西华大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名：程政　　指导教师签名：[illegible]

日期：2011年6月7日　　日期 2011.6.7

西华大学学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，在校攻读学位期间论文工作的知识产权属于西华大学，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，西华大学可以将本论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复印手段保存和汇编本学位论文。(保密的论文在解密后遵守此规定)

学位论文作者签名：程政　　指导教师签名：[illegible]

日期：2011年6月7日　　日期 2011.6.7

摘　要

安全性评价是一种现代化的管理手段，是企业管理的重要内容。对电力系统进行安全性评价，其目的在于夯实电力系统的安全基础。对电网运行、生产环境等进行安全性评价有利于实现对电网事故的超前预测和控制，进而减少和消灭事故的发生，保证电力系统的安全运行。

目前，安全性评价已经在电力系统中得到广泛应用，并取得了良好的效果。但还存在很多不足：①查评的结果数据未得到充分的分析和管理；电力系统在一次查评结束后，只对数据进行了简单的处理，没有做深入的分析和研究。②安全性评价未与“危险点分析”相结合。安全性评价主要用于一个企业全局宏观上的风险评估，危险点分析则是主要用于作业现场、生产岗位局部微观上的评估，二者各有特点，互为补充，不可替代，应结合进行。

针对上述问题，本文设计了一套电力系统安全性评价辅助决策系统，系统中运用了数据仓库和数据挖掘技术：安全性评价自查评数据通过分析和整理放入到数据仓库，通过数据仓库，用户可以查询到自己想要的信息，并通过直观图展示出来；数据挖掘技术的核心是：采用 Apriori 算法查找关联规则从而判断影响电网安全性的危险点。通过数据挖掘技术对数据仓库中的数据进行挖掘，得到电力系统中存在的危险点并找出影响危险点的各种因素，找出自查评项目存在的“危险点”，针对“危险点”，企业管理者可以加强此项目的管理，从而指导决策者消除危险点。

通过辅助决策系统，把电力系统的宏观评价和微观评价结合起来，对电力系统的安全稳定运行必定产生良好的影响。

关键词：安全性评价；数据仓库；数据挖掘，Apriori 算法；辅助决策

Abstract

Safety evaluation is a modern management tools and an important part of business management.The power system safety evaluation,which aims to lay a basis solid for power system security.Safety evaluation to grid operation and production environment conducive to realize the ahead prediction and control on grid accidents,reduce and eliminate accidents and ensure the safety operation of power system.

Currently,the safety evaluation has been widely used in power systems,and have achieved good results.But there are still many deficiencies:first,The results of assessment data has not been sufficient analysis and management;power system assessment in an investigation after the data were only simple processing,did not do in-depth analysis and research.The second, safety evaluation is not combine with the"dangerous points analysis".Safety evaluation is mainly used for a macro risk assessment for an enterprise in terms of overall situation,while dangerous points analysis is applied for micro partial assessment on working site and position.Both methods have their own characteristics which are complementary and cannot be replaced by each other and thus should be taken together according to realistic.

In response to these problems,we designed an Assistant Decision-making System for Power Safety Evaluation,the system using data warehouse and data mining techniques:safety evaluation of self-examination and assessment by analyzing and compiling into the data warehouse,though the data warehouse,users can query the information they want,and displayed through an intuitive map;The core of the tool of data mining is judging the dangerous points that may affect power network security by using association rules based on Apriori algorithm.We can find out the dangerous pionts and the various affecting factors that might exist in power system by mining the data of warehouse on the basis of data mining technology.Managers can enhance the management of this project from dangerous points,and guide decision-makers to eliminate the dangerous points.

The Assistant Decision-making System which analyzes on the electric power system by using the macro evaluation and micro evaluation together will certainly has positive effects on the safety of the system.

Key Words: safy evaluation; data warehouse; data mining; Apriori algorithm; assistant decision-making

目 录

1 绪论

1.1 系统的研究背景及意义

1.1.1 系统研究的背景

随着社会经济的发展，高新技术的大量应用，电力建设也在迅速发展，电力企业要适应当今的发展趋势，要在激烈的竞争中追求成功，不断增强实力和竞争力，最重要的、本质的要求就是发展安全生产。电力安全生产是关系国计民生的大事，对国民经济和人民生活关系极大，电力安全生产的好坏是社会经济的晴雨表，提高电力行业的安全生产水平，有利于促进社会和谐和社会稳定，促进我国早日全面实现小康社会。电力要先行，电力安全生产是保证。

然而，一方面由于我国电力工业的特点及设备、人员、管理以及运行环境等诸多方面的原因，目前许多不安全的因素普遍存在于电力系统中；另一方面由于管理者对各种风险评价方法的认识存在局限性或对评价方法选用不当，造成那些潜在的不安全因素未被人们所辨识，因此难以采取必要的措施、安全对策加以预防，从而导致安全事故的发生，这些事故对职工的生命健康和人身安全造成危害，甚至形成灾难。

因此，为了突出“安全第一，预防为主”，提高电力系统安全事故的可预见性，达到对可能发生的事故实现超前控制、将各种事故消火在隐患之中的“预防为主”的现代文明生产思想的目的，如何正确、合理的选用一种科学的方法对电力系统中可能发生的事故及其概率的高低进行分析和预测，保障电力工作者在生产过程中的人身安全和健康，成为了整个电力系统必须解决的重要问题。

安全性评价作为一门新兴的软科学已经被安全管理时间证明是一种有效落实“安全第一，预防为主”的科学管理方法，它对系统存在潜在的危险进行分析和度量，即对系统危险程度用安全标准来衡量，得出系统安全水平的估计，对企业系统地分析安全隐患，提高企业安全管理水平发挥了越来越大的作用。

目前，安全性评价已经用于电力系统的各个方面，并取得了巨大成功。电力系统逐渐形成了“自查-整改-复查”的良好机制，安全性评价的应用使电力系统的事故率呈逐年下滑趋势。但是，安全性评价工作并非一劳永逸，评价工作不能仅仅停留在表面，应该对安全性评价遗留下来的历史数据进行整理和充分研究，从数据中真正挖掘出电力系统存在的危险点。

随着形势的变化，人们对安全和健康的认识日益加深，新技术的应用、人员素质的提高也为电力系统安全管理提出了更高要求，必须完善和改进安全性评价体系。

基于上述背景，本文作者认为研究并开发一套具有辅助决策作用的风险评价软件系统，是解决背景中出现的关键问题的一条有效途径。

1.1.2 系统研究的意义

安全性评价通过评价一个系统或一个企业安全水平或风险程度，预知和掌握客观存在的危险因素及其严重程度，明确预防事故的重点和需要采取反事故的措施，实现超前控制，减少和杜绝事故。电力系统安全性评价，就是采用系统分析的科学方法，确认电力系统中存在的危险因素，并根据其形成事故风险的大小，采取相应的设备整改和安全措施，来达到电力系统的过程评定或全面评定。安全性评价运用安全系统工程的方法对系统的安全性进行预测和度量，通过对系统存在的危险性进行定性和定量的分析，确认系统发生危险的可能性及其严重程度，提出必要的措施，以寻求最低的事故率、最小的事故损失和最优的安全投资效益[1,2]。开展安全性评价工作，通过对人的不安全因素的全面查评、物的不安全情况的全面查评、环境的不安全问题的全面查评，发现问题、解决问题，克服人的不安全行为、消除物的不安全状态、改变环境的不安全因素，最终达到提高企业安全管理水平的目的。

电力系统安全性评价辅助决策系统致力于电力系统的微观评价，以电力系统安全性评价为基础，对电力系统中的安评数据进行分析，通过数据挖掘技术得到电力系统中存在的危险点并找出影响危险点的各种因素，从而指导决策者消除危险点。

通过安全性评价和辅助决策系统的共同使用，把电力系统的宏观评价和微观评价结合起来，对电力系统的安全稳定运行必定产生良好的影响。

1.2 系统的国内外发展现状和发展趋势

风险评估起源于20世纪30年代的美国保险行业，经过几十年的发展，形成了各种风险评估的理论、方法和应用技术。20世纪60年代，美国道化学公司开发了以火灾爆炸指数为依据的评价办法，推动了评价工作的发展[3]。自1997年以来，北京电机工程学会组织华北网50多名专家开展安全性评价查评工作，对华北、东北、华东、华中、西北、广东等发供电、调度单位进行150余次安全性评价工作，发现了近万个影响安全生产的隐患和问题，通过整改，减少了电力安全生产事故，提高了电力企业安全生产基础水平。国家电力公司在2000年5月1日新颁的《安全生产工作规定》中明确将安全性评价列为供电企业安全管理例行工作的范畴，并多次提及该项工作。随着我国电力工业改革的不断深入，厂网分开，安全性评价是知识经济产品，越来越被各独立发电公司重视和应用。安全性评价能摸清和夯实企业的安全基础，针对特大、重大设备损坏事故，频发性事故、恶性事故及人身伤亡事故隐患，以及安全基础管理方面存在的问题，从生

产设备、劳动安全作业环境、安全管理三个方面进行查评、诊断，为企业领导和工程技术人员提供一种生产管理、技术管理和安全管理的指导方法。

文献[4]讲述了华北电力集团公司认真贯彻了“安全第一，预防为主”的方针，从1990年开始，在发供电企业安全性评价标准制订、时间方法等方面进行了成功的探索。并相继编写了电力系统相关的安全性评价手册，如《供电企业安全性评价》、《水电厂安全性评价》、《火电厂安全性评价》、《发电厂并网运行安全性评价》、《输电网安全性评价》、《电网调度系统安全性评价》等国家标准，为电力行业全面开展安全性评价工作奠定了基础。文献[5]通过调研广西岩滩电厂安全性评价工作的建立与实践，从安全性评价体系的构成、目的、及实施等方面，论述在电力企业开展安全性评价工作的必要性及意义。文献[6]介绍了河南省电力公司安全性评价开展的情况，取得了良好的成效。文献[7]阐述了安全性评价的含义及发展概况，说明安全性评价工作是对电力企业安全生产行之有效的管理方法；指出了准确理解和把握安全性评价的动态管理特性，是安全性评价工作有效实施和管理优化的重要前提；分析了安全性评价的动态管理特性必然存在的原因；提出了电力企业在安全性评价工作实施过程中应实行以人为本的管理，持续完善和更新安全性评价工作的评价内容和标准，实行全过程闭环动态管理，注重推动安全性评价的优化与发展。

电力企业通过安全性评价的查评、整改和复查评，有力夯实了各电力企业的安全生产基础，落实了安全生产责任，健全了安全生产管理制度，规范了安全作业措施，促进企业真正构建长期有效的安全生产机制，保持电力安全生产稳定进行。自开展安全性评价工作以来，电力行业总体事故率呈下降趋势，一些可能引发重大事故的安全隐患得到了及时的发现并加以排除。电力企业形成了“自查、评价、整改、复评”一套完整的事故隐患排查，事故治理和事故预防控制的自我约束、自我发展的安全机制。

目前，安全性评价已经在电力系统中得到广泛应用，并取得了良好的效果。但还存在以下不足：

①查评的结果数据未得到充分的分析和管理。电力系统在一次查评结束后，只对数据进行了简单的处理，没有做深入的分析和研究。

②安全性评价未与“危险点分析”相结合[8]。安全性评价主要用于一个企业全方面宏观上的风险评估，危险点分析则是主要用于作业现场、生产岗位等局部微观上的评估，二者各有特点，互为补充，不可替代，应该结合进行。或者说安全性评估也应该对危险点分析工作进行评价。

