epcc

Data Mining

Julian Parker, Terence Sloan and Hon Yau

Edinburgh Parallel Computing Centre

The University of Edinburgh

Version 1.0

Available from: http://www.epcc.ed.ac.uk/epcc-tec/documents/


epcc

Table of Contents

1 Introduction

1.1 Motivation

In the commercial and financial sector very large online databases, necessary for the daily operation of an organisation, are continually updated.However, the contents of these databases, collected over a period of time, also provide an important strategic resource. Product life-cycles, customer needs and other such trends are all present within this data. Analyses of the databases can therefore produce useful information in, for example, new product identification or in determining future strategy for an organisation.

Data mining is a concept that has been establishing itself since the late 1980's. It covers a range of techniques for the efficient discovery of this valuable, non-obvious information from such large collections of data. Essentially, data mining is concerned with the analysis of data and the use of software techniques to find patterns and regularities in datasets.

The motivation for commercial data mining is clearly that the knowledge derived from it may be used for competitive advantage. The most familiar application of data mining is probably the selection of customers most likely to respond to a targeted direct marketing. Other organisations with large operational databases that may use data mining include government departments and health services. Their motivations would be to improve the targeting of service delivery and management of costs.

For scientists in disciplines that are struggling to cope with the analysis of huge amounts of raw data (*e.g.*, geophysicists and astrophysicists), there are opportunities for traditional data reduction methods (*e.g.*, statistics and image processing) to be complemented by data mining techniques such as feature extraction, classification and symbolic reasoning.

1.2 Aims

This report aims to give a brief, cohesive overview of the breadth of data mining techniques. Data mining is clearly of great commercial interest, and the description of the techniques reflects this. However, this report also aims to make readers aware of the possibilities for applying data mining ideas to scientific data analysis. If some analysis tasks can be automated, or some discrimination applied to reduce dataset size, then more time available for higher level data analysis and assimilation of knowledge.

1.3 Overview

The report has two main sections. Section 2 gives an overview of data mining techniques. References to particular packages or proprietary techniques are omitted, and the reader is referred to the plentiful, up-to-date information available on the Internet for details of software packages. Suggested URLs are listed in Section 2.3

Section 3 gives a brief description of the application of data mining to scientific data analysis, and has some general observations in that topic.

Appendix A: gives an overview of neural networks, which are referred to often throughout the data mining literature.

2 Concepts and Techniques

Fayyad, Piatestsky-Shapiro and Smyth [2] use the term Knowledge Discovery in Databases (KDD) to refer to the overall process of discovering useful knowledge from large datasets. They classify data mining as the particular step in this process dealing with the application of specific algorithms for extracting patterns (models) from data.

Data mining utilises techniques such as neural networks, rule-based systems, case-based reasoning, machine learning and statistical programs, either alone or in combination [4], to analyse and extract patterns from the data.

The KDD process involves the additional steps of data preparation, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, all of which help to ensure that useful knowledge is derived from data. Blind application of data mining methods can be a dangerous activity leading to the discovery of meaningless patterns.

Figure 1 illustrates the steps in the KDD process.

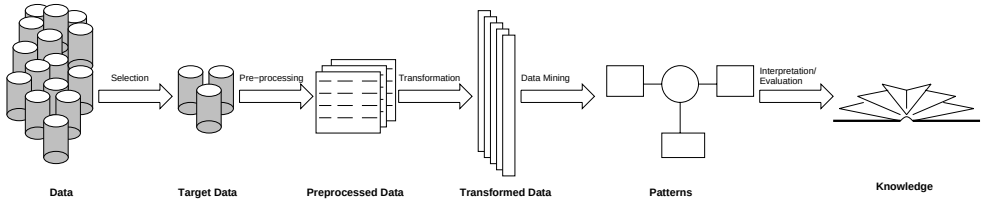

Figure 1: Overview of the steps constituting the knowledge discover in databases process (based on [2])

The next section give a brief overview of the date preparation stage, and then a summary of data mining techniques follows. References to proprietary software packages and algorithms have been specifically avoided. For information on public domain and commercial software, the reader is referred to the plentiful and detailed information available on the Internet. Some WWW sites are listed in Section 2.3.

2.1 Data Preparation and Warehousing

For a successful data mining operation the data must be consistent, reliable and appropriate. The appropriate data must be selected, missing fields and incorrect values rectified, unnecessary information removed, and where data comes from different sources, the format of field values may need to be altered to ensure they are interpreted correctly. In addition to this data cleansing and preprocessing, some transformation of the data may be necessary. For example, some fields may be reduced to a single data item or extra fields may even be generated.

