Aleem Ali, Naresh Kumar, Sanyogita Chouhan / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 3, May-Jun 2012, pp.1417-1421

ANN: A Novel Technique in Data Mining

Aleem Ali[1], Naresh Kumar[1], Sanyogita Chouhan[2]
1). F/O Engg, University Polytechnic, JMI New Delhi-110025
2). M.Tech(CS) Scholar, Department Of Computer Science, JamiaHamdard(Hamdard University)-110065

Abstract
Data mining is a multidisciplinary field, can be described as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from dataand the science of extracting useful information from large data sets or databases. We present technique for the discovery of patterns hidden *in large data sets*, focusing on issues relating to their feasibility, usefulness, effectiveness, and scalability.Predictive data mining is the most common type of data mining and one that has the most direct business applications.Thispaperisanoverviewofartificialneuralnetworks and questions theirpositionas apreferred toolbydata mining practitionersandthekeytechnologyandwaysto achievethedatamining basedonneuralnetworksarealsoresearched.

Keywords:Artificial NeuralNetwork (ANN), Advantages, Back propagation algorithm,Data mining,Data modeling.

1. INTRODUCTION
Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related) in search of consistent patterns and systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data[1]. The ultimate goal of data mining is knowledge discovery and prediction. Knowledge discovery provides explicit information that has a readable form and can be understood by a user[2]. Forecasting, or predictive modeling provides predictions of future events and may be transparent and readable in some approaches (e.g. rule based systems) and opaque in others such as neural networks. Moreover, some data mining systems such as neural networks are inherently geared towards prediction and pattern recognition, rather than knowledge discovery[3].

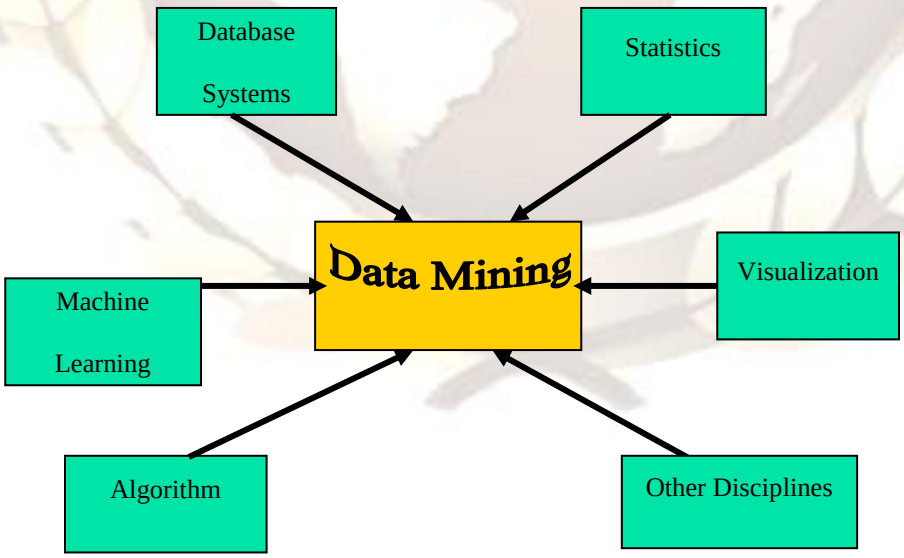


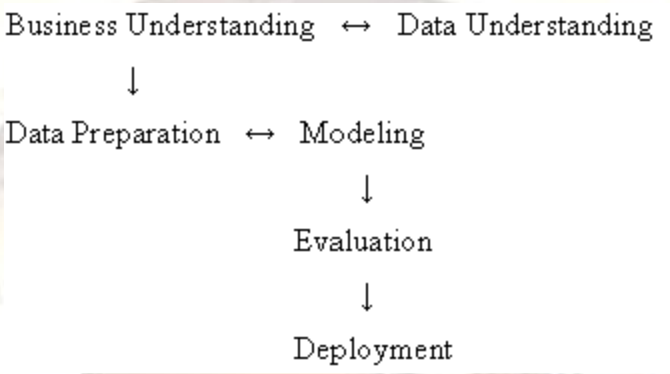
Fig. 1

Data mining relies on the use of real world data(typically in business applications). This data is extremely vulnerable to collinearity precisely because data from the real world may have unknown interrelations.

