

<<现代计算技术与中医药信息处理>>

图书基本信息

书名：<<现代计算技术与中医药信息处理>>

13位ISBN编号：9787308084574

10位ISBN编号：7308084574

出版时间：2012-8

出版时间：浙江大学出版社

作者：吴朝晖，陈华钧，姜晓红 编

页数：233

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

内容概要

Elsevier Insights provides high quality specialized content , across a range of disciplines including life sciences , physical sciences , social sciences , engineering , computing , and finance. Through fast-track publication , Elsevier Insights offers the reader cutting-edge information , available in eBook or print format.

书籍目录

Preface
 1 Overview of Knowledge Discovery in Traditional Chinese Medicine
 1.1 Introduction
 1.2 The State of the Art of TCM Data Resources
 1.2.1 Traditional Chinese Medical Literature Analysis and Retrieval System
 1.2.2 Figures and Photographs of Traditional Chinese Drug Database
 1.2.3 Database of Clunese Medical Formulae
 1.2.4 Database of Chemical Composition from Chinese Herbal Medicine
 1.2.5 Clinical Medicine Database
 1.2.6 TCM Electronic Medical Record Database
 1.3 Review of KDTCM Research
 1.3.1 Knowledge Discovery for CMF Research
 1.3.2 Knowledge Discovery for CHM Research
 1.3.3 Knowledge Discovery for Research of TCM Syndrome
 1.3.4 Knowledge Discovery for TCM Clinical Diagnosis
 1.4 Discussions and Future Directions
 1.5 Conclusions
 2 Integrative Mining of Traditional Clunese Medicine Literature and MEDLINE for Functional Gene Networks
 2.1 Introduction
 2.2 Connecting TCM Syndrome to Modern Biomedicine by Integrative Literature Mining
 2.3 Related Work on Biomedical Literature Mining
 2.4 Name Entity and Relation Extraction Methods
 2.4.1 Bubble-Bootstrapping Method
 2.4.2 Relation Weight Computing
 2.5 MeDisco/3SSystem
 2.6 Results
 2.6.1 Functional Gene Networks
 2.6.2 Functional Analysis of Genes from Syndrome Perspective
 2.7 Conclusions
 3 MapReduce-Based Network Motif Detection for Traditional Clunese Medicine
 3.1 Introduction
 3.2 Related Work
 3.3 MapReduce-Based Pattern Finding
 3.3.1 MRPF Framework
 3.3.2 Neighbor Vertices Finding and Pattern Initialization
 3.3.3 Pattern Extension
 3.3.4 Frequency Computing
 3.4 Application to Prescription Compatibility Structure Detection
 3.4.1 Motifs Detection Results
 3.4.2 Performance Analysis
 3.5 Conclusions
 4 Data Quality for Knowledge Discovery in Traditional Chinese Medicine
 4.1 Introduction
 4.2 Key Data Quality Dimensions in TCM
 4.2.1 Representation Granularity
 4.2.2 Representation Consistency
 4.2.3 Completeness
 4.3 Methods to Handle Data Quality Problems
 4.3.1 Handling Representation Granularity
 4.3.2 Handling Representation Consistency
 4.3.3 Handling Completeness
 4.4 Conclusions
 5 Service-Oriented Data Mining in Traditional Chinese Medicine
 5.1 Introduction
 5.2 Related Work
 5.2.1 Traditional Data Mining Software
 5.2.2 Data Mining Systems for Specific Field
 5.2.3 Distributed Data Mining Platform
 5.2.4 The Spora Demo
 5.3 System Architecture and Data Mining Service
 5.3.1 Hierarchical Structure
 5.3.2 Service Operator Organization
 5.3.3 User Interaction and Visualization
 5.4 Case Studies
 5.4.1 Case 1: Domain-Driven KDD Support for TCM
 5.4.2 Case 2: Data Mining Based on Distributed Resources
 5.4.3 Case 3: Data Mining Process as a Service
 5.5 Conclusions

 6 Semantic E-Science for Traditional Clunese Medicine
 7 Ontology Development for Unified Traditional Chinese Medical Language System
 8 Causal Knowledge Modeling for Traditional Chinese Medicine Using OWL
 9 Dynamic Subontology Evolution for Traditional Chinese Medicine Web Ontology
 10 Semantic Association Mining for Traditional Chinese Medicine
 11 Semantic-Based Database Integration for Traditional Chinese Medicine
 12 Probabilistic Semantic Relationship Discovery from Traditional Chinese Medical Literature
 13 Deriving Similarity Graphs from Traditional Chinese Medicine Linked Data on the Semantic Web

章节摘录

to use different notions and expressions to describe one concept. When the information in different dynasties is collected in databases, the problem of representation inconsistency is also introduced. For instance, ginseng Panax and Radix ginseng could both refer to the Chinese herbal medicine ginseng in English. This situation becomes more complex in Chinese: there are 10 aliases for ginseng. Another example is the inconsistent weight units used in different Chinese medical formulae, which has been mentioned before. Before data analysis and knowledge discovery can be carried out on these data, such representation consistency issues must be addressed to ensure the final reliability.

4.2.3 Completeness One of the biggest problems hampering the effective usage of TCM resources is the incompleteness of data. Take DCMF for an example; two crucial attributes of DCMF are ingredients and efficacy. The attribute ingredients have already been described, and efficacy is a textual attribute containing a description of the remedy principle in the TCM background. Due to historical reasons, among 85,917 valid records wherein the attribute value of ingredients is not null, only 15,671 records are stored with efficacy not null. That is to say, 81.76% of data in attribute efficacy is missing. Identifying such phenomenon in TCM data and treating this problem is an important task in data analysis.

4.3 Methods to Handle Data Quality Problems Due to the existence of data quality problems mentioned previously, it is extremely important to conduct necessary data preprocessing activities for data analysis and knowledge discovery. Jiang et al. indicated that data preprocessing was the key to the knowledge discovery of the compatibility rule of TCM formulae. Thus, it is of vital necessity to explore preprocessing methods of TCM data. The data quality problems mentioned in the last section are the main obstacles in TCM on the way to high data quality. In this section, we introduce the preprocessing methods used to handle these problems.

4.3.1 Handling Representation Granularity The procedure we conduct to treat the representation granularity problem is called structuring, i.e., to structure a data field with multiple data elements into multiple separate data fields. To handle the example problem of representation granularity mentioned in the previous section, a concept of a herb information unit (HIU) is defined, which is the name of Chinese herbal medicine, followed by the preparation method, dosage, and weight unit. With this perspective, we could see that the attribute ingredients usually consist of multiple HIUs separated by commas. To effectively use all information in this field, we should first split ingredients into multiple HIUs. Secondly, for each HIU, we further divide it into four fields: the name of Chinese herbal medicine, preparation method, dosage, and weight unit. To perform this two-step extraction, there are a lot of details and exceptions that should be noticed in practice. For instance, in many records, the delimiter comma might be replaced by semicolon/period, or even be missing; the preparation method/dosage/weight unit is also missing or misspelled in many records. To implement the two-step splitting, a splitting-rule-based system named field splitter was developed in 2003 to handle this problem. Tens of specific splitting rules, such as "keep between A and B" and "replace A with B", are defined. Users can form their own splitting setting by organizing these rules. The system field splitter is found to work well for these years. This is the structuring method we use to fight representation granularity problems in TCM.

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>