While viewed by some as a bothersome preliminary step (sort of like scraping away the old paint before applying a fresh coat of paint), data preparation is crucial to success. Indeed IBM Consulting and independent consultants confirm estimates that data

preparation and pre-processing might consume anywhere from 50 to 80 percent of the resources spent in a data mining operation [1].

Typically, the collected and cleaned data are then stored in some form of data warehouse or datamart [4] for analysis. Data warehousing refers to the current business trend of collecting and cleaning transactional data to make them available for online analysis and support. Such warehouses are updated periodically ---monthly, weekly, or even daily depending on an organisation's needs [5].

Traditional on-line transaction processing (OLTP) systems insert data into databases quickly, safely, and efficiently [6]. However, these systems do not provide the means for the effective derivation of strategic knowledge from large masses of collected data. Products from vendors such as Redbrick are designed specifically to service data warehouse applications like data mining and data analysis, and are able to offer the performance that traditional OLTP-based databases cannot produce. Mainstream relational database managements system (RDBMS) vendors such as Oracle have also introduced new products like Oracle Express Server directed at data warehousing. Other vendors offering such products include Informix, Tandem Computers and Sybase.

Within the data warehouse, the data are organised by subject rather than application, with the different data types forming dimensions. This allows multidimensional views to be constructed where a number of data types can be combined. Some products use a method known as the star schemata [7]. This mimics multi-dimensionality by creating special tables that roll-up data. For example, you may have a central fact table surrounded by star tables with location, time and product data.

There is some argument among vendors as to who provides products specifically designed for data warehousing and who is merely re-badging existing OLTP-focused software for the new market. Either way, these products often exploit parallel processing platforms to provide the power to analyse very large collections of data

2.2 Survey of Techniques

Statistics is fundamentally important in the extraction of meaning from large datasets, and is also a tools for testing hypothesis against data. It is one of the range of data mining techniques, but as it's a well established, familiar discipline, it is not covering in the following survey. The same is true fro data visualisation and again, visualisation is not described further in this report.

To extract information from the prepared data, a variety of data mining tools are available ranging from those that include some artificial intelligence (AI) techniques, through the on-line analytical processing (OLAP) tools to more traditional query and report methods. Various taxonomies have been proposed to describe and categorise the numerous data mining techniques and approaches. Two of these are described. Given the novelty and rapid development of the field, some further techniques are then described in subsequent subsections.

Watterson [7] describes three categories of data mining tools.

- **Intelligent Agents** are capable of sifting through data, hypothesising connections and reporting discoveries. Some are launched manually to perform specific queries to search for patterns in data. Others fire off automatically at predefined intervals, performing a task or monitoring a condition in the background and returning an alert as required. Most intelligent agents are simply short programs that say "if this happens, do that". They are generally considered as information discovery tools which take large amounts of detailed transaction-level data and apply mathematical techniques against these, finding or discovering new knowledge. They require some expertise to set up but once activated need little direction. Some work directly on text. Best used for turning up

unsuspected relationships.

- **Multidimensional analysis (MDA) tools** Here data is represented as n-dimensional matrices called hypercubes, with users able to explore combinations of one or more dimensions of the data. This is sometimes also referred to as online analytical processing or OLAP. Essentially the tool is loaded with all the various types of data whose relationships you wish to examine. For example, imagine all the possible ways of analysing clothing sales: brand name, size, colour, location, advertising and so on. A multidimensional hypercube is filled with this data, and different views of the data can then be generated and examined. In the clothing example, brand name versus advertising costs versus location or any other such combination of the data types. Generally the views are restricted to two or three dimensions. Even then an n-dimensional hypercube has *n(n-1)* two-dimensional views. Usually these tools have simple graphical interfaces for non-expert use. They are generally considered to be good at iterative, interactive exploration of data.
- **Query-and-reporting tools** These tools construct quite complex SQL queries to analyse the contents of a database. However, such queries can have a drastic effect on the performance of a production system and hence are usually made on a data warehouse. These tools require close direction to frame SQL queries. Many tools simplify query generation by providing graphical interfaces. They require a database structure and are useful for asking specific questions to verify hypotheses. However, they are legendary for their ability to drastically slow down production systems.

Dilly describes some of the various techniques used by data mining tools. These are Cluster Analysis, Induction, Neural Networks and On-line Analytical Processing. The following sections describe each of these techniques and are based on extracts from [6].