文献[9]提出了“危险点”的观念，并对电力企业安全性评价改进做了深入研究，提出建立“危险点分析”制度。文献[10]分析了危险点预控的基本思想，讨论了危险点预

控的涵义和实施思路，并开发了供电企业危险点预控管理系统。文献[11]对火力发电厂安全性评价与危险点预控进行了分析，提出安全性评价不能只追求分数不注重效果，企业要正确对待安全性评价后所得出的分数。文献[12]从危险点分析与危险点控制的概念出发，概述了危险点的产生及其特点，并通过分析阐述了捕捉及控制危险点的方法，提出了危险点控制的基本原则及一般步骤，最后梳理了危险点控制工作中应注意的几个问题。

辅助决策系统作为一种新兴的信息技术，能够为企业提供各种决策信息以及许多商业问题的解决方案，从而减轻了管理者从事低层次信息处理和分析的负担。目前，辅助决策系统已经应用于电力系统负荷预测、电力市场金融风险等多个方面[13,14]。它利用数据仓库和数据挖掘技术对历史数据整理和分析，用来预测负荷、预测变电站设备缺陷等等。最近几年，辅助决策系统逐渐在安全性评价中提了出来。

电力系统安全性评价辅助决策系统致力于对电力系统进行微观评价，弥补了以前安全性评价只对电力系统进行宏观评价的不足。电力系统安全性评价辅助决策系统中运用了数据仓库和数据挖掘技术，安全性评价自查评数据通过整理放入到数据仓库，通过数据挖掘技术对数据仓库中的数据进行挖掘，通过数据挖掘技术得到电力系统中存在的危险点并找出影响危险点的各种因素，找出自查评项目存在的“危险点”，针对“危险点”，企业管理者可以加强此项目的管理，从而指导决策者消除危险点。通过辅助决策系统的使用，电力系统的宏观评价和微观评价结合起来对电力系统进行综合评价，它对电力系统的安全稳定运行必定产生良好的影响。

我们把安全评价称为前者，辅助决策系统称为后者，那么两者是相辅相成的，前者为后者提供历史数据，后者通过对历史数据的分析处理来指导前者的进行。

文献[15]通过“企业之星”业务平台中的非结构化数据建立数据仓库，并建立了数据仓库分析模块，为企业提供了统一的、多维度的数据查询和分析平台。文献[16]总结了供电网及其调度管理的特性，介绍了一个通用的供电网调度辅助决策专家系统外壳(PSDMAES)的构造和性能。文章还提出了解决专家系统和 SCADA 实时数据在线接口的一种新方法，从而使专家系统投入在线应用真正成为可能。文献[17]采用最新的 Microsoft SQL Server 2000 建立了电力调度数据仓库，提供浏览器和电子表格的联机分析工具，在电网线损分析和电力负荷方面进行了数据挖掘的初步尝试。文献[18]基于数据挖掘技术建立了电力负荷分析系统，首先利用粗糙集理论和遗传算法选取与负荷相关的预测变量，然后选取与预测日相似的训练模式，最后用神经网络对负荷进行预测，进而辅助调度员进行决策。文献[19]介绍了数据挖掘技术在电力系统中的应用情况，结合电力系统的自身特点，分析数据挖掘在电力系统当前的主要应用动态及应用前景。

随着科技的发展，安全性评价不仅仅局限于对电力系统的宏观评价，微观评价也逐渐被提出来。两者的结合将是电力系统安全性评价发展的必然趋势。

1.3 论文的主要工作

本文对电力系统安全性评价进行了较深入的研究，通过对电力系统安全性评价的现状分析，建立了基于数据挖掘的电力系统安全性评价辅助决策系统，并在此基础上建立了数据挖掘模型。利用数据挖掘技术挖掘电力系统中存在的危险点。利用挖掘结果来有效指导管理者对电力系统中存在的危险项目进行整改，从而消除安全隐患。电力系统安全性评价辅助决策系统是在电力系统安全性评价基础之上建立的辅助决策模块。针对不同的安全性评价体系，只需更改数据库路径，便可以嵌套在安全性评价中使用，具有一定的通用性。此模块包括数据仓库、数据挖掘和危险点汇总三个部分。论文主要工作如下：

（1）安全性评价研究

笔者对电力系统安全性评价做了深入的研究，包括安全性评价对象、目的、形式、意义等，将自己对安全性评价的理解做为论文单独的一章归纳出来。具体内容将在第 2 章中介绍。

（2）数据仓库的建立

安全性评价的数据通过数据预处理后放入到数据仓库，这些数据为以后的数据挖掘做准备。数据仓库中的数据可以以表格或者柱状图直观的展示给用户。具体内容将在第 4 章中介绍。

（3）数据挖掘

利用数据挖掘技术对数据仓库中的安全性评价数据进行深入挖掘，挖掘出存在危险点的项目。企业管理人员可以对存在危险点的项目提出预防措施，在今后的安全性评价工作中加大检查的力度。具体内容将在第 4 章中介绍。

（4）加权关联规则运用

提出传统关联规则的不足和局限性，并对其做了改进，即在辅助决策系统中引入了加权的概念，通过设置安全性评价查评项目中属性所占的比重，来决定项目所占的权值。具体内容将在第 5 章介绍。

1.4 本文的具体组织结构

本文共分为七章：

第一章为绪论。介绍了论文选题的背景、当前的国内外现状，并阐述了课题研究的必要性和可行性。

第二章主要对笔者研究的电力系统安全性评价做了详细的介绍，包括安全性评价的定义、目的和对象；安全性评价的主要过程、方法和意义；最后介绍了我国安全性评价的现状。

第三章介绍了系统需要运用的数据仓库及数据挖掘的基础知识，并重点对数据挖掘中的关联规则理论以及所用的 Apriori 算法做了详细的介绍。

第四章对系统的整体设计做了介绍，给出了系统的模型，对数据挖掘的过程作了详细的介绍。

第五章对辅助决策系统运用的算法进行了改进，在系统中引入了加权的概念，进一步完善了辅助决策系统模块。

第六章对整个系统的界面进行了设计，包括登录界面、数据仓库界面、数据挖掘界面等等，并对各个模块的功能进行了详细的说明

第七章为文章的总结和展望。总结了论文的主要工作，说明系统自身的特点，并对系统的进一步研究做了展望。

2 电力系统安全性评价介绍

2.1 概述

安全性评价在国外又称“风险评估”，起源于20世纪30年代的美国保险业，发展到今天，它已经与现代企业管理科学技术相结合，形成了许多关于安全性评价的理论和技巧[20]。其具体含义是：综合运用安全系统工程的方法对系统的安全性进行度量和预测，通过对系统存在的危险性进行定性和定量的分析，确认系统发生危险的的可能性及其严重程度，提出必要的措施寻求最低的事故率、最小的事故损失和最优的安全投资效益。它是现代化安全管理的一项重要内容。

安全性评价已经用于电力系统的多各方面，比如发电厂、供电企业、输电网等等。目前，电力系统安全性评价已经形成了各种国家标准供企业参考。

2.2 电力系统安全性评价的目的和对象

电力安全生产的方针是“安全第一、预防为主”。“预防为主”实质是要夯实安全基础，实现对事故的预先控制，是现代安全管理的基本原则。我国已用法律形式将“安全第一、预防为主”确定为劳动保护方针，也是电力安全生产和建设管理的基本方针。

电力系统安全性评价是一种能从多个方面对电力系统安全生产进行查评、诊断，为企业领导和工程技术人员提供一种生产管理、技术管理和安全管理的方法[21,22]。他能够摸清和夯实企业的安全基础，针对电网运行、设备工况、生产环境、作业过程等进行安全性评价，有利于辨识和治理事故隐患，实现对事故的超前预测和控制，以达到减少和消除事故的目的。

电力系统安全性评价的评价对象是一个庞大的人、物、环境和管理系统的统一体。电力系统现存的处于变化中的危险因素，不论是主观存在还是客观原因形成的都应该列入评价范围，其设计的评价内容、影响因素等都较为广泛。电力系统安全性评价的内容包括生产设备、劳动安全和作业环境、安全管理三个方面[23]。而所涉及的评价因素则包括以下8个方面：

（1）生产设备是否符合安全条件；

（2）主要生产工具、机具是否符合安全条件；

（3）部、局反事故技术措施的落实情况；

（4）生产设备、工机具的管理水平；

（5）生产、安全主要规章制度建立、健全和贯彻执行情况；

（6）人员技术素质是否符合安全要求；

（7）劳动环境是否符合安全要求；

（8）重大自然灾害抗灾、防灾措施的落实情况。

电力系统安全性评价是一项系统工程，评价的形式是企业自我查评、整改与专家评价相结合，形成由企业自查、整改、专家评价、再整改、复查、巩固等不同环节组成的电力企业自我约束、自我发展的安全机制。在实际的查评工作中，查评技术人员可采用查现场、查资料、实物抽查、现场调查和询问、实验测试等方法，根据现场实际情况灵活使用。电力企业本身自觉地开展安全性评价工作，必将不断夯实安全基础，超前控制事故的发生。

2.3 电力系统安全性评价方法

当前，国内外用于安全性评价的方法很多，按照评价结果的量化程度可以分为定性安全评价方法和定量安全评价方法[24]。

（1）定性安全评价方法

主要是根据经验和直观判断能力对生产系统的工艺、设备、设施、环境、人员和管理等方面的状况进行定性的分析，安全评价的结果是一些定性的指标，如是否达到了某项安全指标、事故类别和导致事故发生的原因等。属于定性安全评价方法的有安全检查表、专家现场询问观察法、因素图分析法、事故引发和发展分析、作业条件危险性评价法、故障类型和影响分析、危险可操作性研究等。

（2）定量安全评价方法

运用基于大量的实验结果和广泛的事故资料统计分析获得的数学模型，对生产系统的工艺、设备、设施、环境、人员和管理等方面的状况进行定量的计算，安全评价的结果是一些定量的指标，如事故发生的概率、事故的伤害范围、定量的危险性、事故致因因素的事故关联度或重要度等。

考虑到科学性和实用性，根据电力系统实际的安全性评价工作的需要，企业选用定量与定性相结合的逐项赋值评分法，简称“评分法”。这种方法首先根据评价对象和目的，按照“系统”分解（分割）的方法，编制安全检查表，确定查评项目，然后根据查评项目所涉及的危险因素的重要程度，逐项赋以同重要程度相对应的分值。

具体做法上一般都先确定一个总分，然后逐级确定子系统的重要程度，即确定权重系数，根据权重系数将总分分配到各个子系统，如果需要还可按此法将子系统的总分再分配到下一级子系统，然后再按权重系数将子系统（或下一级子系统）的总分分配到各个评价项目。权重系数的确定，以往都是通过专家讨论或通过对专家的调查和咨询进行的，主要是依靠专家群体的知识和经验。

2.4 电力系统安全性评价的具体过程

电力系统安全性评价是一项很复杂的工程，如何合理安排安全性评价的工作流程，是安全性评价工作能否达到预期目标的关键所在，如图 2.1 以供电企业安全性评价为例说明安全性评价在执行自查评和整改时的具体流程。

