Chapter 11

Knowledge-Based Decision Support System for Analyzing the Relevancies of Various Attributes Based on Their Characteristics

C. Madana Kumar Reddy
Annamacharya Institute of Technology and Sciences, India

ABSTRACT

Data mining extracts novel and useful knowledge from large repositories of data and has become an effective analysis and decision means in any organization. The resource of the World Wide Web is almost infinite. The growing importance of electronic media for storing and disseminating text documents has created an urgent need for tools and techniques that assist users in finding and extracting relevant and previously unknown information from massive collection of documents available in the web. Thus the development of techniques for mining unstructured, semi-structured, and fully structured textual data has become quite important in both academia and industry. Information management of well organized databases has been a focus of the Data mining research. When to specify too many attributes, system will slow down thus exclude irrelevant or weakly relevant attributes. The general idea behind attribute relevance analysis is to compute some measure that is used to quantify the relevance of an attribute with respect to a given class or concept.

INTRODUCTION

Plato said that "necessity is the mother of invention". Data mining has attracted the attention of many researchers due to the wide availability of huge amounts of data and the need for turning such data into useful information and knowledge. Data mining is applicable to any kind of data repository like relational databases, data warehouses, transactional databases, flat files, data streams, world wide web, object-

relational databases and specific application-oriented databases, such as spatial databases, time-series databases, text databases, and multimedia databases. The elimination of irrelevant or weakly relevant attributes with its characterization or comparison is referred to as analytical characterization or analytical comparison respectively. We are data rich, but information poor. Data mining extracts novel and useful knowledge from large repositories of data and has become an effective analysis and decision means in any organization. The resource of the world-wide web is almost infinite. The growing importance of electronic media for storing and disseminating text documents has created an urgent need for tools and techniques that assist users in finding and extracting relevant and previously unknown information from massive collection of documents available in the web (Agrawal, 1993). Thus the development of techniques for mining unstructured, semi-structured, and fully structured textual data has become quite important in both academia and industry. Information management of well-organized databases has been a focus of the Data mining research. Data mining is defined as the process of finding useful patterns from data, which is already existed in the database, like knowledge discovery in artificial intelligence. In our daily life we receive a lot of information every day and some of them are valuable, therefore we save it in data-base for future reference. But since the information's growth is very fast, may be one week or two this data-base is become huge in size, if we don't find a way to make it easy to handle and manage, we may have a lot of trouble when we reference it later. The knowledge discovery in databases is an iterative process. Once the discovered knowledge is presented to the user the evaluation measures can be enhanced, the mining can be further refined, new data can be selected or further transformed, or new data sources can be integrated in order to get different, more appropriate results. An attribute or dimension is considered as highly relevant with respect to a given class if it is likely that the values of the attribute or dimension are used to distinguish the class from others. Class characterization includes the analysis of attribute or dimension relevance is called analytical characterization. Include pictures also as the part of the relevance of the objects, because it will give the quick idea regarding the concept what you are expressing in your way related to the context. Plato said that "Picture is worth more than thousands of words". Thus include the pictures also as the part of the attribute relevance analysis of various attributes based on their characteristics for the analysis of the data. When to specify too many attributes, system will slow down thus exclude irrelevant or weakly relevant attributes. The general idea behind attribute relevance analysis is to compute some measure that is used to quantify the relevance of an attribute with respect to a given class or concept (David Hand, 2005). The measures in the data mining are scalability, interoperability flexibility. For mining class characteristics, there is only one class to be characterized; the contrasting class is taken to be the set of comparable data in the database. Users can set attribute generalization thresholds. In general generalize dimension place to the country level. Even without explicit user instruction, a default value can be set by the data mining system, which allows each dimension to be generalized to a level that contains very few i.e., two to eight distinct values. By calculating information gain we can remove the less informative attributes by retaining the more informative attributes for use in the concept description analysis. Analytical characterization should be performed if the mined concept descriptions involve many attributes. As the part of it first remove weakly relevant or irrelevant attributes before performing the generalization. Also compute the information gain for each attribute and the result is sorted in increasing order. Users may not be interested in having a single class. Class discrimination or comparison mines descriptions that distinguish a target class from its contrasting classes. Target & contrasting classes must share similar dimensions and attributes. This chapter presents various techniques and their effectiveness with respect to the attribute relevance.

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/knowledge-based-decision-support-system-for-analyzing-the-relevancies-of-various-attributes-based-on-their-characteristics/173831

Related Content

Performance Measurement: From DEA to MOLP
João Carlos Namorado Clímaco, João Carlos Soares de Melloand Lidia Angulo Meza (2008). *Encyclopedia of Decision Making and Decision Support Technologies (pp. 709-715).*
www.irma-international.org/chapter/performance-measurement-dea-molp/11312

A Novel Query Method for Spatial Database Based on Improved K-Nearest Neighbor Algorithm
Huili Xiaand Feng Xue (2024). *International Journal of Decision Support System Technology (pp. 1-15).*
www.irma-international.org/article/a-novel-query-method-for-spatial-database-based-on-improved-k-nearest-neighbor-algorithm/332773

Argument Mapping and Content Fusion to Support the Analysis and Synthesis of Information in Online Discussions
Ali Gürkanand Luca Iandoli (2014). *International Journal of Decision Support System Technology (pp. 14-33).*
www.irma-international.org/article/argument-mapping-and-content-fusion-to-support-the-analysis-and-synthesis-of-information-in-online-discussions/117686

Methodological Aspects of the Evaluation of Individual E-Banking Services for Selected Banks in Poland
Witold Chmielarz (2010). *Infonomics for Distributed Business and Decision-Making Environments: Creating Information System Ecology (pp. 201-216).*
www.irma-international.org/chapter/methodological-aspects-evaluation-individual-banking/38423

Notifiable Disease Databases for Client Management and Surveillance
Ann M. Jollyand James J. Logan (2017). *Handbook of Research on Data Science for Effective Healthcare Practice and Administration (pp. 386-416).*
www.irma-international.org/chapter/notifiable-disease-databases-for-client-management-and-surveillance/186949