applying association rule mining 19 Another improved version of the Apriori algorithm is the Predictive Apriori algorithm, which automatically resolves the problem of balance between these two parameters, maximizing the probability of making an accurate prediction for the data set. In order to achieve this, a parameter called the exact expected predictive accuracy is defined and calculated using the Bayesian method, which provides information about the accuracy of the rule found. In this way, the user only has to specify the maximal number or rules to discover. In experimental tests were performed on a Moodle course by comparing the two previous algorithms. The final results demonstrated better performance for Predictive Apriori than Apriori-type algorithm using the Δs factor. 3.2 Discovering too many rules. The application of traditional association algorithms will be simple and efficient. However, association rule mining algorithms normally discover a huge quantity of rules and do not guarantee that all the rules found are relevant. Support and confidence factors can be used for obtaining interesting rules which have values for these factors greater than a threshold value. Although these two parameters allow the pruning of many associations, another common constraint is to indicate the attributes that must or cannot be present in the antecedent or consequent of the discovered rules. Another solution is to evaluate, and post-prune the obtained rules in order to find the most interesting rules for a specific problem. Traditionally, the use of objective interestingness measures has been suggested, such as support and confidence, mentioned previously, as well as others measures such as Laplace, chi-square statistic, correlation coefficient, entropy gain, gini, interest, conviction, etc. These measures can be used for ranking the obtained rules in order than the user can select the rules with highest values in the measures that he/she is more interested. Subjective measures are becoming increasingly important, in other words measures that are based on subjective factors controlled by the user. Most of the subjective approaches involve user participation in order to express, in accordance with his or her previous knowledge, which rules are of interest. Some suggested subjective measures are:

– Unexpectedness: Rules are interesting if they are unknown to the user or contradict the user's knowledge.
– Actionability: Rules are interesting if users can do something with them to their advantage. The number of rules can be decreased by only showing unexpected and actionable rules to the teacher and not all the discovered rules.

In an Interestingness Analysis System (IAS) is proposed. It compares rules discovered with the user's knowledge about the area of interest. Let U be the set of user's specifications representing his/her knowledge space, A be the set of discovered association rules, this algorithm implements a pruning technique for removing redundant or insignificant rules by ranking and classifying them into four categories: • Conforming rules: a discovered rule Ai∈A conforms to a piece of user's knowledge Uj if both the antecedent and the consequent parts of Ai match those of Uj∈U well.

III. ARTIFICIAL NEURAL NETWORKS

The study of the human brain is thousands of years old. With the advent of modern electronics, it was only natural to try to harness this thinking process. The first step toward artificial neural networks came in 1943 when Warren McCulloch, a neurophysiologist, and a young mathematician, Walter Pitts, wrote a paper on how neurons might work. They modeled a simple neural network with electrical circuits. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. Other advantages include:

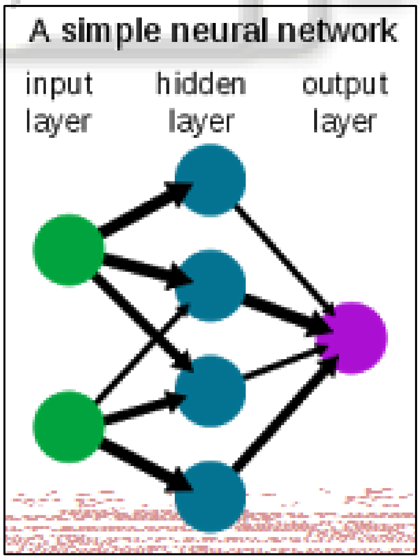
1) Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2) Self-Organization: An ANN can create its own organization or representation of the information it receives during learning time.
3) Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.
4) Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage.

Neural networks take a different approach to problem solving than that of conventional computers. Conventional computers use an algorithmic approach i.e. the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem. That restricts the problem-solving capability of conventional computers to problems that we already understand and know how to solve. But computers would be so much more useful if they could do things that we don't exactly know how to do. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable. On the other hand, conventional computers use a cognitive approach to problem solving; the way the problem is to solved must be known and stated in small unambiguous instructions. These instructions are then converted to a high-level language program and then into machine code that the computer can understand. These machines are totally predictable; if anything goes wrong is due to a software or hardware fault. Neural networks and conventional algorithmic computers are not in competition but complement each other. There are tasks are more suited to an algorithmic approach like arithmetic operations and tasks that are more suited to neural networks. Even more, a large number of tasks, require systems that use a combination of the two approaches (normally a conventional computer is used to supervise the neural network) in order to perform at maximum efficiency.

A. *What is Artificial Neural Network?*

Artificial Neural Networks are relatively crude electronic models based on the neural structure of the brain. The brain basically learns from experience. It is natural proof that some problems that are beyond the scope of current computers are indeed solvable by small energy efficient packages. This brain modeling also promises a less technical way to develop machine solutions. This new approach to computing also provides a more graceful degradation during system overload than its more traditional counterparts. These biologically inspired methods of computing are thought to be the next major advancement in the computing industry. Even simple animal brains are capable of functions that are currently impossible for computers. Computers do rote things well, like keeping ledgers or performing complex math. But computers have trouble recognizing even simple patterns much less generalizing those patterns of the past into actions of the future. Now, advances in biological research promise an initial understanding of the natural thinking mechanism. This research shows that brains store information as patterns. Some of these patterns are very complicated and allow us the ability to recognize individual faces from many different angles. This process of storing information as patterns, utilizing those patterns, and then solving problems encompasses a new field in computing. This field, as mentioned before, does not utilize traditional programming but involves the creation of massively parallel networks and the training of those networks to solve specific problems. This field also utilizes words very different from traditional computing, words like behave, react, self-organize, learn, generalize, and forget.

