

学校编码: 10384

分类号_____密级_____

学号: X2013232308

UDC_____

廈門大學

工程硕士学位论文

数据挖掘分析技术在商标侵权案例分析
中的应用研究

Study on the Application of Data Mining Technology in
Analysis of Trademark Infringement Cases

肖冰

指导教师姓名: 陈海山 教授

专业名称: 软件工程

论文提交日期: 2015 年 09 月

论文答辩时间: 2015 年 11 月

学位授予日期: 年 月

指导教师: _____

答辩委员会主席: _____

2015 年 月

厦门大学博硕士学位论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外，该学位论文为()课题(组)的研究成果，获得()课题(组)经费或实验室的资助，在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士学位论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

- ()1.经厦门大学保密委员会审查核定的保密学位论文，于
 年 月 日解密，解密后适用上述授权。
- ()2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年 月 日

厦门大学博硕士学位论文摘要库

摘要

随着我国司法公开的逐步推进，以及法院信息化建设工作的逐步开展。各级人民法院在多年的司法审判活动中所积累的大量案例，成为公开的、可供研究与分析的案例信息。也是由于法院判决书中所记载的各项内容，具有非常高的准确性。因此，能够为各类研究提供极为可靠的数据来源。

由于研究技术和方法的局限，判决书中所包含的高质量数据信息并没有被完全发掘。一般的统计分析方法，并不能够全面发现案例数据背后所隐藏的知识与规律。因此，如何充分利用这些已公开的案例信息数据，就是一个非常值得研究的问题。而数据挖掘技术，正是在信息技术发展的背景下，对海量数据进行分析与研究过程而产生的。因此，数据挖掘的定义，就是从大量的、不完全的、有噪声的甚至模糊的数据中，提取我们所感兴趣的知识。于是，在上述背景下，将数据挖掘技术，应用于司法案例信息的研究之中。通过相关的分析方法对特定领域中的案件特征属性加以提取，分析可能影响法院判决结果的因素，挖掘并发现与案件审判结果存在一定关联的规则。进而生成用于特定类型案件特征分类和判决结果预测的模型，服务于司法实践与法学研究。为现实中当事人的诉讼策略选择、法院审理案件的标准以及理论研究提供辅助决策。

面对现有司法案例数据库只提供了基本案例检索和查看功能的现实情况，本论文通过将数据仓库技术与数据挖掘技术相结合的方法，对案例信息数据进行管理和分析。通过数据挖掘的分析方法，分析并归纳出具有实用价值与研究价值的知识，实现辅助决策的目的。具体而言，本文研究内容分为以下方面：

1.通过对以公开的判决书进行整理与收集，作为数据源。在此，本文选择了案由为“商标侵权”的案件作为研究对象；

2.对上述判决书的内容进行分析，并按照本研究的需求，对案例信息数据进行预处理，建立相应的案例信息数据仓库；

3.采用决策树分析、聚类分析以及关联分析的方法，对案例信息数据进行研究与分析。

4.综合上述研究与分析的结果，形成能够服务于司法实践与法学理论研究的结论，实现为法律的制定、实施与修改提供辅助决策的目的。

关键词：司法案例；数据挖掘；数据仓库

厦门大学博硕士学位论文摘要库

Abstract

With the development of information technology and the reform of the judicial system has been promoting in recent years. The judgments which made by Local people's courts at various levels has been become a part of publicly available information. Due to the contents of judgment are strictly verified, it can be regard as an accuracy and reliability source of date. Therefore the judicial precedent or the content of judgments has become an important research object in field of legal research.

Limited by the methodology and technique of the research, the high quality data included in the judgments were not fully analyzed and utilized. Consequently, it is worth studying on how to take full advantage of this data and information. In consideration of both feasible in practice and scientific in theory, the Data Mining technology which the goal is the extraction of patterns and knowledge from large amount of data, is satisfy the needs of research.