2. MODELS FOR DATA MINING

In the business environment, complex data mining projects may require the coordinate efforts of various experts, stakeholders, or departments throughout an entire organization. In the data mining literature, various "general frameworks" have been proposed to serve as blueprints for how to organize the process of gathering data, analyzing data, disseminating results, implementing results, and monitoring improvements.

One such model, CRISP (Cross-Industry Standard Process for data mining) was proposed in the mid-1990s by a European consortium of companies to serve as a non-proprietary standard process model for data mining. This general approach postulates the following (perhaps not particularly controversial) general sequence of steps for data mining projects:



Another approach - the Six Sigma methodology - is a well-structured, data-driven methodology for eliminating defects, waste, or quality control problems of all kinds in manufacturing, service delivery, management, and other business activities. This model has recently become very popular (due to its successful implementations) in various American industries, and it appears to gain favor worldwide. It postulated a sequence of, so-called, DMAIC steps -

Define → Measure → Analyze → Improve → Control

that grew up from the manufacturing, quality improvement, and process control traditions and is particularly well suited to production environments (including "production of services," i.e., service industries).

Another framework of this kind (actually somewhat similar to Six Sigma) is the approach proposed by SAS Institute called SEMMA -

Sample → Explore → Modify → Model → Assess

which is focusing more on the technical activities typically involved in a data mining project.All of these models are concerned with the process of how to integrate data mining methodology into an organization, how to "convert data into information," how to involve important stake-holders, and how to disseminate the information in a form that can easily be converted by stake-holders into resources for strategic decision making.

3. ARTIFICIAL NEURAL NETWORKS

Neural Networks are analytic techniques modeled after the (hypothesized) processes of learning in the cognitive system and the neurological functions of the brain and capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called *learning* from existing data[4]. Neural networks is one of the Data mining techniques.

Some authors stress the fact that *neural networks* use, or we should say are expected to use, massively parallel computation models. For example Haykin (1994) defines *neural network* as: *"a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: (1) Knowledge is acquired by the network through a learning process, and (2) Interneuron connection strengths known as synaptic weights are used to store the knowledge".*

The first step is to design a specific network architecture (that includes a specific number of "layers" each consisting of a certain number of "neurons")[5]. The size and structure of the network needs to match the nature (e.g., the formal complexity) of the investigated phenomenon. Because the latter is obviously not known very well at this early stage, this task is not easy and often involves multiple "trials and errors." (Now, there is, however, neural network software that applies artificial intelligence techniques to aid in that tedious task and finds "the best" network architecture.)

The new network is then subjected to the process of "training." In that phase, neurons apply an iterative process to the number of inputs (variables) to adjust the weights of the network in order to optimally predict (in traditional terms, we could say find a "fit" to) the sample data on which the "training" is performed. After the phase of learning from an existing data set, the new network is ready and it can then be used to generate predictions.

One of the major advantages of *neural networks* is that, theoretically, they are capable of approximating any continuous function, and thus the researcher does not need to have any hypotheses about the underlying model, or even to some extent, which variables matter. An important disadvantage, however, is that the final solution depends on the initial conditions of the network, and, as stated before, it is virtually impossible to "interpret" the solution in traditional, analytic terms, such as those used to build theories that explain phenomena.

Neural networks are essentially comprising three pieces: the architecture or model; thelearning algorithm; and the activation functions. (Fausett (1994)) Neural networks are programmed or "trained" to " store, recognize, and associatively retrieve patterns or database entries; tosolve combinatorial optimization problems; to filter noise from measurement data; to control illdefinedproblems; in summary, to estimate sampled functions when we do not know the form of thefunctions" (Kosko (1992), p.13) . It is precisely have these two abilities pattern recognition and function.