2.2.1 Cluster Analysis

Clustering and segmentation basically partition the database so that each partition or group is similar according to some criteria or metric. Clustering according to similarity is a concept which appears in many disciplines. If a measure of similarity is available, there are a number of techniques for forming clusters. Membership of groups can be based on the level of similarity of members and from this the rules of membership can be defined. Another approach is to build set functions that measure some property of partitions as functions of some parameter of the partition. This latter approach achieves what is known as optimal partitioning.

Many data mining applications make use of clustering according to similarity to segment for example a client / customer database. Clustering according to optimisation of set functions is used in data analysis, for example, when setting insurance tariffs the customers can be segmented according to a number of parameters and the optimal tariff segmentation achieved.

Clustering / segmentation in databases are the processes of separating a data set into components that reflect a consistent pattern of behaviour. Once the patterns have been established they can then be used to reconstruct data into more understandable subsets and also to provide subgroups of a population for further analysis or action. This extraction of sub-groups from a population is important when dealing with large databases. For example, a database could be used for profile generation for target marketing. Previous responses to mailing campaigns can be used to generate a profile of people who responded. This profile can then be used to predict responses and filter mailing lists to achieve the best response.

2.2.2 Induction

A database is a store of information but more importantly information can also be inferred from it. There are two main techniques available are deduction and induction.

- **Deduction** infers information that is a logical consequence of information in a database. For example, the join operator applied to two relational tables where the first concerns employees and departments and the second departments and managers infers a relation between employee and manager. That is, deduction works on existing facts and deduces new knowledge from old [8].
- **Induction** infers information that is generalised from a database. That is, rather than starting with existing knowledge, it takes examples and generalises. This is higher level information or knowledge in that it is a general statement about objects in the database. This involves searching the database for patterns or regularities.

Within data mining, induction has been used in decision trees and in rule induction.

2.2.3 Decision Trees

Decision trees are a simple knowledge representation that classify data objects to a finite number of classes. The nodes of the tree are labelled with attribute names, the edges with possible values for this attribute and the leaves with different classes. Objects are classified by following a path down the tree, by taking the edges corresponding to the values of the attributes in an object.

Figure 2 is an example of a decision tree that classifies objects describing the weather at a given time. The objects contain information on outlook, humidity, etc. Some objects are positive examples denoted by *P* whilst others are negative denoted by *N*. The diagram can be used to classify the objects correctly.

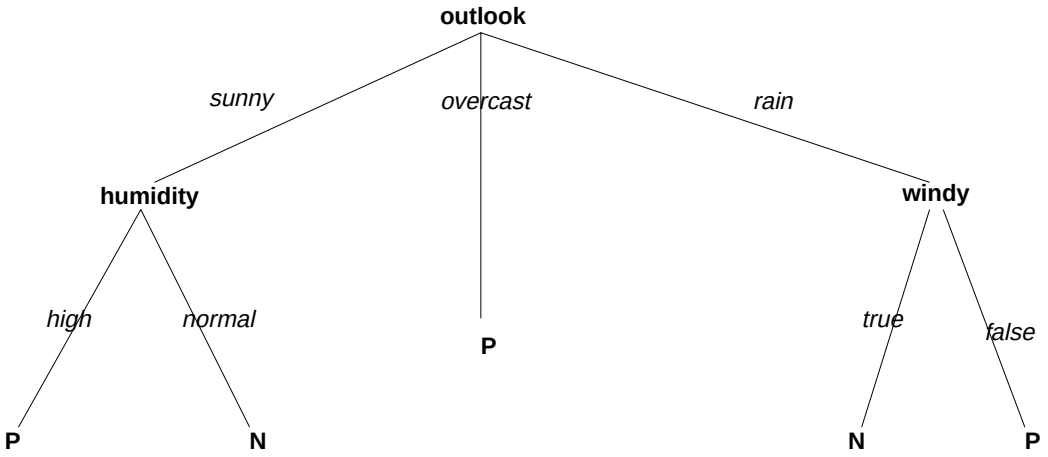

Figure 2: Example of a decision Tree structure for weather description (based on [6])

2.2.4 Rule Induction

Rule induction attempts to infer some statistically meaningful patterns about a portion of the data. The rules are in the form *if P then Q*, where *P* is a clause matching some portion of the database and *Q* is the prediction clause. For example, if the price of apples is lowered by 10 percent, then sale increase by 5 percent. Such rules have been widely used to represent knowledge in expert systems and they have the advantage of being easily understood by human experts because of their modularity. That is, a single can be understood in isolation and does not need to reference other rules.