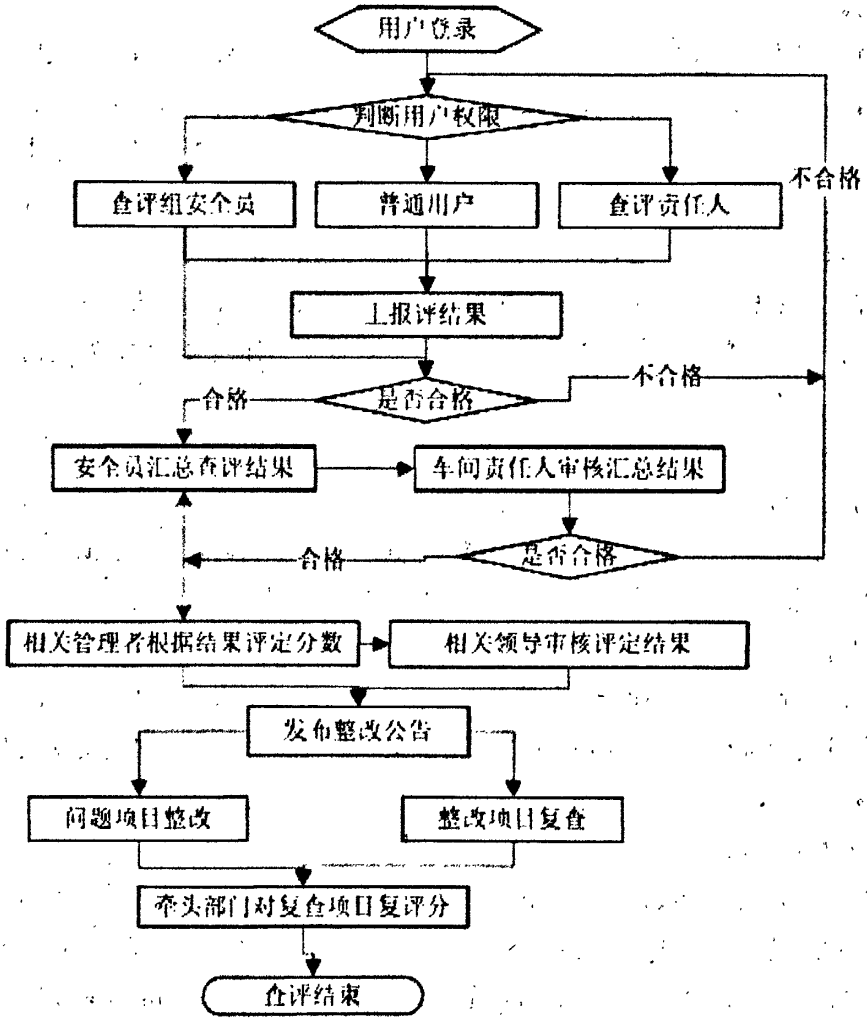


图 2.1 安全性评价流程

Fig.2.1 The process of safty evaluation

下面简单介绍一下供电企业安全性评价工作的主要过程：

（1）自查评阶段

①成立安全性评价查评组。为了让企业领导、技术人员、工人了解清楚安全性评价

工作的目的、意义和方法，为正确评价创造有利条件，企业应当首先成立一个临时的领导机构，负责安全性评价工作的宣传、组织学习和实施评价等。

②查评项目任务分解。安全性评价的项目应当按照专业项目和系统设备分解到各个车间、部门，然后再细分到各个班组，依次“层层分解”，最终将每个查评项目明确到个人。查评项目分解后，车间、部门应分别组织好安全性评价项目、查评方法等有关部分的学习。

③有计划地组织车间、班组进行“自查”。自查一般需要两周或多一些的时间，车间、班组检查中发现的问题，要登记在“安全性评价检查发现问题及整改措施”专用登记表上，经车间汇总报上级部门，查评专业组要在有关班组自查结果的基础上，进一步查看设备系统，查阅技术资料，评价过程中随时将评价结果或检查发现的问题记录在“查评扣分记录”表上。

④评价结果总结。查评组汇总评价结果，并提出安全性评价相关报告。安全性评价报告应包括：文字总结、安全性评价结果明细表、查评扣分记录和检查发现问题整改措施表等。

（2）整改阶段

①问题项目整改。若是单纯的对系统存在问题进行查询评价，那么评价就没有任何的意义，因此，安全性评价的目的在于整改。在明确了安全问题所在以后，企业将成立专门的整改组织体系，明确工作职责，具体落实整改部门、整改措施、整改责任人以及整改期限等。

②整改项目复查。企业将指定专人（一般为安监部负责人）定期检查整改措施落实情况，尽量做到每月检查一次项目的完成情况，统计完成率，并及时提交安全性评价查评组审核。

（3）复查阶段

通过一段时间（一般在一年左右）的项目整改以后，企业邀请相关专家组，对企业安全生产状态进行复查评，对复查评中仍存在的问题，企业将再次按照“自查—整改—复查”的程序进行评价工作，再次循环完成安全性评价工作，直到消除系统中存在的安全隐患。

2.5 电力系统安全性评价的意义

电力系统开展安全性评价工作，能够分析当前电力生产的安全状况，预测危险性发展为事故的概率以及事故造成损失的严重程度，全面地评价电力系统各部分的危险程度和安全管理状况，真正落实“安全第一，预防为主”方针，使安全管理变事后处理为事

先预测、预防，改变了事后补救处理的传统管理方式，具有重大意义。具体体现在以下几个方面[25,26]：

（1）安全性评价是为企业引进最先进的管理方法

加强和改进安全工作，需要借助现代安全管理方法。而安全性评价法作为一种现代安全管理方法，在事先预测系统存在的危险性，做出定性和定量的评价，并有针对性地提出控制措施上，有着明显的适用性、可靠性和有效性，是其他科学方法所不能替代的。进一步地说，安全评价法体现了安全管理方法的科学化，它采用了系统论等现代科学原理，全面系统地进行安全管理，能达到以较少的投入取得较佳的安全效果、预防或减少事故的发生目的。

（2）安全性评价工作落实了安全生产的标准

为了实现安全生产，国家对电力系统制定了安全生产方针，并相继颁发了一系列安全规章、规定和标准。这些安全生产方针、规章、规定和标准，是有效地遏制事故、保持系统安全的重要保证，也是开展安全性评价工作的基本依据和必须坚持的准则。企业开展安全性评价的过程，实际上就是贯彻落实安全生产方针、规章、规定和标准的过程。

（3）安全性评价使电力系统防患于未然

一般来说，系统中存在的危险分为两大类：一类是显现的，容易觉察和控制的；另一类是潜在的，不易觉察的。如果潜在的危险失去控制，就会造成事故的发生，因此，潜在的危险危害性更大。潜在的危险存在于系统的内部，人们不可能直接地进行观测。开展安全评价工作，通过对系统的全面分析、判断和评价，就能及时地发现系统存在哪些危险、处于哪些部位、危险的严重程度如何，从而采取措施加以控制，把危险降低到社会允许的标准。

长期以来，为查找和控制电力系统存在的危险、避免事故的发生，采取了多种措施。但这些措施与安全性评价相比较，都有一定的局限性。安全性评价则不然，它的着眼点放在分析和评价系统的安全基础上，能够通过量化方法，向管理人员提供哪些方面、哪些专业安全基础较为薄弱，薄弱到什么程度，哪些重大事故和恶性事故发生的可能性较大，存在的主要问题和隐患是什么以及需要采取什么样的方法加以防范等等。

（4）安全性评价增加了电力系统的可靠性

安全性评价工作从评价的对象、运用的手段到参评的人员，都体现了增加科技含量的特点，因而做出的评价会更加公正准确，适合现代化大生产的要求。开展安全性评价工作，对系统的安全性不仅做出定量分析，同时又做出定性分析，所获得的情况来自实践，企业安全管理者依据安全性评价结果做出的决策，就能实现主观与客观相一致，具有科学的指导价值。

随着社会对电力需求的快速增长，电网规模迅速扩展，技术的复杂性相应增加，客观上要求管理创新，需求探索并逐步建立与现代电力工业相适应的现代化管理体系。安全性评价势必对供电企业的安全生产管理产生深远的影响。

2.6 我国电力系统安全性评价的现状

电力生产过程采用安全性评价这一种现代化的管理方式，是实现电力安全生产管理创新并与国际安全生产管理先进水平接轨的重要体现[27,28]。我国国家电网公司从 90 年代末期开始研究如何对电力系统进行安全基础的评价工作，至今已经经历了十几年的历程。期间借鉴了国外的风险评估现代化的管理方法，总结了我们国家十几年来在安全生产上积累的宝贵经验，培养了一批业务素质精良的安全生产及监督管理队伍，形成了一套电力安全生产规章制度，建立了电力生产安全管理机制，设备基础和人员素质有了较大提高，电网安全生产局面稳定，安全生产指标稳步提高，人身安全、电网事故、设备事故率逐年下降，事故造成的损失逐年下降。

自开展安全性评价工作以来，我国电力系统连续几年电力事故和一类障碍呈下降趋势，可能引发重大事故的隐患及时得到发现和消除，企业的经济技术指标和事故率有了明显改观，安全生产水平得到显著提高，“安全性评价”开始转化为巨大的生产力。于此同时，电力系统也形成了“自查—整改—复查”的一套事故隐患辨识和预防控制的自我约束、自我发展的安全机制[29]。

但是，由于安全性评价的历史局限性以及电力系统本身的特点，目前的电力系统安全性评价工作还难免存在一些问题：

（1）对查评的国家标准理解有差异，评分中的随意性难以避免；

（2）评价工作依然为一次性行为，查评的结果数据未能得到充分的分析和管理，没有对数据进行完全的整理；

（3）安全性评价工作没有纳入日常安全管理体系，没能充分挖掘和利用安全性评价工作对于安全管理的巨大的价值。

由于以上这些问题的存在，电力系统安全性评价工作往往只停留在表面，用户没有意识到查评下来的安评历史数据对企业长期安全管理的作用和意义。

为了解决电力系统安全性评价工作中的诸多问题，建立一个基于数据仓库和数据挖掘技术的安全性评价辅助决策系统成了必然。系统利用数据仓库技术对日常的数据进行统一分析和管理，剔除错误和空白数据。利用数据挖掘技术对数据仓库中的数据进行分析，及时发现电力系统中存在的人们表面上难以察觉的问题，决策者可以针对这些问题提出重点预防以及改进措施。

实践证明全面推动、规范实施安全性评价工作是十分必要的，但是由于我国开展安全性评价工作的时间较晚，安全信息不够齐全，尽管目前我国电力系统安全性评价工作取得了一定成果，但其间仍然存在较多的不足之处，需要在以后的实践中不断的提高和完善，需要有一个逐步成熟的过程，尤其是在提高我国安全性评价管理工作的科学化、简单化、智能化方面任重而道远。

3 数据仓库、数据挖掘技术及其在电网安全性评价中的应用

数据仓库在技术上为数据挖掘提供了最基础、最理想的数据。数据仓库的数据是经过清理、转换、集成、标准化和初步分析处理了的，它为数据挖掘提供了干净的、一致的、有质量保证的、准确的数据，并提供了有助于对数据及信息进行准确理解的环境。因此，数据仓库为数据挖掘提供了原始的数据，为高效执行数据挖掘算法准备了条件、减少了障碍。总的来说，数据仓库是数据挖掘的基础，他为数据挖掘提供数据支持。数据挖掘技术对数据仓库中的数据做更深入的分析处理，挖掘出里面隐藏的、对人们感兴趣的知识或者模式，为用户辅助决策提供支持。

