**分类号：TP315 U D C：D10621-408-(2013)1104-0**

**密 级：公 开 编 号：2009051013**

**成都信息工程学院**

**学位论文**

**拟周期及其关联规则采掘的研究与实现**

|  |  |
| --- | --- |
| **论文作者姓名：** | **\*\*\*** |
| **申请学位专业：** | **计算机科学与技术** |
| **申请学位类别：** | **工学学士** |
| **指导教师姓名（职称）：** | **\*\*（讲师）** |
| **论文提交日期：** | **2013年06月02日** |

**拟周期及其关联规则采掘的研究与实现**

**摘 要**

数据采掘是数据库技术、人工智能、机器学习、统计分析、模糊逻辑、模式识别、人工神经网络等多个学科相结合的产物。其中的时序数据（Time Series）采掘因其广泛的应用性与极高的商业价值成为了当前数据采掘研究的热点之一。 本文的主要工作集中在：

 1. 述了当今国际上的一些成熟的时序采掘的产品和时序采掘的研究现状并提出了自己的看法。

2. 针对以上的工作盲点提出拟周期等六个概念、抗干扰势态等五个算法和两个定理，建立了拟周期及其关联规则的采掘模型。在此模型上进行拟周期及其关联规则采掘系统RPMiner的结构和模块设计。

3. 使用Visual C ++ 中的ODBC技术实现了RPMiner的各个功能模块。自行设计的源程序共有850K。

4. 对安宁河断裂带地震数据库数据进行试采掘，分析其采掘结果得出了两个出人意外的结果，一个是∶安宁河断裂带以5周为小活动周期，而5个月为较大的活动周期；另一个是∶在安宁河断裂带的北南方向，地形形变与地震的同步性比较明显。

本文组织如下：

* 第一章介绍了数据采掘的基本概念和有关技术。
* 第二章介绍了在数据采掘中当今时序采掘的产品方面和研究方面的情况，并总结了其特点与盲点。
* 第三章，介绍了RPMiner系统的设计，包括要实现的目标和提出来的基本概念。
* 第四章，介绍了RPMiner的实现，包括用以实现RPMiner的Visual C++的关于ODBC的一些技术，以及PRMiner的使用介绍。第五章给出了对安宁河地震断裂带的地震数据库数据的试采掘结果分析。最后在第六章浅谈了我对数据采掘的发展趋势的一些看法。

**关键词：**数据采掘 时序采掘 拟周期 关联规则

**Research and Implementation of Mining Relaxed Periods and their Association Rules**

**Abstract**

Data Mining is a main step of KDD processes, and draws upon many techniques from diverse fields, such as database technology, artificial intelligence, machine learning, statistics, fussy logic, pattern recognition, and artificial neural network, etc. Mining on Time Series becomes a hot topic of Data Mining due to its wide applications and high commercial value. The main contribution of this paper includes:

1. Survey the current mature products and research harvests internationally;
2. Propose six concepts of “Relaxed-Period” etc, five algorithms of  “anti-noise tendency” etc. and two theorems to fill the blind spot of the above researches, forming the model of Mining Relaxed Periods and their Association Rules;
3. Based on the previous model, design the system structure and all the sub-models of RPMiner; A prototype called RPMiner is implemented based on ODBC and Visual C ++. All the codes written by myself are almost 850K.
4. Mining the seismic data of the fault belt along the River ANNING and analyzing the results, two surprising results are uncovered, one is that 5 weeks is the shorter periodicity while 5 months is the longer periodicity, the other one is that the reform in the North-South direction gives a remarkable contribution the earthquake magnitude.

 The theses is organized as follows:

* Section 1 introduces some basic concepts and technology about data mining. Section 2 gives the survey on currently international products and research harvests, summarizing their characters and blind spots.
* Section3 is all about the design of the system RPMiner, including the actualization goals and the basic concepts.
* Section 4 tells the whole procedure of actualizing the RPMiner, including the synopsis of the Visual C++ ODBC technology and the overlook of RPMiner. Section5 gives the background of the mining data of the fracture belt along the River ANNING as well as the analysis of the mining results. At last, in Section 6, some of my personal opinions of the developing trend of data mining are     proposed.

**Key words:** Data Mining, Time Series, Relaxed Period, Association Rule