2.2.5 Neural Networks

Neural networks are an approach to computing that involves developing mathematical structures with the ability to learn. The methods are the result of research to model learning in the nervous system. Neural networks can derive meaning from complicated or imprecise data and 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 analyse during training.

Appendix A describes neural networks in more detail, and reference [22] provides a thorough introduction to the field.

2.2.6 On-line Analytical Processing

The term OLAP refers to the dynamic synthesis, analysis and consolidation of large volumes of multidimensional data. It is essentially a way to build associations between dissimilar pieces of information using predefined business rules about the information. The advantages of a dimensional system are the freedom they offer a user to explore the data and format reports without being restricted in their format.

OLAP database servers use multidimensional structures to store data and relationships between data. These structures can best be visualised as cubes of data, and cubes within cubes of data where each side of the cube is considered a dimension. Figure 3 illustrates such a structure.

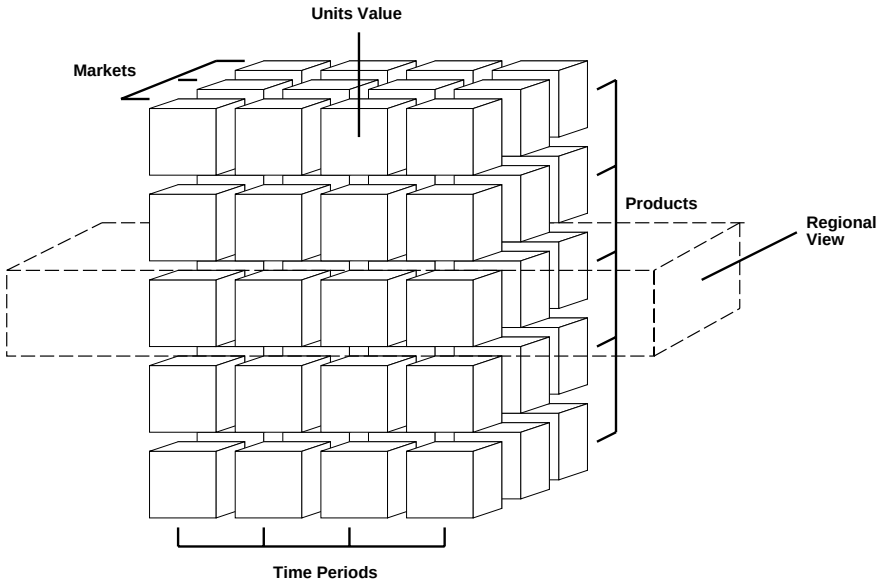

Figure 3: Hypercube structure representing a multidimensional database containing information on products sold in different markets over a number of time periods.

Each dimension represents a different category such as product type, region, sales channel and time. Each cell within the multidimensional structure contains aggregated data relating elements along each of the dimensions. For example, a single cell may contain the total sales for a given product in a region for a specific sales channel in a single month. Multidimensional databases are a compact and easy to understand means for visualising and manipulating data elements that have many inter relationships.

OLAP database servers support common analytical operations including: consolidation, drill-down and ``slicing and dicing''.

- **Consolidation** involves the aggregation of data such as simple roll-ups or complex expressions involving inter-related data. For example, sales offices can be rolled up to districts and districts rolled-up to regions.
- **Drill-Down** OLAP data servers can also go in the reverse direction and automatically display detail data which comprises consolidated data. This is called drill-down. Consolidation and drill-down are an inherent property of OLAP servers.
- **"Slicing and Dicing"** refers to the ability to at the database from different viewpoints. One slice of the sales database might show all sales of product type within regions. Another slice might show all sales by sales channel within each product type. Slicing and dicing is often performed along the time axis in order to analyse trends and finds patterns.

Essentially, OLAP servers logically organise data in multiple dimensions and so allow users to quickly and easily analyse complex data relationships. The database itself is physically organised in such a way that related data can be rapidly retrieved across multiple dimensions. OLAP servers are very efficient when storing and processing multidimensional data. In contrast, RDBMSs have been developed and optimised to handle OLTP applications. Relational database designs concentrate on reliability and transaction processing speed, instead of decision support need.