Whenever we talk about a neural network, we should more popularly say ―Artificial Neural Network (ANN)‖, ANN are computers whose architecture is modelled after the brain. They typically consist of hundreds of simple processing units which are wired together in a complex communication network. Each unit or node is a simplified model of real neuron which sends off a new signal or fires if it receives a sufficiently strong Input signal from the other nodes to which it is connected.



Traditionally neural network was used to refer as network or circuit of biological neurons, but modern usage of the term often refers to ANN. ANN is mathematical model or computational model, an information processing paradigm i.e. inspired by the way biological nervous system, such as brain information system. ANN is made up of interconnecting artificial neurons which are programmed like to mimic the properties of m biological neurons. These neurons working in unison to solve specific problems. ANN is configured for solving artificial intelligence problems without creating a model of real biological system. ANN is used for speech recognition, image analysis, adaptive control etc. These applications are done through a learning process, like learning in biological system, which involves the adjustment between neurons through synaptic connection. Same happen in the ANN.

IV. Combination of Both Techniques

In "An expert system for detection of breast cancer based on association rules and neural network", Murat Karabatak, M. Cevdet Ince have used association rule mining and neural networks for detection of cases of breast cancer.

In order to see how AR can be used in breast cancer data with NN, first of all it is needed to define AR. AR find interesting associations and/or relationships among large set of data items. AR show attributes value conditions that occur frequently together in a given dataset. They allow capturing all possible rules that explain the presence of some attributes according to the presence of other attributes. A typical and widely-used example of association rule mining is Market Basket Analysis (Agrawal et al., 1993).

A. Apriori algorithm

The Apriori algorithm is a state of the art algorithm most of the association rule algorithms are somewhat variations of this algorithm (Agrawal et al., 1993). The Apriori algorithm works iteratively. It first finds the set of large 1-item sets, and then set of 2-itemsets, and so on. The number of scan over the transaction database is as many as the length of the maximal item set. Apriori is based on the following fact: The simple but powerful observation leads to the generation of a smaller candidate set using the set of large item sets found in the previous iteration.

Apriori first scans the transaction databases D in order to count the support of each item i in I, and determines the set of large 1-itemsets. Then, iteration is performed for each of the computation of the set of 2-itemsets, 3-itemsets, and so on.

In this study, an automatic diagnosis system for detecting breast cancer based on association rules (AR) and neural network (NN) is presented. Feature extraction is the key for pattern recognition and classification. The best classifier will perform poorly if the features are not chosen well. A feature extractor should reduce the feature vector to a lower dimension, which contains most of the useful information from the original vector. So, AR is used for reducing the dimension of breast cancer database and NN is used for intelligent classification. The proposed AR + NN system performance is compared with NN model. The dimension of input feature space is reduced from nine to four by using AR. In test stage, 3-fold cross validation method was applied to the Wisconsin breast cancer database to evaluate the proposed system performances. The correct classification rate of proposed system is 95.6% for four inputs and 97.4% for eight inputs. This research demonstrated that the AR can be used for reducing the dimension of feature vector and proposed AR + NN model can be used to obtain efficient automatic diagnostic systems for other diseases.

V. CONCLUSIONS

Both association rule mining and neural networks have their own benefits and limitations. The combination of the two results in a unique and more relevant technique to extract patterns and hidden anomalies or trends in big datasets. This technique can also be used to predict the association of events with critical alarms or stoppages in a machine which have low frequency of occurrence in large data. Thus, this review can be used for a new research technique of using both algorithms together.

REFERENCES

[1] Ms. Sonali. B. Maind and Ms. Priyanka Wankar, Research Paper on Basic of Artificial Neural Network, International Journal on Recent and Innovation Trends in Computing and Communication, Vol 2, Issue 1, Pg. No. 96-100, January 2014.

[2] Dr. M. Renuka Devi and Mrs. A. Baby Sarojini, Applications of Association Rule Mining In Different Databases, Journal of Global Research in Computer Science, Vol 3, No. 8, Pg. No. 30-34, August 2014.

[3] Murat Karabatak and M. Cevdet Ince, An expert system for detection of breast cancer based on association rules and neural network, Expert Systems with Applications, Vol. 36, Pg. No.-3465-3469, 2009.

[4] Heling Jiang, An Yang, Fengyun Yan and Hong Miao, Research on Pattern Analysis and Data Classification Methodology for Data Mining and Knowledge Discovery, International Journal of Hybrid Information Technology Vol.9, No.3, Pg. No. 179-188, 2016.

[5] Sung-Bae Cho, Neural-Network Classifiers for Recognizing Totally Unconstrained Handwritten Numeral, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 8, NO. 1, JANUARY 1997.

[6] Enrique García, Cristóbal Romero, Sebastián Ventura, Toon Calders, Drawbacks and solutions of applying association rule mining in learning management systems