4. NEURAL NETWORK IN DATA MINING

Neuralnetworkmethodis usedforclassification,clustering, featuremining,predictionandpatternrecognition[6].Itimitates theneuronsstructureof animals,basesontheM-Pmodeland Hebb'slearningrule,soinessenceit isa distributedmatrix structure.Throughtrainingdatamining,the neuralnetwork method gradually calculates(includingrepeatediterationor cumulativecalculation)theweightsthe neuralnetwork connected.Theneuralnetworkmodelcanbebroadlydivided intothefollowingthreetypes:

(1) Feed-forward networks: it regards the perception back-propagationmodelandthe functionnetworkas representatives,andmainlyusedin theareassuchasprediction andpatternrecognition;

(2)Feedback network:itregardsHopfielddiscretemodel andcontinuousmodelasrepresentatives,andmainlyusedfor associativememoryandoptimizationcalculation;

(3) Self-organization networks: it regards adaptive resonancetheory(ART)modeland Kohonenmodelas representatives,andmainlyusedforclusteranalysis.

At present,the neuralnetworkmostcommonlyusedindata miningisBPnetwork[7].Ofcourse,artificialneuralnetworkis thedevelopingscience,andsometheorieshave notreallytaken shape,suchasthe problemsof convergence,stability,local minimumandparametersadjustment.FortheBPnetworkthe frequentproblemsitencounteredarethatthetrainingisslow, mayfallintolocalminimumandit is difficulttodetermine trainingparameters.Aimingat theseproblemssomepeople adoptedthemethodofcombiningartificialneuralnetworksand geneticgenealgorithmstoachievebetterresults.

Artificial neural network has the characteristics of distributed information storage, parallel processing, information,reasoning,andself-organizationlearning,andhas thecapabilityofrapidfittingthenon-lineardata,so itcansolve manyproblemswhicharedifficultforothermethodstosolve.

4.1 FEEDFORWARDNEURAL NETWORK

One ofthesimplestfeedforward neural networks (FFNN),suchas inFigure,consistsof threelayers:aninputlayer, hiddenlayerandoutput layer[8]. In each layer there are one or more processing elements (PEs). PEs is meant to simulatetheneuronsinthe brainandthisiswhy theyareoften referredtoas neuronsor nodes.A PE receivesinputsfromeitherthe outsideworldorthepreviouslayer.Thereareconnections between the PEs in each layer that have a weight (parameter) associated with them. This weight is adjusted duringtraining. Informationonly travels inthe forwarddirectionthroughthenetwork-thereare no feedback loops.

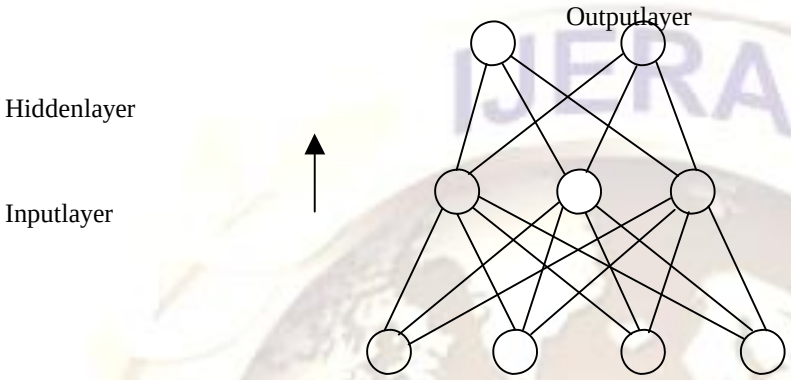


Fig 2.Multilayered feed-forwardneuralnetwork(ANN)

ThesimplifiedprocessfortrainingaFFNNisas follows:

1. Inputdata is presented tothe network and propagatedthroughthe networkuntilitreaches the output layer.Thisforward process produces a predicted output.
2. The predicted output is subtracted from the actual output and an error value for the networks is calculated.
3. The neural networkthenusessupervisedlearning, whichinmostcasesisback propagation, totrainthenetwork.Backpropagationisa learning algorithm for adjustingthe weights. Itstartswiththe weights betweenthe outputlayer PE's andthe last hiddenlayer PE's and works backwards throughthe network.
4. Once back propagation has finished, the forward process starts again, and this cycle is continueduntiltheerrorbetweenpredictedandactual outputsis minimized.