2.2.7 Score Cards

Score carding is a data modelling technique which tries to partition the data into just two sets. That is, for a given a binary objective function which appears to depend on a set of analysis field values, the algorithm will attempt to best separate those data with one objective value, from those with another. Conceptually, a hyperplane[1] is drawn through the dataset to separate those fields which give the two responses. Since it is highly unlikely that all the data points can be separated into two regions by a hyperplane, a means of measuring the quality of the split is usually defined; for example, some measure of the number of records correctly categorised by the algorithm divided by the total number of records used to build the Score card model. Having produced a Score card model with one dataset, it can then be very quickly applied to other data, to predict their outcome by evaluating the objective function for each record in the new dataset.

The mathematical operation performed is also identical to training a single layer neural network with a linear response, for each item of data. Because of this architecture, the weights for from each input neuron can be calculated by standard matrix linearisation techniques such as Gaussian elimination.

2.2.8 Bayesian Networks

Bayesian networks (also known as Belief Networks, Causal Probabilistic Networks and Link Analysis), are complex diagrams that organize the body of knowledge in a given domain by mapping out cause-and-effect relationships among key variables and encoding them with numbers that represent the extent to which one variable is likely to affect another. Bayesian networks combine the disciplines of classical statistics with expert knowledge.

1. A hyperplane is a generalisation of a 2-D straight-line or 3-D flat-plane into arbitrary multi-dimensional space, where each field attribute in the dataset is represented on each spatial dimension.

Programmed into computers, these systems can automatically generate optimal predictions or decisions even when key pieces of information are missing. This differentiates Bayesian networks from other machine learning techniques such as neural networks which cannot predict unforeseen conditions.

Typical application areas are in safety and risk assessment, and in the control of complex machinery (*e.g.*, reactor and unmanned underwater vehicle control systems).

2.2.9 Text Mining

This topic covers a range of commercial software tools rather than particular techniques. Data mining tools are now available which are adapted to search, analyse and classify unformatted textual data. Application areas are the legal profession, intelligence gathering and market research. In the later case, text mining was used to analyse the response to open questions in a market survey. The possibilities of using text mining to do more sophisticated literature searches of research abstracts databases seems clear.

2.2.10 Deviation Detection

Deviation detection covers a range of techniques uses for data auditing, to reveal errors and suspect cases in data. Typical application areas are in data cleansing, and fraud detection (*e.g.*, in illegal use of credit cards).

2.2.11 Genetic Algorithms

Genetic Algorithms are primarily an optimisation technique, so it isn't immediately clear why they are often associated with data mining. The connection between the techniques is that optimisation features in some data mining techniques such as rule induction and decision trees. As described earlier, rule induction attempts to model the data using rules, which contain parameters determined by the knowledge discovery algorithm. An optimisation stage is need to determine the parameter values that create the "best" rules to model the data, and hence Genetic Algorithms have been incorporated in such data mining tools.

For a detailed description of Genetic Algorithms, see the EPCC Technology Watch Report "Genetic Algorithms for Optimisation".

2.3 References

- `http://pwp.starnetinc.com/larryg/` -- The Data Warehousing Information Center has comprehensive, tabulated pages on warehousing and OLAP products, and data mining.
- `http://www.kdnuggets.com/` -- has a wealth of pointers to software in all the categories of data mining techniques, and the KD Nuggets newsletter.
- `http://scanner-group.mit.edu/DATAMINING/` -- well structured presentation of more commercially relevant data mining topics.
- `http://www.kdd.org/` -- Knowledge Discovery and Data Mining Foundation
- `http://www.olapreport.com/` -- Another site address at serious commercial users.
- `http://www.yahoo.co.uk/Computers_and_Internet/Software/Databases/`

3 Scientific Applications

The discussion of data mining and its applications, in the previous section and in published material generally, focuses mostly on commercial business sectors. Here, the opportunity to exploit existing data that is collected during business operations is being taken. If this data is then analysed using data mining, there are usually few hypothesis about what information (knowledge) is contained in the data. Hence business data mining tends to benefit the most from the sophistication of the various knowledge discovery techniques. In contrast, scientific data is collected specifically to analysis it. The structure data is understood better, so selecting the analysis techniques will usually be more straightforward. The main tools are well established statistical analysis and visualisation techniques. However, the emergence of novel data mining techniques is an opportunity to apply them in the analysis of some types of scientific data.

The remainder of this section outlines case studies that were described by Fayyad, Haussler and Stolorz [3], and the conclusions that they drew about scientific data mining.