Face the fact of existing systems could only provide the function of judgments search and browse. In order to taking full use of the information in cases and discover the influence factors of judgments even its association rules which behind the data. This dissertation, data warehouse and data mining technologies are applied to analysis the data in judgments and cases. The thesis mainly engaged in the work as the following aspects:

First, it takes the trademark infringement cases as the object of the research. Collation the data of judgments cases and judicial precedent and creating a data source.

Second, it through a process of extraction, cleaning, conversion and loading of data. Then building judgments and cases data warehouse.

Third, it uses the way of decision tree, correlation analysis and cluster analysis, to identify the relationship of mutual relations between each factors and judicial decisions.

Finally, Analysis and summary of the results of the research and stated the conclusion of this thesis.

Keywords: Judicial Precedent; Data mining; Data Warehouse

厦门大学博硕士学位论文摘要库

第一章 绪论	1
1.1 研究背景	1
1.2 研究意义	1
1.3 研究现状	2
1.4 研究目标	3
1.5 论文结构	4
第二章 数据仓库与数据挖掘技术	5
2.1 数据仓库	5
2.2 数据挖掘技术	7
2.3 本章小结	12
第三章 司法案例数据仓库设计	13
3.1 设计方法	13
3.2 设计流程	13
3.3 数据仓库建模	14
3.4 数据处理	18
3.5 本章小结	22
第四章 决策树算法在案例分析中的应用	23
4.1 决策树的概念	23
4.2 决策树生成算法	23
4.3 模型设计	25
4.4 模型实现	30
4.5 本章小结	35
第五章 聚类算法在案例分析中的应用	37
5.1 聚类概述	37
5.2 模型设计	38
5.3 模型实现	39

5.4 本章小结	43
第六章 关联规则算法在案例分析中的应用	45
6.1 关联规则概述	45
6.2 关联规则生成算法	46
6.3 模型设计	48
6.4 模型实现	49
6.5 本章小结	51
第七章 总结与展望	53
7.1 总结.....	53
7.2 展望.....	53
参考文献	55
致谢.....	57

Contents

- **Chapter One Introduction1**
 - 1.1 Research Background.....1
 - 1.2 Research Meaning.....1
 - 1.3 Research Status2
 - 1.4 Research Objective3
 - 1.5 Thesis Structure4
- **Chapter Two Data Warehouse and Data Mining Technology5**
 - 2.1 Data Warehouse5
 - 2.2 Data Mining.....7
 - 2.3 Summarization12
- **Chapter Three Design and Realization Data Warehouse of Judgment13**
 - 3.1 Design Method.....13
 - 3.2 Design Process13
 - 3.3 Data Warehouse Modeling14
 - 3.4 Data Processing18
 - 3.5 Summarization22
- **Chapter Four Application of Decision Tree Algorithm in Case Analysis23**
 - 4.1 Conception of Decision Trees23
 - 4.2 Decision Tree Algorithm23
 - 4.3 Model Design25
 - 4.4 Implementation Model30
 - 4.5 Summarization35
- **Chapter Five Application of Clustering Algorithm in Case Analysis37**
 - 5.1 Conception of Clustering.....37
 - 5.2 Model Design38
 - 5.3 Implementation Model39
 - 5.4 Summarization43

- **Chapter Six Application of Association Rules Algorithm In Case Analysis45**
 - 6.1 Conception of Association Rule 45
 - 6.2 Algorithm of Association rule 46
 - 6.3 Model Design 48
 - 6.4 Implementation Model 49
 - 6.5 Summarization 51
- **Chapter Seven Conclusion and Prospect.....53**
 - 7.1 Conclusion 53
 - 7.2 Prospect..... 53
- **References55**
- **Acknowledgements57**

第一章 绪论

1.1 研究背景

判决书指法院依照判决而写成的一种法律文书，包括案件所涉及的争议事项、案件事实、证据和裁判结果等内容。现实中，全国各级人民法院也在多年的司法审判活动中积累了数量可观的判决书。但长期以来，受制于技术的发展与司法改革的进程，这些海量的判决书并没有被集成为一个完整的案例数据库。近年来，随着信息技术的发展及其在司法审判活动中的应用，伴随着司法判决上网公开工作的展开与深入。使得以法院判决书为代表的司法文书，逐渐成为公开的、通过信息化途径可以获取的信息数据。在上述背景下，相关的司法判例数据库也逐渐问世，比较知名的有中华人民共和国最高人民法院制作的“裁判文书网”和北大英华科技有限公司出品的“北大法宝”。

