Chapter 1
Advances in Intelligent Data Mining

Dawn E. Holmes, Jeffrey W. Tweedale, and Lakhmi C. Jain

1 Department of Statistics and Applied Probability,
University of California Santa Barbara,
Santa Barbara, CA 93106-3110, USA
holmes@pstat.ucsb.edu
2 Defence Science and Technology Organisation,
PO. Box. 1500, Edinburgh
South Australia SA 5111, Australia
jeffrey.tweedale@dsto.defence.gov.au
3 School of Electrical and Information Engineering,
University of South Australia, Adelaide,
Mawson Lakes Campus, South Australia SA 5095,
Australia
lakhmi.jain@unisa.edu.au

1 Introduction

The human body is composed of eleven sub-systems. These include the: respiratory, digestive, muscular, immune, circulatory, digestive, skeletal, endocrine, urinary, integumentary and reproductive systems [1]. Science shows how complex systems interoperate and have even mapped the human genome. This knowledge resulted through the exploitation of significant volumes of empirical data. The size of medical databases are many orders of magnitude those of text and transactional repositories. Acquisition, storage and exploitation of this data requires a disparate approach due to the modes and methods of representing what is being captured. This is significantly important in the medical field. As we transition from paper or film capture across to digital repositories, the challenges grow exponentially. The technological challenges compel the industry to undergo a paradigm shift that has resulted from the volume and bandwidth demanded of radiological imaging. Again, society is demanding instant access and analysis of diagnostic equipment to enable timely management of medical conditions or treatment. Such treatment also requires access to patient records, regardless of their source or location. This book examines recent developments in Medical, Health, Social and Biological applications.

More healthcare related data is being collected than ever before; for example, medication records for individual patients, epidemic statistics for public health specialists, data on new surgical techniques and drug interactions. Data mining techniques, such as clustering, k-means and neural networks, now hold a key role in knowledge discovery from medical data. Using prediction and classification techniques has enabled researchers to investigate the success of surgical procedures and the efficacy of medication. Much useful work has already been accomplished in the field of medical data.
mining; for example, heart attack prediction using neural network technology and classification of breast tumors. Mining EEG and EKG images is also a growth area. As well as technical issues, ethical, social and legal questions, such as data ownership and confidentiality arise from this area of research, making it truly interdisciplinary. The development of clinical databases has led to the burgeoning field of healthcare data mining, which we now explore.

2 Medical Influences

Science and technology has stimulated medicine to evolve diagnostic and treatment regimes. Information technology has enabled humanity to capture and store virtually every aspect of its existing and influence on life. Photography demonstrated the effectiveness of capture and storage in diagnosing or recording changes in our bodies. This method of diagnosis provides snapshots and enables comparative analysis. Medical practitioners have embraced this concept by embracing equipment with digital capture. The transition from paper to screen has forced the data mining community to adapt text based paradigms to data mining and data fusion technology in order to extract meaningful knowledge from diagnostic images and complex data structures [2].

3 Health Influences

eHealth has emerged as a significant benefit to society, allowing doctors to monitor, diagnose and treat medical condition more effectively than ever before. There are obvious benefits to hospitals and cost significant savings that flow on to patients. Medical information has unique characteristics, with significant duplicity and redundancy [3,4]. It is plagued by multi-attribution, incomplete information and time sensitive data which influences diagnostic techniques. Significant research continues to evolve methodology to expose viable and verifiable knowledge about specific conditions. Obviously the intent is to isolate each condition for individuals patients and concentrate on creating a healthy and productive society.

4 Social Influences

Data mining has influenced preventative health regimes by enabling improved coordination of immunisation, pandemic control and public or social policy. Technology enables improved access to societal data with respect to public health trends, attitudes and perception. The community can monitor and engage in effective public health campaigns or treatment. The recent swine flu pandemic is a prime example.

4.1 Information Discovery

In the field of Information Discovery, a family of automated techniques are used to sift through volume of data available in order to find knowledge. These nuggets can be deployed directly by applications to help focus a more appropriate outcome. Examples