3.1 Sky Survey Cataloguing

The 2nd Palomar Observatory Survey had the huge task of analysis 3 terabytes of data on 3,000 digitised photographic images, to find and classify an estimated 2 billion sky objects (*e.g.*, stars, galaxies, clusters). The initial processing sequence was basic image segmentation to select sky objects and then measuring 40 attributes of each one. The difficult problem was then classifying the object based on its attributes. A decision tree learning algorithm was used. Details of the method of this Sky Image Cataloguing and Analysis Tool (SKICAT) are given in [19]. The accuracy achieved by SKICAT in classifying objects was 94 percent. Reliable classification of faint objects increased the number of objects classified (usable for analysis) by 300 percent.

Key points given for the success of SKICAT are:

- The astronomers solved the feature extraction problem (the transformation from pixel data to feature space). This transformation implicitly encodes a significant amount of prior knowledge.
- Within the 40-dimensional feature space, at least 8 dimensions are needed for accurate classification. For a human, choosing which 8 of the 40 to use, and then actually classifying the object, is difficult. Data mining methods contributed by solving the classification problem.

3.2 Biosequence Databases

Interpreting the DNA sequence of the human genome and relating it to the structure and function of the proteins that make up a human cell is a huge endeavour. The

genome is about 3 billion “letters” long, and there are about 100,000 different human proteins. Data mining techniques are been used in two tasks:

- Gene-finding -- Finding the parts of the DNA sequence that “code” particular genes. Lots of effort has already been devoted to developing gene-finding programs, which use neural nets and other artificial intelligence or statistical methods. Improving these automatic or computer-assisted methods is an area of much research.
- Database searching -- search methods are used to match a query sequence to stored sequences with similar high-order structure or function. The matching involves more than just sequence correlation, since the 3d, folded structure of biomolecules is crucial to their function.

A method for dealing with sequence data is to infer transition probabilities between process state variables from the observed data. A particularly successful class of techniques used for modelling this is hidden Markov models (HMM) [20]. Development of algorithms based on HMM took place for speech processing, and this effort has been successfully re-applied to protein modelling.

3.3 Extraction of Atmospheric Features

As with the case of the sky object catalogue, atmospheric scientists are interested on having access to a database of certain categories of atmospheric features for detailed analysis. For example, tornados or storm episodes. The difficulty is to extract the interesting features from the mass of observation data provided by remote sensing instruments. A team at UCLA and JPL developed the CONQUEST system [21] to extract and store data on instances of a type of cyclone. The method of feature extraction involved singular value decomposition and neural net pattern recognition algorithms. Parallel computer implementations of these algorithms were used to handle the large spatial/temporal input datasets. The selected features were then stored in a relational database, where they could then be retrieved using content-based indexing.

3.4 Key Issues

In the conclusions in their paper [3], Fayyad, Haussler and Stolorz identified key issues that differential scientific data mining from its business application:

- Ability to use prior knowledge and understanding during mining
- More stringent requirements for accuracy, such as in the classification of objects requiring an accuracy better than 90 percent for SKICAT.
- Issues of scalability of computers and algorithms (e.g., use of parallel computers in CONQUEST).
- Ability to deal with low-probability class, which may be significant and interesting for research.

4 Bibliography

[1] Bigus J. P. *Data mining with neural networks*. McGraw-Hill, 1996.

[2] Fayyad U., Piatesky-Shapiro G, and Smyth P. *The KDD Process for Extracting Useful Knowledge from Volumes of Data*. Communications of the ACM, 39(11):27--34, November 1996.

[3] Fayyad, Haussler and Stolorz. *Mining Scientific Data*. Communications of the ACM 39, 11 (1996), 51-57.

[4] Hedberg S. R.. *The Data Gold Rush*. Byte, pages 83--88, October 1995.

[5] Watterson C. D.. *Data-Mining Dynamite*. Byte, pages 97--103, October 1995.

[6] Dilly R. *Data Mining: An Introduction - Student Notes*. Parallel Computer Centre, The Queen's University of Belfast, December 1995.

[7] Watterson K. *A Data Miner's Tools*. Byte, pages 91--96, October 1995.

[8] Finlay J. and Dix A. *An introduction to Artifical Intelligence*. UCL Press, 1996.

[9] James M. *Neural networks make predictions for science*. Scientific Computing World, pages 29--30, November 1996.

[10] Beale R. and Jackson T. *Neural Computing: An Introduction*. Adam Hilger, IOP Publishing Ltd., 1990.

