Chapter XIII

Heuristics in Medical Data Mining

Susan E. George
University of South Australia, Australia

This chapter presents a survey of medical data mining focusing upon the use of heuristic techniques. We observe that medical mining has some unique ethical issues because of the human arena in which the conclusions are outworked. The chapter proposes a forward looking responsibility for mining practitioners that includes evaluating and justifying data mining methods–a task especially salient when heuristic methods are used. We define heuristics broadly to include those methods that are applicable for large volumes of data, as well as those specifically directed to dealing with uncertainty, and those concentrated upon efficiency. We specifically consider characteristics of medical data, reviewing a range of mining applications and approaches. We conclude with some suggestions directed towards establishing a set of guidelines for heuristic methods in medicine.

INTRODUCTION

Deriving—or discovering—information from data has come to be known as data mining. It is a popular activity in many domains, from stock market prediction to healthcare. There are varied and diverse applications of the knowledge derived from mining as conclusions are utilised in various capacities. Knowledge derived from medical mining has been used to assist tasks as diverse as patient diagnosis and inventory stock control; it has formed the knowledge behind intelligent interfaces for patient record systems and been the tool of medical discovery. This chapter

reviews some of the applications that have been made and identifies the need for a set of guidelines for heuristic methods in medicine.

Data mining presents many challenges since data is often present in huge volumes, distributed across many sources and highly complex in nature with many hundreds of variables and relationships among those variables varying in time, space or both often with a measure of uncertainty. Data mining is sometimes distinguished from statistics because of the vastness of the data sets with which data miners are concerned. Large data sets mean an increased algorithmic complexity and with that complexity comes the corresponding need to address issues in efficiency, hence the focus of this volume on a synergy of heuristics and data mining.

Artificial intelligence gives a very specific meaning to the term "heuristic." A heuristic is something that aids discovery of a solution. Heuristic methods are typically employed within the context of problem solving, where the solution to a problem must be found via a process of graph search (elements in the graph represent problem states, or operations that transform problem states). In search heuristic methods are able to guide exploration in an appropriate way and achieve faster solutions or more optimal solutions; occasionally the heuristic prevents a solution. Thus there are classic graph search algorithms such as the A* algorithm which is heuristic search (under the right conditions). However, we may broaden the definition of heuristics to include those techniques that are specifically relevant to dealing with large volumes of data or uncertainty within data. In doing so we can move away from traditional heuristic methods and encompass a wider range of techniques that may be regarded as "heuristic" because they "aid discovery," the crux of data mining.

In the remainder of this chapter we consider the nature of medical data mining and identify it as a unique arena for heuristic techniques. This is not just because the data is sensitive or highly private or that ethical concerns shadow every stage of mining from data collection to analytic procedures, from storage to appropriate data access; privacy, accuracy and security are issues in other domains (Walhstrom, Roddick, & Sarre, 2000). Rather, medical data mining is unique because of the implications for human life that may arise from the conclusions drawn. We propose that a set of guidelines for medical data mining is necessary, and that this is a natural consequence of forward-looking responsibility within this field. Forward-looking responsibility is accountable for high quality products and methods and requires appropriate evaluation of results and justification of conclusions. The proposed guidelines relate to the evaluation and justification of data mining results (so important when heuristic "aids to discovery" are utilised that "may" benefit a solution) and extend to both where and how the conclusions may be utilised and indeed where heuristic techniques are relevant in this field.

This chapter continues by providing a short review of some techniques that may be regarded as heuristic, including the classic A* search algorithm, tabu search, genetic approaches, fuzzy logic, rough sets, case-based reasoning and neural networks. We also examine the nature of medical data that present particular challenges for mining tasks and the diverse range of medical mining applications

Related Content

Spatial Navigation Assistance System for Large Virtual Environments: The Data Mining Approach

Mehmed Kantardzic, Pedram Sadeghianand Walaa M. Sheta (2008). *Mathematical Methods for Knowledge Discovery and Data Mining (pp. 265-283).*

www.irma-international.org/chapter/spatial-navigation-assistance-system-large/26145

Mining Unexpected Sequential Patterns and Implication Rules

Dong (Haoyuan) Li, Anne Laurentand Pascal Poncelet (2010). *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection (pp. 150-167).*

www.irma-international.org/chapter/mining-unexpected-sequential-patterns-implication/36905

Conceptualization of the Dike Distribution Analysis Aiming at Identification of Eruptive Centers

Sophie Konevaand Cyril Pshenichny (2018). *Dynamic Knowledge Representation in Scientific Domains (pp. 119-154).*

www.irma-international.org/chapter/conceptualization-of-the-dike-distribution-analysis-aiming-at-identification-of-eruptive-centers/200174

Mining Clinical Trial Data

Jose Ma. J. Alvirand Javier Cabrera (2007). *Knowledge Discovery and Data Mining: Challenges and Realities (pp. 31-47).*

www.irma-international.org/chapter/mining-clinical-trial-data/24900

Optimizing and Managing Digital Telecommunication Systems Using Data Mining and Knowledge Discovery Approaches

Adnan I. Al Rabeaand Ibrahiem M. M. El Emary (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains (pp. 66-84).*

www.irma-international.org/chapter/optimizing-managing-digital-telecommunication-systems/46891