由于法院判决书中所包含的信息本身具有非常高的准确性和可靠性，因而具有极高的理论研究价值。在实现了信息公开之后，就可以对判决书中的数据进行研究与分析。但是，目前已有的数据库只能提供最基本的司法判例信息检索功能，而对于法学专业的研究人员而言，应该对这些数据的进行更为全面与深入分析，以期挖掘出更多具有研究价值的信息，实现对这些数据的充分利用。通过这些信息，对法院审判改革与立法等方面提供决策建议与意见。正如《人民法院报》在2013年的一篇报道中指出：大数据的关键在于通过对海量数据的分析加工实现数据的“二次利用”。强化海量数据的集中存储，做到安全保存和灵活调用；加强对海量数据的自动挖掘和分析，特别是在大量裁判案例基础上，运用大数据分析方法对法院各类案件的审判工作建构裁判模型，为法官裁判案件提供参考。

本文提出以数据挖掘技术为基础对司法案例信息进行分析的设想，旨在通过对司法案例信息数据进行统计、分析和推理。使用数据仓库和数据挖掘技术，通过对司法判决案例进行数据挖掘分析，发现司法审判工作的一般规律和趋势，总结影响司法审判结果的各种条件。寻找出数据信息之间的内在联系以及规律，以期实现为提高司法审判效率，法律修改等方面提供辅助决策的效果。

1.2 研究意义

1.对司法案例进行分析的重要性

司法案件的审判结果不仅与当事人的利益直接相关，也是直接反映我国司法是否公平、公正的重要标志。全国各级人民法院在多年的审判活动中，形成了数量庞大的案例。但由于信息化建设滞后的原因，大量的司法案例并未能够转化为可供研究与分析的信息数据。近年来，在新的指导思想要求下，司法判例逐步在网上公开，成为了可供广大法律工作者与法学研究者分析与参考的信息。但是，目前对于已经公开的司法案例，相关研究仅停留在较为初级的描述性分析阶段。在理论研究领域，案例审判结果分析的研究也较为缺乏。仅从原告是否胜诉的角度进行统计，而对影响案件审判结果的因素没有进行深入分析。这样就导致了大量的案例信息数据没有得到有效与充分的利用，更没有基于对相关数据的分析，发现我国司法审判工作现实情况以及可能存在的问题。

随着数据挖掘技术的兴起，也有学者将相关技术应用于法律问题的分析之中。例如通过聚类分析的方法，对特定违法犯罪行为的发生加以识别；通过决策树技术，对特定行为的法律风险或道德风险进行评估等等。这些研究对认识我国司法现状起到了一定的指导作用。因此，将这项技术应用法学实证研究领域，可以对司法案例进行准确、全面与客观的定量研究。尽可能地挖掘出司法审判的结果与诸多因素，特别是人为因素之间所隐藏的内在联系，并从中提炼出各项具体的分类规则。进而，发现司法审判活动中所存在的现实问题，借此帮助法学研究学者甚至立法机关制定相应的措施。推动我国司法审判活动沿着正确的方向发展。

2.现有案例分析技术的不足与应对方案

目前，在法律研究领域中的实证分析仍停留在简单的数据统计阶段。不仅研究样本数量相对偏少，同时数据分析技术也较为简易。正是由于研究方法的不足，导致了相关研究仅能利用最基本的统计量进行分析，而无法对影响司法审判判决结果的因素进行深入与具体的分析。

1.3 研究现状

所谓数据库中发现知识这一概念，最早出现于 1989 年 8 月召开的第一届 KDD 国际会议上^[1]。相比之下，国内对数据挖掘的研究是在 90 年代中期开始起步的。相应的，数据挖掘技术在法学研究领域的运用则是近年来才逐渐出现的。较具有代表的有 2010 年浙江省高级人民法院，以全省法院数据为基础，通

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.