[11] Jain A. K. and Mao J. *Artificial Neural Networks: A Tutorial*. Computer, pages 31- 44, Mar 1996.

[12] Pal S. K. and Srimani P. K. *Neurocomputing - Motivation, Models and Hybridization*. Computer, pages 24--28, Mar 1996.

[13] Holland J. H. *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, Michigan, U. S. A., 1975.

[14] Michalewicz Z. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 3rd edition, 1996.

[15] Kirkpatrick S., Gelatt G. D., and Vecchi M. P.. *Optimisation by Simulated Annealing*. Science, 220(4598):671--680, 1983.

[16] Wolfpert D. H. and Macready W. G. *No Free Lunch Theorems for search*. Technical Report SFI-TR-95-02-010, Santa Fe Institute, New Mexico, U. S. A., 1995.

[17] Fishwick P. A.. *Computer Simulation: The Art and Science of Digital World Construction*. IEEE Potentials, 15(1):24 --- 27, Feb / Mar 1996.

[18] Smith R. *Simulation*. In Encyclopedia of Computer Science. IEEE Press, 1998.

[19] Weir N., Fayyad U.M., and Djorgovski S.G. *Automated star/galaxy classification for digitized POSS-II*. Astron. J. 109, 6 (1995), 2401-2412

[20] Rabiner L.R. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proc. IEEE 77 (1989), 257-286.

[21] Stolorz P, et al. *Fast spatiotemporal data mining of large geophysical datasets*. In Proceedings of the 1st Int. Conference. on Knowledge Discovery and Data Mining (Montreal, Aug. 1995), AAAI Press, Menlo Park, Calif. 1995.

[22] Hertz J., Krough A, and Palmer R. G. *Introduction to the Theory of Neural Computation*. Studies in the Sciences of Complexity. Santa Fe Institute, Lecture Notes Volume 1, Addison-Wesley Publishing Company, 1991.

Appendix A: Neural Networks

An artificial neural network is a structure built from elements whose behaviour resembles that of biological neurons [9]

The basic model of a neuron, as described in Beale and Jackson [10], is shown in figure f:basic neuron . It performs a weighted sum of its inputs, compares this to some internal threshold level, and turns on only if this level is exceeded. If not, it stays off. Since the inputs are passed through the model neuron to produce the output, the system is known as a feed-forward one.

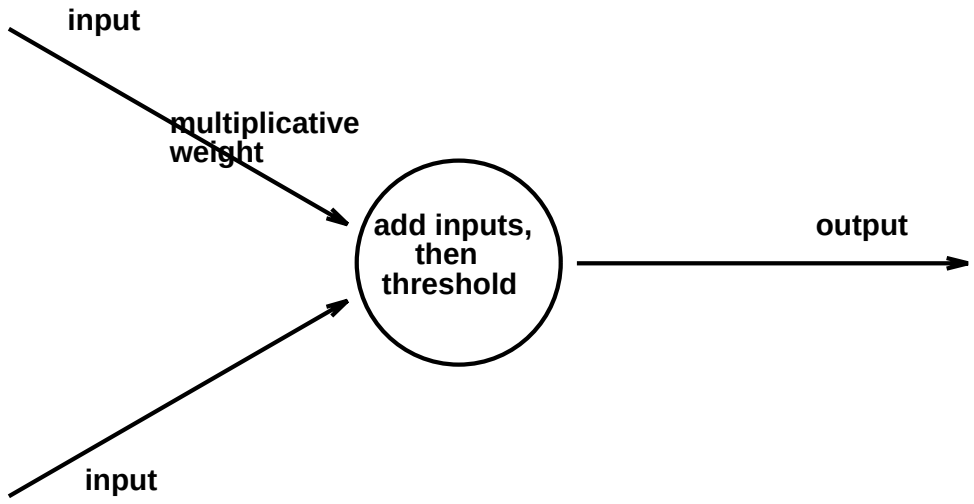

Figure 4: Outline of the processing in single neuron (based on [10]).

As explained in Jain [11], networks of these artificial neurons are capable of performing tasks such as those listed below.

- **Pattern classification** That is, assigning an input pattern to one of a number of pre-specified classes. Well-known applications are character recognition, speech recognition and printed circuit board inspection.
- **Clustering/categorisation** This involves exploring the similarity between patterns, placing similar patterns in a cluster. Applications include data mining, data compression and exploratory data analysis.
- **Function approximation** Given a number of input-output pairs generated from an unknown function, this involves finding an estimate of this function. Various engineering and scientific modelling problems require this.
- **Prediction/forecasting** Given a set of samples in a time sequence, the task is to predict the sample at some future time. Stock market and weather prediction are typical applications.
- **Optimisation** The goal here is to find a solution satisfying a set of constraints such that an objective function is maximised or minimised. The Travelling Salesrep Problem (TSP) is the classic example.