4.2. THE BACK PROPAGATIONALGORITHM:

Backpropagation,or**propagationoferror**,isa common method of teaching artificial neural network how to performa given task.The back propagation algorithm isused inlayered feed- forward ANNs. This means that the artificial neurons areorganized in layers, and send their signals "forward", and then theerrorsare propagated backwards[9]. The back propagation algorithmusessupervisedlearning,whichmeans thatweprovidethealgorithmwithexamplesofthe inputs and outputs wewant the network to compute, and then the error (difference between actualandexpectedresults)iscalculated.Theidea ofthebackpropagationalgorithmistoreducethis error,until theANN*learns* the training data[10]. **Summary ofthe technique:**

1. Presenta training sample tothe neuralnetwork.
2. Comparethenetwork'soutputtothedesired outputfrom thatsample.Calculatetheerrorin eachoutputneuron.
3. For eachneuron, calculate what the output shouldhavebeen,anda*scalingfactor*, how muchlower orhigherthe outputmustbe adjustedtomatchthedesiredoutput.Thisis the local error.
4. Adjusttheweightsofeachneurontolowerthe local error.

ActualAlgorithm:

1. Initialize the weights in the network (oftenrandomly)
2. repeat
   *for each example e inthe training setdo
      1.O=neural-net-output(network,e);
        forwardpass

2. T=teacheroutputfor e
3.Calculateerror(T- O)attheoutput units
4. Computedelta_wi for allweights fromhiddenlayer tooutputlayer ;
backward pass
5. Computedelta_wi for allweights frominput layer to hidden layer ; backward passcontinued
6.Updatetheweightsinthenetwork
*end
3. until all examples classified correctly or stoppingcriterion satisfied
4.return(network)

5. ADVANTAGES

Artificial neural networks(ANN) are just one of the tools used to find patterns in the data and to infer rules from them. Neural networksare useful in providing information on associations, classifications, clusters, andforecasting. Both the IRS and Wrangler have used neural networks in a data miningsituations with good success. More examples would have been given, but the Internetsearch of data mining and neural networks revealed only these cases. We anticipate as timepasses, and data mining grows more case studies will become available.We have also seen that for best results with neural networks a working knowledge ofstatistical models is desired. With all the common material between the two disciplines,neural networks and statistics, better communication between them would be advantageousto both. Computers have a long way to go before they can rival the human brain on thesame parallel scale but neural networks are a start in the right direction.

REFERENCES

[1] Aleem Ali, A Concise Artificial Neural Network in Data Mining, *International Journal of Research in Engineering & Applied Sciences, Feb 12, 2(2),*p. 418-428, 2012.

[2] O. Maimon and Rokach (eds.), Data Mining and Knowledge Discovery Handbook: *A Complete Guide for Practitioners and Researchers(*Kluwer Academic Publishers, 2007).

[3] AgrawalR andSrikantR, *Fastalgorithmsformining associationrules, Proceedingsofthe20th VLDBConference*:,1994,487-499.

[4] *http://en.wikipedia.org*

[5] Fausett, Laurene, *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*(Prentice-Hall, New Jersey, USA, 1994).

[6] G Towell, J W Shavlik. The extraction of refined rules from knowledge based neural networks, *Machine Learning*, 1993(13):71-101.

[7] Rumelhart, D. E and Zipser, D, Feature discovery by competitive learning. *Cognitive Science*, 1985,9:75-112.

[8] Zurada J.M, *An introduction to artificial neural systems*(St. Paul: West Publishing, 1992).

[9] Aleksander I. and Morton H, *An introduction to Neural Computing*,(2nd edition,1998).

[10] Maulik V. Dhamecha and Bhavesh R. Akbari, COMPREHENSIVE STUDY OF MODIFIED ARTIFICIAL NEURAL NETWORK ALGORITHM AND COMPARE EFFICIENCY AND PERFORMANCE WITH BASIC ALGORITHM OF NEURAL NETWORK IN DATA MINING,*International Journal of Research in Engineering & Applied Sciences*, 2(2), 2012, pp. 29-39.