3.1 数据仓库

3.1.1 数据仓库定义及特征

社会的迅速发展导致了市场竞争的加剧，从大量数据中提取（查询、检索等）制定市场策略信息显得尤为重要。这种需求既要求联机服务，又涉及大量用于辅助决策的数据，然而传统的数据库系统已无法满足这种要求。具体体现在以下三个方面：

（1）历史数据量很大

一个企业存在很多管理系统，例如对于一个学校，它包括：办公自动化系统、邮件管理系统、教务系统等等。那么它一天的数据量是很大的，对于传统的数据库系统，很容易储存这些数据，但是无法实现决策功能。

（2）辅助决策信息涉及许多部门的数据，而不同系统的数据难以集成

数据库中储存着大量的历史数据，这些历史数据具有不同的特征和属性，不利于用户进行辅助决策。

（3）它对大量数据的访问能力不足

当数据库访问大量数据时，所耗时间相对较长，访问的数据不能直观的展现给用户。

C/S技术的成熟以及并行数据库的发展，提高了数据分析和决策的有效性和效率，把大量的事务型数据库中抽取数据，并将其清理、转换为新的存储格式是当今信息处理技术的发展趋势，即把数据聚合成一种特殊的格式以便用户进行决策。随着此过程的发展和完善，这种用于决策支持系统（DSS－Decision Support System）的、特殊的数据存储即被称为数据仓库[30,31]。

数据仓库之父 W.H.Inmon 认为“数据仓库是一个面向主题的、集成的、不可更新的且随时间不断变化的数据集合，用来支持管理人员的决策”。该定义表明了数据仓库是一个处理过程，该过程依据主题对若干个分布的、异质的信息源中的历史数据进行组织和存储，并能集成地进行数据分析，它有着比一般数据库系统更大的数据规模。数据仓库不仅具有传统数据库管理系统的共享性、完整性、数据独立性等基本特点，还具有主题、集成性、不可更新性、时间属性等独有特征。

（1）数据仓库是面向主题（Subject-Orientation）的

它是根据最终用户的观点来组织和提供数据的，其目的是尽快而全面的提供用户所需要的信息，很少或几乎不用做数据的更新操作。因此，数据仓库中数据的组织形式是按照各种主题的方式来进行的，主题在数据仓库中的物理实现是一系列相关的数据表，这与面向应用环境有很大的区别。比如一个物流公司按照应用组织可能是社会物流、企业物流、国际物流、区域物流、特殊物流，而数据仓库环境是按照客户、需求、质保金、索赔来组织数据。这样就决定了数据仓库将重点集中在数据建模和数据库设计上，而不像面向应用的环境还需要关心过程设计；数据仓库中细节的数据还摒弃了仅用于操作而对决策支持没有用处的数据。

（2）数据仓库是集成（Integration）的

由于要管理大量的历史数据，这些数据与多个领域或者多个数据库相关，需要收集和组织多个应用程序来处理在各种地方获取的数据。

（3）数据仓库是不可更新（non-renewal）的

数据仓库中的数据反映的是一段相当长的时间内的历史数据，主要供企业辅助决策分析之用。与面向应用的事务数据库相比，数据仓库中的数据所涉及的主要操作是查询和新数据的导入，很少进行修改操作。而不需要对数据作频繁的插入、更新操作。

（4）数据仓库是随时间不断变化（varies with time）的

针对数据仓库应用来说的它里面的数据是不可更新，即数据仓库的用户进行分析处理时不进行数据更新操作。但数据仓库并非永远不变的，数据仓库的数据是随时间不断变化的。数据变化有三种方式：

1. 数据仓库随时间变化不断增加新的数据内容。
2. 数据仓库随时间变化不断删去旧的数据内容。
3. 数据仓库要随时间变化不断对数据进行重新综合。

3.1.2 数据仓库的体系结构

数据仓库是多种技术的综合体，数据仓库的实施分数据获取、数据组织、数据应用和数据展示四个功能区。其体系结构如图 3.1 所示：

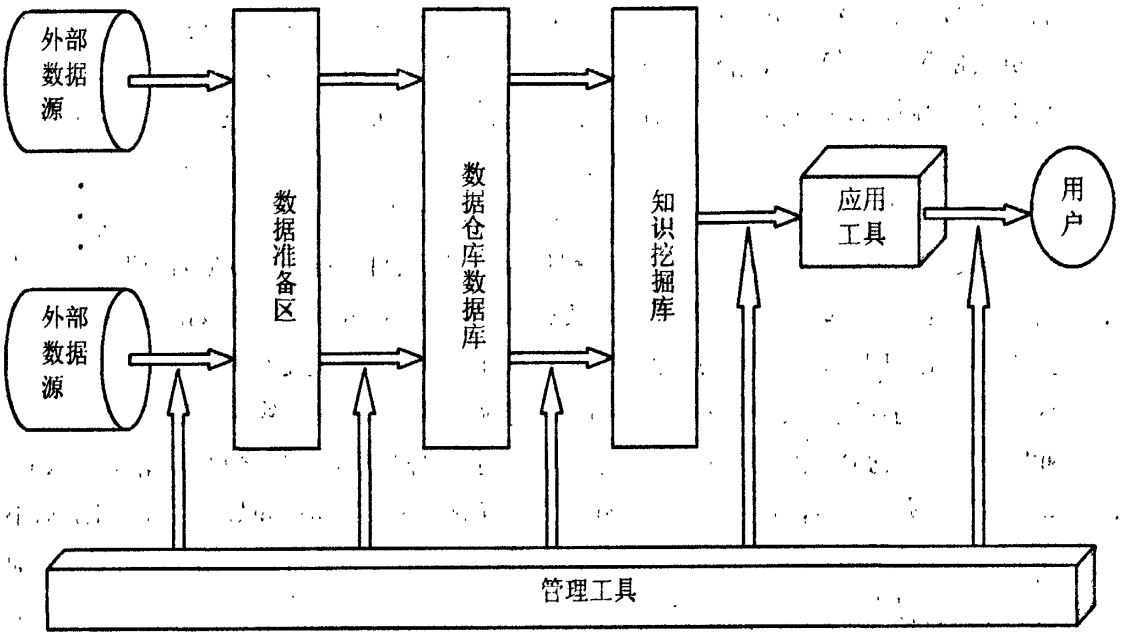


图 3.1　数据仓库体系结构

Fig.3.1　Data Warehouse Architecture

从图 3.1 可以看出，可以利用各种数据仓库管理工具辅助完成数据仓库的创建。数据仓库建立过程如下：首先要从“数据源”中提取相关的数据到“数据准备区”；数据在“数据准备区”中经过一系列（包括清理、集成、净化）处理后，再加载到“数据仓库数据库”；最后根据用户的需求将数据导入到“知识挖掘库”中。用户在使用数据仓库时，主要利用两类应用工具：一类是联机分析处理（On-Line Analvtical Processing），主要用于分析历史发展变化；另一类是数据挖掘（Date Mining），主要用于预测未来趋势走向。

3.2　数据挖掘

3.2.1　数据挖掘定义

什么是数据挖掘（Date Mining）？关于定义取决于定义者的观点和背景，各人的说法不一样[32,33]。

Fayyad 提出数据挖掘是一个确定数据中有效的、新颖的、潜在有用的，以及最终可理解的模式的非平凡过程。

Zekulin 认为数据挖掘是一个从大型数据库中提取以前未知的、可理解的、可执行的信息，并用它来进行关键的商业决策的过程。

Ferruzza 给出数据挖掘是用在知识发现过程，来辨识存在于数据中的未知关系和模式的一些方法。

Jonn 提到数据挖掘是发现数据中有益模式的过程。

Parsaye 定义数据挖掘是我们为那些未知的信息模式而研究大型数据集的一个决策支持过程。

从以上的表述可以看出，每个人对数据挖掘的理解是不同的。这里我们主要从技术和商业的角度给出数据挖掘的定义。

从技术的角度看，数据挖掘（Date Mining）就是从大量的、不规则的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程[34-36]。这个定义包括好几层含义：数据源必须是真实的、大量的、含噪声的；发现的是用户感兴趣的知识；发现的知识要是可接受、可理解、可运用的；这些知识是相对的，是有特定前提和约束条件的，在特定领域中具有实际应用价值。

从商业角度看，数据挖掘是一种新的商业信息处理技术。其主要特点是对商业数据库中大量业务数据进行提取、转换、分析和其他模型化处理，从中提取辅助商业决策的关键性知识，即从一个数据库中发现相关商业模式。例如对于一个商场，通过数据挖掘技术可以分析出乒乓球拍销售量对乒乓球销售量的影响。多年以来，统计学家开始手工挖掘数据库，从数据库中寻找符合统计学规律的有意义的模式或规则。这是统计学类型的数据挖掘技术，是目前数据挖掘技术中最为成熟的技术。

综上所述，数据挖掘是一门交叉学科，它把人们对数据的应用从低层次的简单查询，提升到从数据中挖掘知识，提供辅助决策。在这种需求的引导下，汇聚了不同领域的研究者，尤其是数据库技术、人工智能技术、数理统计、可视化技术、并行计算等方面的学者和工程技术人员，投身到数据挖掘这一新兴的研究领域，形成新的技术热点。

3.2.2 数据挖掘中的关联规则理论

关联规则是形如下面的一种规则："在购买面包和啤酒的顾客中，有80 %的人同时也买了牛奶"（面包+啤酒=>牛奶）。关联规则挖掘是数据挖掘中的一种重要的数据挖掘模式，最初用于挖掘大型事务数据库中项与项之间的关系，近年来已经被广泛用于文本信息处理中[37]。一般来说，关联规则就是描述数据库中数据项（属性、变量）之间所存在的潜在关系的规则。设$I=\{i_1,i_2,\ldots,i_m\}$是m个不同项目的集合，D是针对I事物的集合，

每一笔事物包含若干项目$i_1,i_2,\ldots,i_k \in I$。关联规则形如X=>Y的蕴含式，其中$X \subset I$，$Y \subset I$，$X \cap Y=\Phi$。对于规则“X=>Y”，一般用支持度、置信度、期望置信度和作用度四个参数来描述一个关联规则的属性。下面将进行详细介绍。

（1）支持度

支持度描述了 X 和 Y 这两个属性集在所有的事务数 D 中同时出现的概率有多大，即

$$\sup port(X => Y) = \frac{\sup port\_count(X \cap Y)}{\sup port\_count(D)} \tag{3.1}$$

其中 support 表示支持度，support_count 表示事物的支持度计数。

例如：某天共有 100 个顾客到某超市购买商品，其中有 20 个顾客同时买了面包与啤酒，那么关联规则“面包=>啤酒”的支持度为 20%（20/100）。

支持度是对关联规则重要性的衡量，支持度说明了这条规则在所有记录中有多大的代表性，显然支持度越大，该关联规则所代表的情况出现的越频繁。

（2）置信度

规则 X=>Y 的置信度是事物集中 X、Y 同时出现的事物数与 X 出现的事物数之比，即

$$confidence(X => Y) = \frac{\sup port\_count(X \cap Y)}{\sup port\_count(X)} \tag{3.2}$$