- **Content-Addressable Memory** Also known as Associative memory. The content in memory can be recalled even by a partial input or distorted content.
- **Control** Here a control input is generated such that a system follows desired trajectory determined by a reference model. An example is engine idle-speed control.

There are various types of artificial neural networks, each with different characteristics, with some types of network more suited to certain applications than others. However, each type of network must undergo some of form of learning. This learning updates the network architecture, altering the weights on the connections between neurons so that the network can efficiently perform a specific task.

Essentially there are three forms of learning which Pal and Srimani [12] describe:

- **Supervised** where the network is provided with the correct answer for every input pattern. Weights are determined to allow the network to produce answers as close as possible to the known correct answers. Reinforcement learning is a variant of supervised learning in which the network is provided only with a critique on the correctness of the network outputs, not the correct answers themselves.
- **Unsupervised** Here a correct answer for each input pattern is not required. This explores the underlying structure in the data, or correlations between patterns in the data, and organises patterns into categories from these correlations.
- **Hybrid** This combines supervised and unsupervised learning with part of the weights determined through supervised and part through unsupervised.

Pal and Srimani go on to describe the more popular types of neural network currently available and their applications. Their descriptions of these are summarised below.

Hopfield Network

This acts as a nonlinear associative memory that can retrieve an internally stored pattern when presented with an incomplete or noisy version of that pattern. It can also be used as an optimisation tool. The Boltzman machine is a generalisation of the Hopfield network which is based on simulated annealing. The Hopfield network, the Boltzman machine, and a derivative known as the Mean-Field-Theorem machine have been used in applications such as image segmentation and restoration, combinatorial optimisation (eg. TSP), in addition to their use as a contents addressable memory.

Multilayer perceptron network

This is a network of elementary processes that can learn to recognise and classify patterns automonously. A single layer perceptron is unable to classify patterns with non-linear separating boundaries, however a multilayer network can. Supervised learning is used, ie. Feed-forward back propagation, where weights are adjusted to minimise the error between the observed and desired outputs. It is believed that an MLP can produce correct (or nearly correct) output for input not used in training. An MLP performs interpolation well, since the continuous activation functions produce continuous output functions. Given a data set however an MLP can pick up one of many possible generalisations corresponding to different minima. Also since learning involves searching over a complex space, it is often time consuming. The MLP has been applied to many applications ranging from classifier design, function approximations and speech identification to scene analysis and military target identification.

Self-Organising Feature Mapping (SOFM)

This unsupervised learning network transforms p-dimensional input patterns to a q-dimensional (usually q = 1 or 2) discrete map in a topologically ordered fashion. Input points that are close in the p dimension are also mapped closer on the q dimensional lattice. Each lattice cell is represented by a neuron associated with a p dimensional adaptable weight vector. The match between each weight vector is computed with every input. The best matching weight vector and some of its topological neighbours are then adjusted to better match the input points. Such networks have been used for generating semantic maps, phonetic typewriters and graph bipartitioning. In a special case of SOFM, the Learning Vector Quantisation (LVQ) network, only the weight vector associated with the winner node is updated with every data point. Such a learning scheme is known as competitive . It is essentially a clustering network that does not preserve topological order. Its main uses are for clustering and image data compression.

Adaptive Resonance Theory (ART) network

In a competitive learning scheme, there is no guarantee that the clusters formed will be stable unless the learning rate gradually approaches zero with iteration. However when this happens, the network loses its plasticity. In ART, a weight vector is adapted only when the input is sufficiently similar to that of the prototype. A vigilance parameter is used to check on similarity.

Radial basis function network

Here, the output nodes linearly combine the basis functions computed in the hidden layer nodes. This type of network is also as known as a localised receptive field network because the hidden layer nodes produce localised responses to the input signals. The most commonly used basis function is the Gaussian kernel. Like MLP, RBF networks can be used for classifier design and function approximation, and they can make an arbitrary approximation to any continuous nonlinear mapping. The main difference between MLP and RBF lies in the basis function used by the hidden layer nodes. RBF nets use Gaussian while MLP nets use sigmoidal functions. The choice between RBF and MLP depends on the problem at hand.