ANN as a tool for medical prognosis

Sameem Abdul Kareem, University of Malaya

Available at: https://works.bepress.com/sameem_abdulkareem/6/
ANN as a tool for medical prognosis
S. Abdul-Kareem, S. Baba, Y.Z. Zubairi and M.I.A. Wahid

HEALTH INFORMATICS J 2000; 6; 162

The online version of this article can be found at:
http://jhi.sagepub.com/cgi/content/abstract/6/3/162

Published by:
SAGE Publications
http://www.sagepublications.com

Additional services and information for Health Informatics Journal can be found at:

Email Alerts: http://jhi.sagepub.com/cgi/alerts

Subscriptions: http://jhi.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav
ANN as a tool for medical prognosis

S. Abdul-Kareem, S. Baba, Y.Z. Zubairi and M.I.A. Wahid

The analysis of cancer survival is used to determine the efficiency of treatment programmes and protocols; it is also used to determine the type of treatment. At the individual level a prediction of cancer survival can help patients make informed decisions with regards to their quality of life and future finances. Currently available prediction methods apply to groups of people, and may not be adequate to predict treatment outcome for individual patients. This paper presents a conceptual model of a cancer knowledge base incorporating a computer-based predictor for survival. It proposes the use of an artificial neural network (ANN) as an alternative tool for investigating cancer survival. The focus of this research will be on nasopharyngeal carcinoma survival data collected in Malaysia.

INTRODUCTION

Medical prognosis is an estimate of cure, complication, recurrence of disease, length of stay in healthcare facilities or survival for a patient or group of patients. The simplistic meaning of prognosis is a prediction of the future course and outcome of a disease and an indication of the likelihood of recovery. It is only a prediction, however, and like all predictions it is not 100 per cent accurate [1].

Statistical analysis of survival data is usually used to help estimate prognosis; survival statistics indicate a cohort of patients with certain types and stages of cancer and the outcome and survival following treatment. However, statistics alone may not be sufficient to predict the outcome of a particular patient, as no two patients are exactly alike. Cancer statistics apply to large groups of people and may be meaningless for an individual patient, as no two patients are exactly alike. As an example, an 80 per cent chance of survival is only a prediction, however, and like all predictions it is not 100 per cent accurate [1].

Nasopharyngeal carcinoma (NPC) is a type of cancer predominant within the Chinese community; research findings have placed the incidence of NPC among the Chinese community of Malaysia only slightly below that of China and Hong Kong [8]. For this reason, we have chosen NPC to investigate prognostic models in cancer.

NPC is a malignant growth of the tissue and the mucosa lining the nasopharynx. The nasopharynx is an area situated behind the nose and the upper part of the throat above the soft palate at the back of the mouth. The eustachian tube and the fossa of Rosenmuller form the lateral walls. The nare through which we breathe lead into the nasopharynx; air and food pass through the pharynx on the way to the trachea (windpipe) or the oesophagus respectively [9] [10] [11] [12] [13] [14].

Survival analysis

Medical prognosis is equally important to healthcare policy makers and administrators, drug manufacturers, clinicians as well as patients. The statistical analysis of survival data is a method that is commonly employed by researchers in estimating cancer prognosis.

Prognostic information is important to healthcare policy-makers and administrators in determining healthcare policies and to monitor the progress of regional and national cancer control programmes. Survival analysis plays an important role in clinical decision-making: as a tool for assessing the efficacy of treatment protocols, to monitor the progress of treatment programmes and as an aid in choosing treatment types and methodologies. Survival outcomes are also used in medical purchasing and medical supplies manufacturing and individual as well as population prognosis.
On the part of patients, prognostic information helps them understand their conditions, thus aiding in making informed decisions with regards to treatment, finance and the quality of life.

A cancer patient’s prognosis can be affected by many factors, such as the type of cancer the patient has, the stage of the cancer, or its grade (how aggressive the cancer is or how closely the cancer resembles normal tissue). Other factors that may also affect a patient’s prognosis include his or her age and general health or the effectiveness of treatment.

Survival analysis in cancer is difficult because different patients cannot be observed for the same length of time. This may be because of the fact that some patients are diagnosed at the beginning of the period under study, some near the end and others may be diagnosed at any other time of the study period. In survival analysis terminology, patients who are observed until they reach the end point (that is, death) are called uncensored cases while those who survive beyond the end or who are lost to follow-up at some point are called censored cases [15].

Based on the success artificial neural networks have shown in prediction we believe that they have the ability to predict medical prognosis, and furthermore they would be able to predict the prognosis at an individual level.

THE CONCEPTUAL MODEL

The project that we are currently undertaking involves the development of a conceptual model of a knowledge base as shown in Figure 1. This model consists of a database component for data storage purposes, and a knowledge discovery component. We propose to link these two components via an Open Database Connectivity (ODBC) interface.

The knowledge discovery component consists of the data mining devices that will be used in the project, namely, statistical tools for data analysis purposes. The knowledge base would also be incorporated with a predictor model, shown in Figure 2, utilizing neural network technology to predict the outcome of cancer survival.