如上所述的面包与啤酒的例子，该参数表述了这样一个问题，如果一个用户买了面包则他同时买啤酒的可能性有多大。若有 60%的用户在买了面包的同时买了啤酒，关联规则“面包=>啤酒”的置信度即为 60%。

置信度是对关联规则的准确度的衡量，有些关联规则置信度很高（例如达到 80%以上），即使其支持度较低，所代表的情况出现的次数有限，但是只要抓住这些规则，成功的可能性就比较高。

（3）期望置信度

X=>Y 期望置信度描述了在没有任何条件影响下，属性集 Y 在总的事务数中出现的概率，即

$$P(Y) = \frac{\sup port\_count(Y)}{\sup port\_count(D)} \tag{3.3}$$

仍然考虑面包与啤酒的例子，如果某一天共有 100 个顾客到超市购买物品，其中有 30 个顾客买了啤酒，则上述关联规则的期望置信度为 30%(30/100)。

（4）作用度

X=>Y 的作用度是置信度和期望置信度的比值，即

$$Lift(X \Rightarrow Y) = \frac{P(Y \mid X)}{P(Y)} = \frac{P(XY)}{P(X)P(Y)} \tag{3.4}$$

则在面包与啤酒的例子中，关联规则“面包=>啤酒”的作用度为 2。

作用度描述属性集 X 的出现对属性集 Y 的出现的影响有多大。置信度与期望置信度的比值反映了在加入“属性集 X 出现”的这个条件后，属性集 Y 的出现的概率发生了多大的变化。由公式 3.4 可知，显然作用度越大，说明属性集 Y 受属性集 X 的影响越大。一般情况，有用的关联规则的作用度都应该大于 1，只有关联规则的置信度大于期望置信度，才说明 X 的出现对 Y 的出现有促进作用，也说明了它们之间某种程度的相关性，如果作用度不大于 1，则此关联规则属于误导或者说错误的规则，也就没有意义了。

为了直观、简便的引入概率公式，以上四个参数可用公式表示如表 3.1。

表 3.1 关联规则参数

Tab.3.1 Parameters of association rules

名称	计算公式	概率表示	意义
支持度 (Support)	$\frac{A和B的记录总数}{记录总数}$	P(A∩B)	属性集 A、B 同时出现的概率
置信度 (Confidence)	$\frac{A和B的记录总数}{包含A的记录总数}$	P(B|A)	在属性集 A 初出现的前提下，属性集 B 出现的概率
期望置信度 (Expected Confdence)	$\frac{包含B的记录总数}{记录总数}$	P(B)	属性集 B 出现的概率
作用度 (Lift)	$\frac{置信度}{期望置信度}$	P(B|A)/P(B)	置信度与期望置信度的比值

从关联规则定义可知，任意给出一个事务的两个属性集，它们之间或多或少都存在某种关联规则，只不过各参数值（支持度、置信度等等）有所不同。如果不考虑关联规则的支持度、置信度与作用度，那么就可能形成关联规则的泛滥，即出现很多有用或者无用的关联规则。然而，事实上人们对满足一定支持度和可置信度，并且作用度大于 1 的关联规则感兴趣。因此为了找出对用户有意义的关联规则，需要给定两个阈值：最小支持度和最小置信度。前者用来描述关联规则的频繁性，如果规则的支持度大于最小支持度则认为此规则是频繁项集，否则为非频繁项集。后者用来描述规则的可信度。同时满足最小支持度与最小置信度的规则称为强关联规则。那我们给出如下定义：

定义 1 强关联规则，是指规则满足用户给定的最小支持度（min_sup）和最小置信度（min_conf）的关联规则。

定义 2 频繁项目集，是大于等于最小支持度的非空子集。

定义 3 最大频繁项目集，是频繁项目集中不被其它元素包含的频繁项目集。

关联规则挖掘的目的就是从数据库中挖掘出满足用户要求的最小支持度与最小置信度的强关联规则。

挖掘关联规则问题一般可以分解为以下两个子问题[38]：

①找出存在于事物数据库中的所有频繁项集，即找出所有支持度满足用户所规定的最小支持度阈值的项集；

②用频繁项集生成候选关联规则，然后验证候选关联规则是否满足用户所规定的最小置信度阈值。若满足，该候选关联规则为要找的关联规则。

3.2.3 关联规则中的 Apriori 算法

关联规则常用算法有很多，这里主要介绍 Apriori 算法。

关联规则在分类上属于单维、单层、布尔型关联规则。Apriori算法是一种最有影响的挖掘布尔关联规则频繁项集的算法。其核心是基于两个阶段频集思想的递推算法。在这里，所有支持度大于最小支持度的项集称为频繁项集，简称频集。该算法的基本思想是：首先找出所有的频繁项集，这些项集出现的频繁性至少和预先设定的最小支持度一样或者更大。然后由频繁项集产生强关联规则，这些规则必须满足最小支持度和最小置信度。一旦这些规则被生成，那么只有那些大于用户给定的最小置信度度的规则才被留下来。为了生成所有频集，使用了递推的方法。Apriori算法使用一种称作逐层搜索的递推方法，用k项集搜索（k+1）项集。首先通过扫描数据库，累积每个项的计数，并收集满足最小支持度的项，找出频繁 1 项集的集合L_1，然后L_1用于找频繁 2 项集的集合L_2，L_2用于找L_3，如此下去，直到不能再找到频繁 k 项集。找每个L_k需要一次数据库扫描。Apriori算法的关键是由L_{k-1}找L_k：先通过L_{k-1}与自身连接产生候选k项集的集合C_k（即连接步），并利用Apriori性质压缩C_k（即剪枝步），然后扫描事物表进行计数，将满足最小支持度的项加入项加入L_k中。然后根据最大频繁项目集和用户给定的最小置信度产生强关联规则。

综上所述，Apriori 算法的设计可以分解为两步骤来执行挖掘：

1、从事务数据库D中挖掘出所有频繁项集。

首先需要挖掘出频繁 1-项集；

然后，继续采用递推的方式来挖掘频繁 k-项集（k>1），具体做法是：

在挖掘出候选频繁k-项集（C_k）之后，根据最小置信度min_sup来筛选，得到频繁k-项集。

最后合并全部的频繁 k-项集（k>0）。

挖掘频繁项集的算法描述如下：

```
L1 = find_frequent_1-itemsets(D); //挖掘频繁 1-项集
for (k=2; Lk-1 ≠Φ; k++)
  {
  Ck = apriori_gen(Lk-1 ,min_sup); //调用apriori_gen方法生成候选频繁k-项集
  for each transaction t ∈D
    {
      Ct = subset(Ck, t);
      for each candidate c∈Ct
      c.count++; //统计候选频繁 k-项集的计数
    }
    Lk ={c ∈Ck|c.count≥min_sup}//满足最小支持度的k-项集即为频繁k-项集
  }
return L= ∪ k Lk; //合并频繁k-项集（k>0）
```

2、基于第 1 步挖掘到的频繁项集，继续挖掘出全部的频繁关联规则。

置信度大于最小置信度 min_conf 的关联规则称为频繁关联规则(Frequent Association Rule)。在这一步，首先需要从频繁项集入手，首先挖掘出全部的关联规则(或者称候选关联规则)，然后根据 min_conf 来得到频繁关联规则。

挖掘频繁关联规则的算法描述如下：

```
初始状态：L = ∪ k Lk; AR = Φ; //L 是频繁项集集合，AR 是频繁关联规则集合
for all λk     //λk 是 L 的元素，是一个 k-频繁项集
 {
   for all αk      //αk 是 λk 的非空真子集
    {
      if(conf(αk → βm)>= min_conf)      //这里，m + k = n，其中 αk →βm 是一个关联规则
        {
          AR = AR ∪ (αk → βm);
        }
    }
```

}

return AR;

此时，我们已经找出来所有的频繁关联规则。

3.3 数据挖掘技术在电力系统中的应用

数据挖掘技术在电力系统的应用中取得了大量的应用成果，主要集中在电力系统故障诊断、负荷预测、用户特征提取等方面。

文献[39]介绍了火电厂常见故障的数据挖掘方法，从火电厂 SCADA 系统历史数据库的大量实时数据中获取故障诊断知识进行故障诊断。该方法避免了为大型设备的故障诊断而附加的专门测试或试验,不但降低了费用, 而且减少了试验对设备造成的潜在的危险。文献[40]结合区域电网气象负荷数据库，设计与实现了决策树形式的数据挖掘模型 并运用于日负荷预测。统计分析结果表明该数据挖掘模型满足实用标准，具有通用可靠以及准确率高等特性。文献[41]把先进的数据挖掘技术应用到变电站设备缺陷预测管理过程中，首先对数据进行预处理，然后建立了多元线性回归模型，最后利用数据挖掘技术从中发现隐含的趋势和规律，从而指导变电站工作人员提前做好缺陷处理工作。文献[42]基于长期运行的电力安全生产事故统计分析系统积累了大量的事实数据，引入了一种基于数据挖掘的关联规则算法来挖掘出电力事故中的相关规则，进而通过这些规则来研究事故发生的规律。

数据挖掘技术在安全性评价中的应用还处于初级阶段，文献[43]最先提出了在安全性评价基础上建立辅助决策系统，结合实际的供电企业安全性评价管理工作的需要，利用.net 服务技术、数据库技术，设计出一套供电企业安全性评价管理（SAM）系统，构建了专家辅助决策数据仓库，运用 OLAP 和数据挖掘技术为安评工作管理人员提供有效的决策信息。但是此文献并未给出具体的算法和模型来处理数据仓库中的数据。接下来的文献并没有对数据挖掘技术在安全性评价中的应用作过多的研究，因此为了更好的达到安全性评价的目的，能深层次的挖掘安全性评价数据，建立基于数据挖掘技术的辅助决策系统成为了必然趋势。

4 电力系统安全性评价辅助决策系统的设计

4.1 系统实现目标

电力系统安全性评价辅助决策系统是在电力系统安全性评价基础之上进行的辅助决策。电力系统安全性评价致力于电力系统的宏观评价，辅助决策系统致力于电力系统的微观评价，它对电力系统安全性评价遗留下的历史数据进行分析、整理，运用数据挖掘技术挖掘电力系统中存在的危险点，利用 Apriori 算法从中找出影响电力系统危险点的关联规则。管理者针对危险点提出预防措施，从而消除电力系统中存在的危险点。通过关联规则的查找，了解影响危险点的各种因素，对管理者决策提供重要的依据。

4.2 系统的总体设计

4.2.1 数据准备

数据准备一般可分为三个步骤：数据源选择、数据预处理和数据变换。数据源选择的目的是确定数据挖掘的对象，即目标数据。数据预处理一般包括消除噪声、数据缺失处理、重复数据处理、数据类型转换等。数据变换的主要目的是消减数据维数或降维，即从初始特征中找出真正有用的特征以减少数据开采时要考虑的特征或变量数。

（1）数据源选择

数据来源于供电企业安全性评价管理系统，此系统现应用于重庆市璧山县电力公司。璧山电力公司每年都会进行安全性评价，因此数据量非常充分，故选取其中的安评数据作为数据源。如图 4.1 所示。

