<<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

内容概要

Elsevifer Insights provides high quality specialized content, across a range of disc, iplines including Iife sciences, pliysical sciences, social scicnces, engineering, compuLing, and flnance.Through fast-track publicalion, Elscvier Insights offers the reader cutting-edge inforniation, available in eBook or print format.

<<现代计算技术与中医药信息处理>>

书籍目录

章节摘录

to use different notions and expressions to describe one concept.When the infor-mation in different dynasties is collected in databases, the problem of representa-tion inconsistency is also introduced.For instance, ginseng Panax and Radixginseng could both refer to the Chinese herbal medicine ginseng in English.Thissituation becomes more complex in Chinese： there are 10 aliases for ginseng.Another example is the inconsistent weight units used in different Chinese medicalformulae, which has been mentioned before.Before data analysis and knowledgediscovery can be carried out on these data, such representation consistency issuesmust be addressed to ensure the final reliability. 4.2.3 Completeness One of the biggest problems hampering the effective usage of TCM resources isthe incompleteness of data.Take DCMF for an example; two crucial attributes ofDCMF are ingredients and efficacy.The attribute ingredients have already beendescribed, and efficacy is a textual attribute containing a description of the remedyprinciple in the TCM background.Due to historical reasons, among 85 , 917 validrecords wherein the attribute value of ingredients is not null, only 15, 671 recordsare stored with efficacy not null.That is to say, 81.76% of data in attribute efficacyis missing.Identifying such phenomenon in TCM data and treating this problem isan important task in data analysis. 4.3 Methods to Handle Data Quality ProblemsDue to the existence of data quality problems mentioned previously, it is extremelyimportant to conduct necessary data preprocessing activities for data analys, e andknowledge discovery.Jiang et al. indicated that data preprocessing was thekey 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 dataquality problems mentioned in the last section are the main obstacles in TCM onthe way to high data quality.In this section, we introduce the preprocessing meth-ods used to handle these problems. 4.3.1 Handling Representation Granularrty The procedure we conduct to treat the representation granularity problem is calledstructurizing, i.e., to structurize a data field with multiple data elements .into multi-ple separate data fields.To handle the example problem of representation granular-ity mentioned in the previous section, a concept of a herb information unit （HIU） isdefined, which is the name of Chinese herbal medicine, followed by the preparationmethod, dosage, and weight unit.With this perspective, we could see that the attri-bute ingredients usually consist of multiple HIUs separated by commas.To effec-tively use all information in this field, we should first split ingredients into multiple HIUs.Secondly, for each CIU, we further divide it into four fields： the name ofChinese herbal medicine , preparation method, dosage, and weight unit.To performthis two-step extraction, there are a lot of details and exceptions that should benoticed in practice.For instance, in many records, the delimiter comma might bereplaced 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 in2003 to handle this problem.Tens of specific splitting rules, such as “ keep betweenA and B ” and “ replace A with B ”, are defined.Users can form their own splittingsetting by organizing these rules.The system field splitter is found to work well forthese years.This is the structurizing method we use to fight representation granu-larity problems in TCM. ……

版权说明