Database

The NPC data that have been collected by the ENT Department of the University Hospital, University of Malaya, since 1969 were initially in a conventional paper format. Subsequently, these data were converted into an electronic media in 1982. For the purpose of this project, the data are now transformed into a relational database that will enable easy manipulation. Operations on the database will include queries, accesses and reports. The use of the relational database also allows for easy linking via ODBC, with data mining devices.

Data mining

Data mining is the process designed for extracting consistent, useful and interesting information patterns inherent in data. Data mining methods, which combine techniques from database research, artificial intelligence and statistics, have been used in a variety of domains. The concept of data mining is becoming increasingly popular as an information management tool as it reveals the knowledge implicit in the data that has been collected by various application areas.

Statistical tools

The analysis of survival data has long been the domain of statistics as can be observed through the number of medical statistics textbooks and journals dedicated to the field [2] [15] [16] [17] [18]. Statistical methods such
ANN

As has been previously mentioned, data mining methods combine techniques from database research, artificial intelligence and statistics. One technique that is increasingly used among data mining researchers is artificial neural network [19].

An artificial neural network (ANN) is an information processing system that tries to simulate biological neural networks. The field of ANN was born in an attempt to overcome the limitations of the computer’s ability to perform certain tasks; these tasks, such as reading a handwritten document or recognizing a face, may seem simple for human beings, but are difficult for even the most advanced computers. Programmers, therefore, began exploring the option of having a technology that mimics the acts of the human brain, with its neurons and synaptic connections [20] [21].

ANN techniques are different from classical data analysis techniques in the sense that ANNs learn about a particular subject from the data provided, rather than being programmed by the user in a conventional sense. Conventional computing relies on programmes that solve a problem using a series of steps or well-defined procedures known as algorithms. The programmes are controlled by a single, complex central processing unit (CPU), and store information at specific locations in memory. The CPU is used to carry out many instructions, which are handled one at a time. Neural networks, on the other hand, analyse data by passing them through several simulated processors that are interconnected and highly distributed. Although the programming and mathematics behind neural network technologies are rather complex, applying the technology for problem solving can be quite simple and the results are extraordinary [22]. One of the key characteristics of neural network technology is that no programming is involved.

Neural network performs its processing by accepting inputs, which are multiplied by a set of weights non-linearly transformed by the neurons into an output value. The output of a neuron depends on the neuron’s input and on its activation function. ANNs gather their knowledge by detecting the patterns and relationships in the data, learning from the relationships and adapting to change. For each new situation, the neural system automatically adjusts itself and eventually generalizes it.

Neural networks are considered one of the most powerful techniques used in the area of data mining. The learning algorithms of neural networks can probe through data and learn relationships, which are not otherwise apparent.

The main advantage of neural network technology is that the internal representation and distribution of data need not be known. Although neural networks have not been tested extensively for modelling survival data, based on its predictive successes in other domains, it is considered a good alternative for the prediction of survival of individual patients. Neural networks also do not offer any obstacle to handling censored data.

Our proposed model, as shown in Figure 2, will take in prognostic factors as inputs to predict survival. The architecture of the network, the algorithm used to train the network, the number of neurons and the number of layers needed to produce the desired output will be the subject of the research. We would also like to use our trained network to predict the survival of individual NPC cases and subsequently the survival of new NPC cases.

SUMMARY

Statistical methods such as the life-table, the Kaplan-Meier method and regression models such as the Cox Proportional Hazard model are used to explain survival data and to model the progression of the disease rather than to make survival predictions for populations or individual patients [1] [3]. The Cox method is usually used to model the progression of disease by revealing the importance of covariates rather than to perform prognosis on new cases.
Statistics has been proven to be a useful tool in analysing large amounts of data especially in the field of cancer survival. However, statistics apply to large groups of people and may be meaningless for an individual. Statistics are based on estimates taken from a sample, and these estimates will differ from the figures obtained if a complete census had been taken using the same survey procedures. Thus, statistics do not and cannot predict events in the future with any certainty.

Neural networks have been shown to perform good results in predicting and we believe they have the ability to predict the prognosis of new cases. Neural network technology is relatively new in the field of cancer survival, however, based on its proven success in other fields neural network has the potential to be explored as an alternative technique to currently used statistical methods.

The next stage of our research will involve the analysis of survival data using conventional methods such as Cox and Kaplan-Meier. Subsequently, we intend to develop a computer-based cancer prognosis predictor at the individual level based on the parameters of individual risk factors, tumour stage, treatment and so on, by using artificial neural network technology. We would then like to show how ANN compares as a prognostic predictor with current statistical methods.

Acknowledgements

We are grateful to Professor U. Prasad of the ENT Department, University Hospital, University of Malaya, for kindly allowing us access to his department’s collection of NPC data and for his role as advisor in this project. One of the problems faced by any Malaysian researcher who would wish to undertake projects in medical informatics would be the availability of data. Most hospital administrators are unwilling to provide their patients’ data and there are no publicly accessible databanks such as the American Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. This project is funded by an IRPA grant from the Ministry of Science and Technology.

REFERENCES