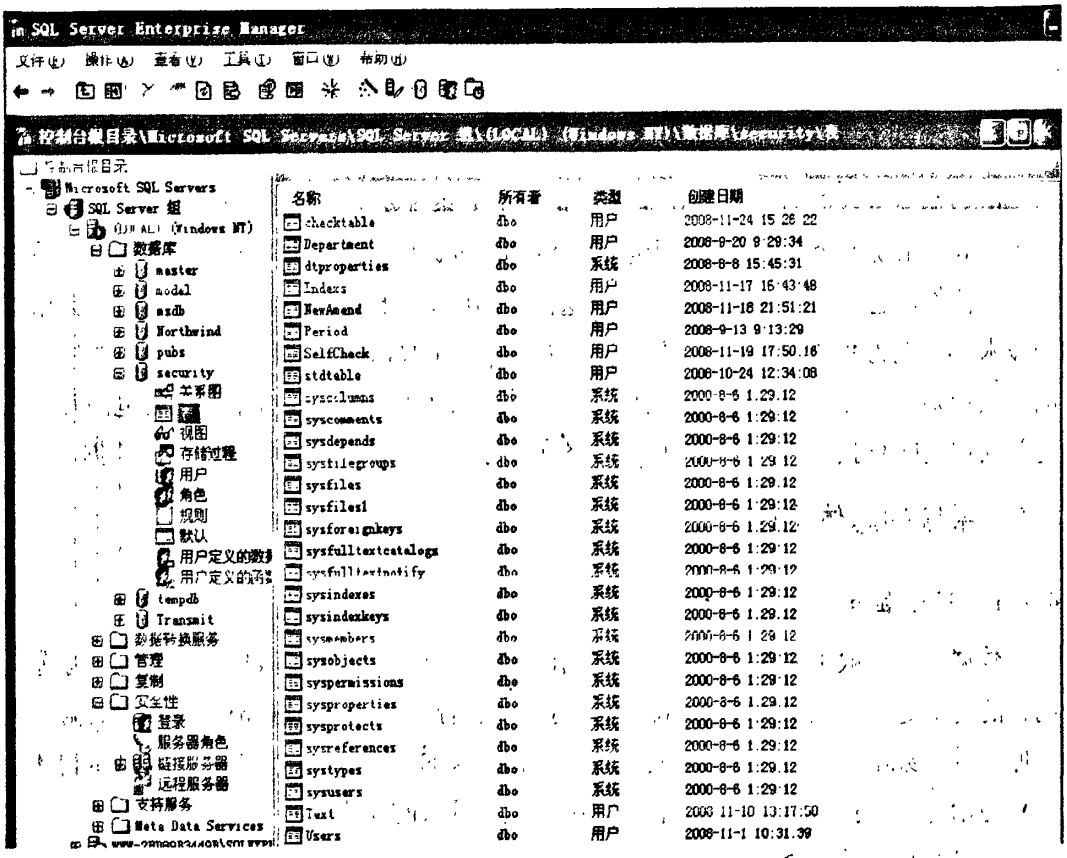


图 4.1 数据源

Fig.4.1 Data Source

如图 4.1 所示，SelfCheck 表示自查评表，其中保存所有用户对项目查评时产生的数据，NewAmend 表示整改表，所有项目的整改信息以及整改记录都保存在里面，Users 表示用户表，用来登录是核对用户信息的。

（2）数据采集、清理[44-47]

数据采集过程往往没有考虑到数据挖掘的目标，因此通过数据采集方法收集到的很多数据对挖掘目标来说是没有意义的；另外，我们也不知道具体的哪些数据对挖掘目标是有用的，为了不遗漏有用的数据，我们只能把所有采集到的数据都收集上来。为了解决以上矛盾，必须进行一定的数据清理。在数据预处理中，工作量最大的就是数据的清理和转换工作。我们此处清理的是自查评表中的数据。首先，对表中的字段进行解释，如图 4.2 所示。

28194	1.1	2	2	4	87	1500	1325	0.88	<NULL>	多项不合格	正常
28195	1.1.1	2	2	2	87	300	156	0.52	<NULL>	进行1空改	严重
28196	1.1.1.1	4	2	5	87	185	87	0.53	<NULL>	问题一般	正常
28197	1.1.1.1.1	4	2	5	87	35	25	0.7	<NULL>	需改	正常
28198	1.1.1.1.2	4	2	3	87	10	8	0.8	<NULL>	问题一般	正常
28199	1.1.1.1.3	3	3	7	87	20	5	0.25	<NULL>	领导视察	严重
28200	1.1.1.1.4	3	3	7	87	15	5	0.33	<NULL>	领导审核	<NULL>
28201	1.1.1.1.5	5	4	2	87	15	0	0	<NULL>	无二下篇	严重
28202	1.1.1.1.6	2	2	7	87	10	12	1.2	<NULL>	<NULL>	正常
28203	1.1.1.1.7	8	5	3	87	15	-10	-0.67	<NULL>	查阅相关项目	严重
28204	1.1.1.1.8	5	2	1	87	20	30	1.5	<NULL>	个问题	正常
28205	1.1.1.1.9	7	5	8	87	15	10	0.67	<NULL>	<NULL>	<NULL>
28206	1.1.1.1.10	2	3	6	87	10	10	1	<NULL>	无需修改	正常
28207	1.1.1.2	4	8	5	87	85	50	0.68	<NULL>	<NULL>	<NULL>
28208	1.1.1.2.1	6	5	7	87	15	5	0.3	<NULL>	查阅图标	正常
28209	1.1.1.2.2	6	5	7	87	15	12	0.8	<NULL>	符合条件	正常
28210	1.1.1.2.3	3	3	5	87	10	2	0.2	<NULL>	领导审查	严重
28211	1.1.1.2.4	8	5	3	87	10	3	0.3	<NULL>	不合格	<NULL>
28212	1.1.1.2.5	5	8	7	87	10	15	1.5	<NULL>	<NULL>	正常
28213	1.1.1.2.6	4	6	7	87	10	9	0.9	<NULL>	合格	正常
28214	1.1.1.2.7	8	5	9	87	10	9	0.9	<NULL>	合格	正常
28215	1.1.1.2.8	8	5	9	87	5	3	0.6	<NULL>	问题一般	严重
28216	1.1.1.3	7	5	2	87	50	11	0.22	<NULL>	<NULL>	<NULL>
28217	1.1.1.3.1	7	5	3	87	10	13	1.3	<NULL>	不合格	正常
28218	1.1.1.3.2	7	5	3	87	10	-8	-0.8	<NULL>	合格	严重
28219	1.1.1.3.3	8	8	9	87	10	4	0.4	<NULL>	有点严重	严重
28220	1.1.1.3.4	5	4	3	87	10	2	0.2	<NULL>	整改	正常
28221	1.1.1.3.5	2	2	3	87	10	10	1	<NULL>	无需修改	正常
28222	1.1.2	<NULL>	<NULL>	<NULL>	87	520	<NULL>	<NULL>	<NULL>		<NULL>

图 4.2　自查评表

Fig.4.2　The form of self-evaluation

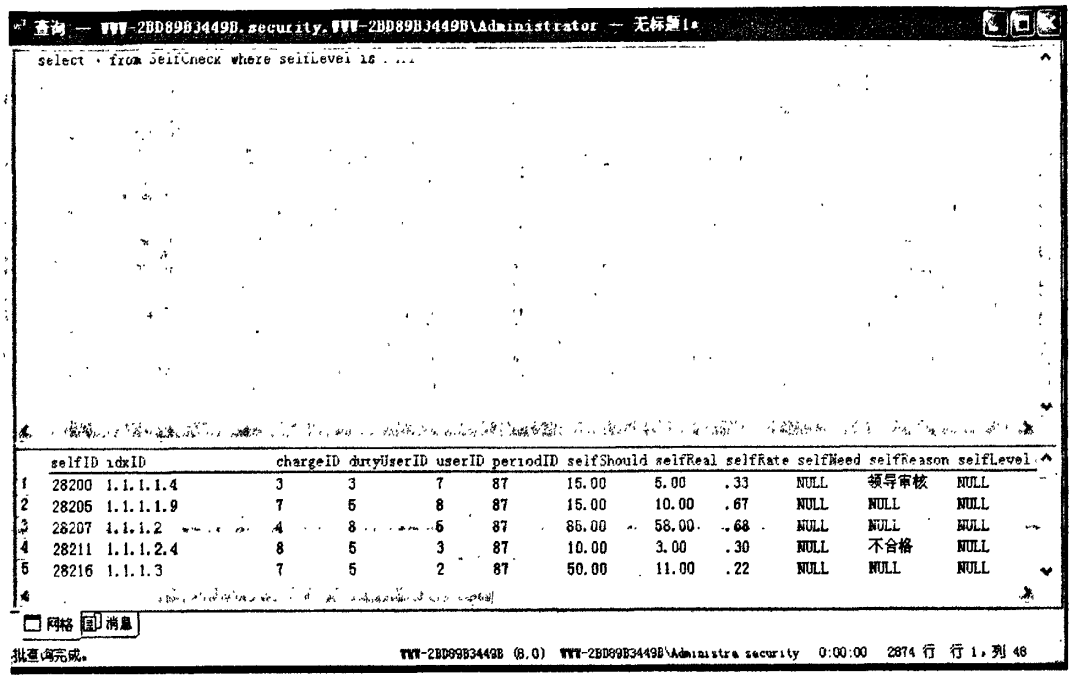
在图 4.2 中，这些数据来自自查评表（SelfCheck）中，idxID 代表查评项目；chargeID 代表部门编号；dutyUserID 代表审核人编号；userID 代表查评人编号；selfShould 代表项目标准分；selfReal 代表项目实际得分，selfRate 代表得分率；selfReason 代表得扣分原因；selfLevel 代表问题严重程度。

从图 4.2 中我们可以初步看出一些数据并不合理，有些项目的字段为空值，有些字段存在错误数据。例如项目 1.1.1.1.7 的得分率为-0.67，这显然不符合实际情况，因此我们对自查评的数据做如下处理：

①数据缺失的处理

数据缺失的原因是用户自查评时疏忽大意忘记填写。下面以 selfLevel 为例进行说明。通过如下语句来查询 selfLevel 的缺失情况：

select * from SelfCheck where selfLevel is null，其中 SelfCheck 表示自查评表，selfLevel 表示问题的严重程度。结果如图 4.3 所示。



select * from SelfCheck where selfLevel is ...

	selfID	idxID	chargeID	dutyUserID	userID	periodID	selfShould	selfReal	selfRate	selfNeed	selfReason	selfLevel
1	28200	1.1.1.1.4	3	3	7	87	15.00	5.00	.33	NULL	领导审核	NULL
2	28205	1.1.1.1.9	7	5	8	87	15.00	10.00	.67	NULL	NULL	NULL
3	28207	1.1.1.2	4	8	5	87	85.00	58.00	.68	NULL	NULL	NULL
4	28211	1.1.1.2.4	8	5	3	87	10.00	3.00	.30	NULL	不合格	NULL
5	28216	1.1.1.3	7	5	2	87	50.00	11.00	.22	NULL	NULL	NULL

图 4.3　缺失数据

Fig.4.3　Missing data

从图 4.3 中我们可以看出，selfLevel 为空值的所有项目均已找出，那么我们就要对这些空值进行挖掘前的处理。我们都知道，开展数据挖掘的前提是需要大量的历史数据，为了为以后的数据挖掘提供充分的数据，因此对于此处的缺失数据，笔者默认为 selfLevel 空缺值的属性为严重。

②数据错误的处理

数据错误产生的原因主要有两个：设计数据时没有进行约束；数据的人为输入错误。前者是在系统设计时没有对用户的输入进行约束，使得用户可以输入不满足要求的数据。后者是由于很多数据都是以字符串的形式来存储的，无法使用约束来保证数据的正确性，而且由于用户知识水平和文化背景的差异，输入时往往会拼写出错或者录入完全错误的数据。下面以 selfRate 为例进行说明，通过下面的语句来查询 selfRate 的错误情况：

select * from SelfCheck where selfRate >1 or selfRate<0，其中 SelfCheck 表示自查评表，selfRate 表示得分率。结果如图 4.4 所示。

查询 — WWW-2BD89B3449B.security.WWW-2BD89B3449B\Administrator — 无标题1*

```
select * from selfcheck where selfRate > 1 or selfRate 0
```

	selfID	idxID	chargeID	dutyUserID	userID	periodID	selfShould	selfReal	selfRate	selfNeed	selfReason	selfLevel
1	28202	1.1.1.1.6	2	2	7	87	10.00	12.00	1.20	NULL	NULL	正常
2	28203	1.1.1.1.7	8	5	3	87	15.00	-10.00	-.67	NULL	查阅相关项目	严重
3	28204	1.1.1.1.8	5	2	1	87	20.00	30.00	1.50	NULL	小问题	正常
4	28212	1.1.1.2.5	5	8	7	87	10.00	15.00	1.50	NULL	NULL	正常
5	28217	1.1.1.3.1	7	5	3	87	10.00	13.00	1.30	NULL	不审核	正常

网格 消息

批查询完成。 WWW-2BD89B3449B (8.0) WWW-2BD89B3449B\Administra security 0:00:00 6 行 行 1，列 56

图 4.4 错误数据

Fig.4.4 Error data

从图 4.4 中可以看出，selfRate 大于 1 和小于 0 的项目都已经找出。查询出 selfRate 产生的原因是用户自查评打分时，打分出现溢出或者输入错误。因为出现这种情况毕竟是少数，笔者将出错的 selfRate 属性全设置为 0。这样设置，我们就能保证不遗漏任何严重的问题。

（3）数据转换

由于selfRate的取值范围为[0,1]，不利于我们进行数据挖掘，那么我们就要对它进行离散化处理。将selfRate的取值范围为[0,0.5]记为A_1，selfRate的取值范围为(0.5,1]记为A_2。同样，对selfLevel也进行离散化处理，selfLevel属性值为“正常”记为B_1，属性值为“严重”记为B_2。离散化的结果可用表 4.1 表示。

表 4.1 数据转换

Tab.4.1 Data Conversion

tid	selfRate	selfLevel
1	A_1	B_1
2	A_1	B_2
3	A_2	B_2
4	A_1	B_2
5	A_2	B_1
…	…	…

我们将转换好的数据放入到数据仓库中，为后面的数据挖掘做准备。

4.2.2 数据挖掘模型的建立

我们要处理的问题是如何从数据源中挖掘到我们想要的危险点。那么我们建立了如下的挖掘模型：

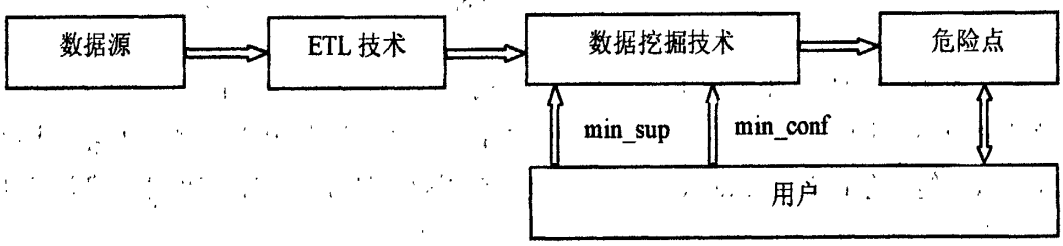


图 4.5 数据挖掘模型

Fig.4.5 Data mining model

模型解释：数据源我们选择是供电企业安全性评价管理系统中的数据；ETL 技术指的是对数据源中数据的清理、转换等；数据挖掘技术用的是关联规则挖掘方法；应用关联规则挖掘方法时用户需对算法中的参数进行设置，其中 min_sup 代表最小支持度，min_conf 代表最小置信度。通过数据挖掘技术挖掘出的危险点，可以直观的展现给决策者。

4.2.3 数据挖掘过程

本文已经将历史数据离散化处理，故主要是挖掘布尔型关联规则，因此采用最典型的 Apriori 算法。

首先采用 Apriori 算法生成频繁项集，然后由频繁项集根据最小支持度和最小置信度产生强规则。基于 Apriori 算法的数据挖掘流程如图 4.5 所示。

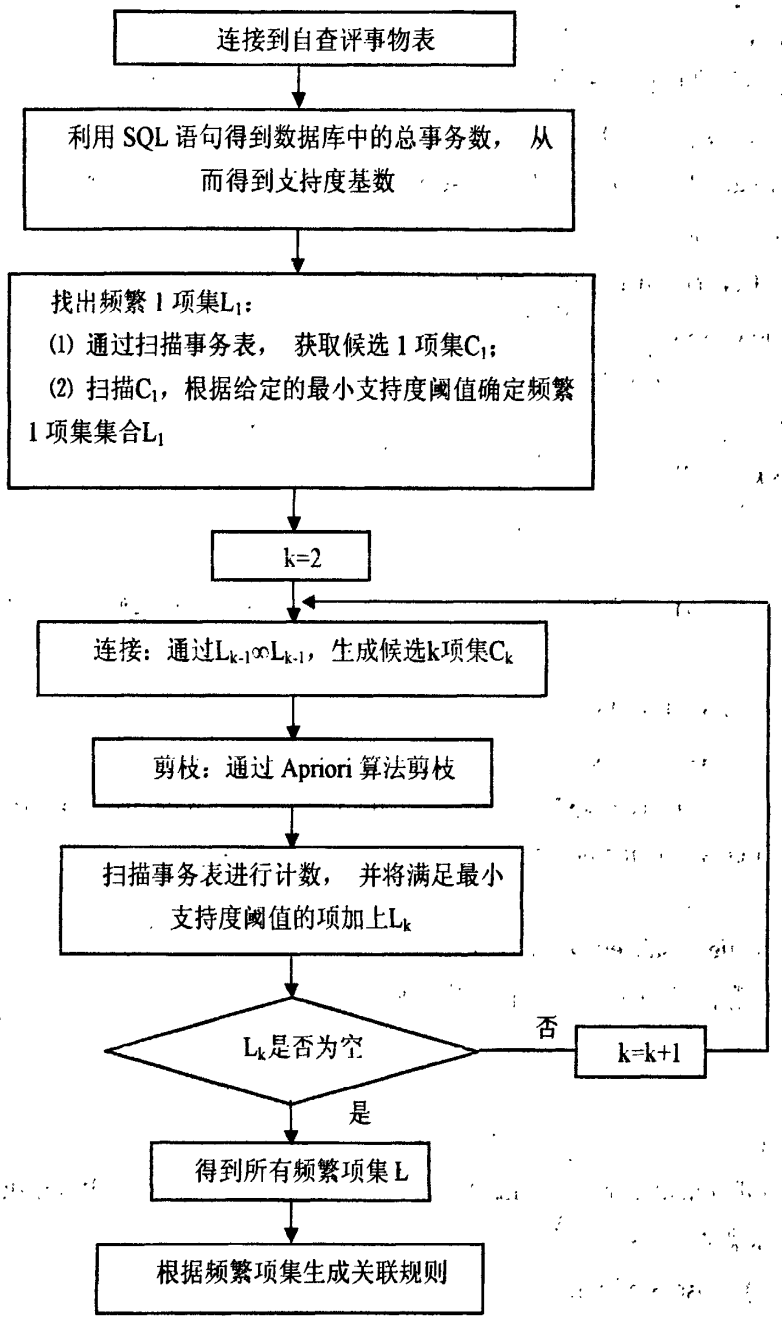


图 4.5 数据挖掘流程

Fig.4.5 The process of data mining

下面给出算法的伪代码：

输入：数据库D，最小支持度min_sup；

输出：频繁项集 L；

size=Scan(D)

L_1={频繁1-项目集}；//所有支持度不小于最小支持度的1-项目集；

for (k=2；$L_{k-1}\neq\varphi$；k++) do begin

C_k=apriori_gen(L_{k-1})；// C_k是k个元素的候选集

for all transactions t∈D do begin

C_t=subset(C_k，t)；//C_k是所有t包含的候选元素集

for all candidates c∈C_t do

c.count++

end

L_k={c∈C_k | c.count >=min_sup}；

end

L=∪L_k；

函数apriori_gen(L_{k-1})的作用是通过(k-1)-频繁项目集L_{k-1}产生k-候选项集C_k，具体描述如下：

for all itemset p∈L_{k-1} do

for all itemset q∈L_{k-1} do

if p.$itemset_1$= q.$itemset_1$ and p.$itemset_2$= q.$itemset_2$ and ... p.$itemset_{k-2}$= q.$itemset_{k-2}$ and p.$itemset_{k-1}$< q.$itemset_{k-1}$ then begin

c=p∞q；

if has_infrequent_subset(c，L_{k-1}) then

delete c；//删除含有非频繁项目子集的候选元素

else add c to C_k；

end

return C_k；

函数has_infrequent_subset(c，L_{k-1}) 作用是判断一个候选项集c是否有子(k-1)-项集不在L_{k-1}中，具体描述如下：

for all (k-1)-subset s of c do

IF s∉L_{k-1} then

return true；

return false；

程序说明：

（1）Scan（D）：安全性评价数据库D为处理对象，扫描数据库，发现其中频繁项集的最大可能长度，返回该值。

（2）c=p∞q：项目集p和项目集q链接。例如项目集p={1,2,4}，项目集q={1,3,4},那么他们链接就生成项目集c={1,2,3,4}.

本模型对自查评表中查评项目 1.1.1.1 项前 30 次的查评历史数据进行分析，设定最小支持度（min_sup）为 0.4，最小置信度（min_conf）为 0.6。项目 1.1.1.1 的事务表如表 4.2 所示。

表 4.2 最终事务表

Tab.4.2 Final transaction table

tid	selfRate	selfLevel
1	A_1	B_1
2	A_2	B_1
3	A_1	B_2
4	A_1	B_1
5	A_2	B_1
6	A_1	B_2
7	A_1	B_2
8	A_1	B_1
9	A_2	B_1
10	A_1	B_2
11	A_1	B_1
12	A_2	B_2
13	A_2	B_2
14	A_1	B_2
15	A_2	B_2
16	A_1	B_1
17	A_1	B_1
18	A_2	B_2
19	A_1	B_2
20	A_1	B_2
21	A_1	B_2
22	A_2	B_2
23	A_1	B_2
24	A_1	B_1
25	A_1	B_2
26	A_1	B_1
27	A_1	B_2
28	A_1	B_1
29	A_1	B_2
30	A_2	B_2

具体实现过程如下：

①扫描自查评最终事务表，得到候选 1-项集C_1，如图 4.6 所示。sup_count表示项集在数据库中的计数，计算A_1的支持度：

$$\sup(A_1) = 21/30 = 0.7 > \min\_sup = 0.5$$

故A_1为频繁项集。

计算A_2的支持度：

$sup(A_2) = 9/30 = 0.3 < min\_sup = 0.5$

所以A_2不是频繁项集。

同理，B_1和B_2均频繁项集，A_2不是频繁项集。

候选 1-项集C_1

item	sup_count
A1	21
A2	9
B1	12
B2	18

比较C_1中各项支持度 ⇒

频繁项集L_1

item	sup_count
A1	21
B1	12
B2	18

图 4.6　频繁项集L_1

Fig.4.6　Frequent itemsets L_1

②频繁项集L_1自连接，得到候选 2-项集C_2，如图 4.7 所示。分别计算A_1B_1、A_1B_2和B_1B_1的支持度，可以得到为A_1B_2为频繁项集。

候选 2-项集C_2

item	sup_count
A1B1	9
A1B2	12
B1B2	0

比较C_2中各项支持度 ⇒

频繁项集L_2

item	sup_count
A1B2	12

图 4.7　频繁项集L_1

Fig.4.7　Frequent itemsets L_1

此时，所有的频繁项集均已找出，即频繁项集L_1和频繁项集L_2，算法结束。

那么，频繁项集L_1和频繁项集L_2我们可以得出如下四条关联规则：A_1=>B_1、B_1=>A_1、A_1=>B_2、B_2=>A_1。

对于所得到的全部频繁项集，并不都是有用的关联规则，即强关联规则。我们要根据置信度来判断该关联规则是否为强关联规则。具体运算情况如下：

对于关联规则A_1=>B_1，置信度

$$\mathrm{conf}(A_1 \Rightarrow B_1) = P(B_1 \mid A_1) = \frac{\sup\_count(A_1 \cap B_1)}{\sup\_count(A_1)} = \frac{9}{21} = 0.43$$

对于关联规则$B_1 \Rightarrow A_1$，置信度

$$\mathrm{conf}(B_1 \Rightarrow A_1) = P(A_1 \mid B_1) = \frac{\sup\_count(A_1 \cap B_1)}{\sup\_count(B_1)} = \frac{9}{12} = 0.75$$

同理：$\mathrm{conf}(A_1 \Rightarrow B_2) = 0.57$；

$\mathrm{conf}(B_2 \Rightarrow A_1) = 0.67$。

根据以上计算，再根据Apriori算法性质可得：$B_1 \Rightarrow A_1$、$B_2 \Rightarrow A_1$为强关联规则。

4.3 数据挖掘结果解释

由上面的挖掘结果可知，关联规则$B_1 \Rightarrow A_1$、$B_2 \Rightarrow A_1$是最终确定的关联规则。B_2代表问题严重程度为“严重”，A_1代表得分率在[0,0.5]之间。对于关联规则$B_1 \Rightarrow A_1$，置信度为 0.75。意味着当问题严重程度为“正常”时，得分率在[0,0.5]的概率为 0.75。这种关联规则就不符合我们平时的生活逻辑，即项目正常，得分率低。因此这条关联规则无疑对管理来说没有意义。对于关联规则$B_2 \Rightarrow A_1$，支持度为 0.4，置信度为 0.67。意味着当问题严重程度为“严重”时，得分率在[0,0.5]之间的概率为 0.67。那么针对这种得分率低而且问题严重程度高的关联规则，管理者可以对 1.1.1.1 项提前提出控制措施，达到预防事故的目的。

5 加权关联规则在本系统中的应用

关联规则可以发现存在于项目之间的潜在关系，但是，并不是所有的关联规则都是用户感兴趣的。比如第4章我们找到的关联规则B_1=>A_1，就是用户不感兴趣或者说错误的。而且Apriori算法是基于两大前提假设[48-50]：（1）数据库中各个项目的性质和作用相同，即具有同等的重要性；（2）数据库中各个项目是均匀分布的，即出现的频率相同或相似。然而，现实生活中往往不是这样的。例如，对于一个商场的管理人员来说，最关心的是商品的利润，一个液晶电视的利润比MP3的利润往往要大出很多倍，但是液晶电视的销售量比MP3小很多，即对于同样的数据库，液晶电视的支持度就很小，那么管理者感兴趣的液晶电视的关联规则很有可能被忽略掉。但是，如果将支持度设置的太小，又会产生很多无用甚至得虚假的关联规则，如何解决这个矛盾，加权关联规则在数据挖掘中提了出来。

加权关联规则在项目中提出了加权的概念，把用户关心的属性赋予较大的权值，不关心的属性赋予较小的权重。通过设置项目中属性的权重，数据库中各个项目的重要程度将会不同，改进Apriori算法存在的缺点。

5.1 加权关联规则介绍

加权关联规则描述这样一个问题：在数据库D中共有p条记录，q个项目。设I={i_1，i_2，…，i_m}是一个数据库，T=（t_1，t_2，…，t_n）表示数据库的属性集，每一个项目均有一个权值与其对应[51-53]。他们的权值分别为{w_1, w_2, …, w_m}，其中$w_i \in [0,1]$。记support(X)为数据库中支持X的交易数，n为数据库中的交易总数，与关联规则一样，加权关联规则也存在支持度与置信度。我们所要讨论的布尔型加权关联规则的形式为“X->Y”。那么给出如下定义。

定义 1 属性集X=(x_1，x_2，…，x_n)的加权支持度为

$$w\sup(X)=\sum_{i=1}^{n} w_i \times \frac{\sup port(X)}{n} \tag{5.1}$$

定义 2 关联规则形如 X=>Y 的加权支持度为

$$w\sup(\mathrm{X} \Rightarrow \mathrm{Y})=\sum_{i=1}^{m+n} w_i \times \frac{\sup port(X \cap Y)}{n} \tag{5.2}$$

定义 3 关联规则形如 X=>Y 的加权置信度为

$$wconf(\mathrm{X} \Rightarrow \mathrm{Y})=\frac{\sup port(X \cap Y)}{\sup port(X)} \tag{5.3}$$

关联规则 Apriori 算法的核心是如果一个属性集是频繁项集，那么它的所有子集也都是频繁项集。当为每个属性分配了权重之后，属性集的加权支持率就不再满足这个性质。但由于属性集的加权支持率总是小于支持率，因此可以给出下面的加权关联规则挖掘算法：

输入：数据库$I=\{i_1, i_2, \ldots, i_m\}$，最小加权支持度，最小加权置信度。

输出：加权关联规则。

（1）先不考虑布尔型属性的权重，利用布尔型属性关联规则 Apriori 算法中找大属性集的方法，找出支持率不小于用户给定最小加权支持率的属性集。由于加权支持率小于支持率，容易知道这些属性集组成的集合是加权大属性集组成集合的超集；

（2）计算这个超集中所有属性集的加权支持率，并把加权支持率小于最小加权支持率的大属性集删除，从而得到所有的加权大属性集。这步的计算中只需乘上相应的属性权重，无须再去扫描数据库，因此只花费很少时间，也就是说发现加权大属性集花费的时间与布尔型属性关联规则中发现大属性集花费的时间差不多；

（3）利用加权大属性集生成所需的加权关联规则，由于加权置信度度的定义与置信度度的定义相同，因此可以直接利用布尔型关联规则的生成算法。

5.2 加权关联规则在系统中的应用

电力系统安全性评价是对国家标准中的项目进行评价，企业将需要评价的项目分配给各个部门，各个部门再将其分给各个查评人。那么，在这个过程中，同一项目很有可能多次分给不同的部门或人。例如，项目 1.1.2 的评价任务可以同时分给安监部门和生技部门，安监部门和生技部门又可以将此项目分给下属不同的人评价。不同的人对项目的理解程度是不一样的，因此在查评的过程中对项目的打分以及对问题是否严重的理解程度也是不一样的。

为了说明各个项目的属性在用户心目中的重要程度是不一样的，我们对第 4 章给出的算例进行分析。$\{A_1,A_2,B_1,B_2\}$是项目属性集，分别表示得分率为[0,0.5]、得分率为(0.5,1]、问题正常、问题严重。我们可以将表中的前 6 条数据提取出来，如表 5.1 所示。其中(1,0,1,0)表示A_1和B_1出现，A_2和B_2不出现。

表 5.1 数据表

Tab.5.1 Data sheet

A_1	A_2	B_1	B_2
1	0	1	0
0	1	1	0
1	0	0	1
1	0	1	0
0	0	1	0
1	0	0	1
…	…	…	…

由表 5.1 可以看出，数据库中各属性的重要程度和出现的概率是不一样的。即在一条记录中，有的出现有的不出现。

我们对 4 章找到的关联规则进行分析，如表 5.2 所示。

表 5.2 关联规则

Tab.5.2 Association rules

关联规则	支持度	置信度
B_1=>A_1	0.3	0.75
A_1=>B_2	0.4	0.57
B_2=>A_1	0.4	0.67

由表 5.2 可得三条关联规则B_1=>A_1、A_1=>B_2、B_2=>A_1。对于关联规则B_1=>A_1，置信度为 0.75。意味着当问题严重程度为“正常”时，得分率在[0,0.5]的概率为 0.75，即项目正常，得分率低，这种关联规则不符合我们平时的生活逻辑。这条关联规则无疑对管理来说没有意义或者说是错误的。对于关联规则A_1=>B_2解释如下：得分率在[0,0.5]之间，项目问题严重的概率为 0.57。显然这条关联规则也是管理者感兴趣的，然而这样的关联规则由于种种原因将其忽略掉了。

综上所述联规则B_1=>A_1是无意义的关联规则，而有意义的关联规则A_1=>B_2却被忽略掉了。因此，传统的Apriori算法已经无法满足用户的需要。为了解决这些矛盾，我们在项目属性值里面引入加权的概念。根据人们日常的生活经验，管理者肯定对得分率低和问题严重的项目更感兴趣，那么笔者对各种状态赋予的权值如表 5.3 所示。

表 5.3 设置权值

Tab.5.3 Set weight

item	weight(权值)
A_1	0.8
A_2	0.3
B_1	0.2
B_2	0.9

文献[54]提出了 K-支持度期望概念，并提出了加权关联规则算法通过算例对算法进行了仿真，但同时也存在小问题，加权关联规则的支持度有可能大于 1，这与人们的实际逻辑思维是不相符的。文献[55]对属性的权重集进行了归一化处理，解决了加权支持率可能大于 1 的不足，提出第 1 类加权关联规则挖掘算法，此算法能有效地考虑属性的权重，并且认为规则的重要性随着规则中所含属性数量的增加而增加。

为了避免支持度大于 1 的情况，我们对表 5.3 中的权值进行归一化的处理。具体过程如下：

设 $\gamma = w_1 + w_2 + \ldots + w_m$，则（$w_1/\gamma$，$w_2/\gamma$，…$w_m/\gamma$）即是归一化后的属性权重集，记为 β =（w_1/γ，w_2/γ，…，w_m/γ），其中 $w_1/\gamma + w_2/\gamma + \ldots + w_m/\gamma = 1$。

对于权重集 w=（0.8，0.3，0.2，0.9）归一化后的权重如表 5.4 所示。

表 5.4 归一化权值

Tab.5.4 Normalized weight

item	weight(权值)
A_1	0.364
A_2	0.136
B_1	0.091
B_2	0.409

现在，再用加权支持率与加权置信度度来分析表 5.2。加权支持度为 β =（0.364，0.136，0.091，0.409），最小支持度为 0.4，最小置信度为 0.6，设定最小加权支持率为 0.2，最小加权置信度为 0.5。

由公式 5.1 和 5.2 计算加权支持度和加权置信度，可得表 5.5 的数据。

表 5.5 加权关联规则

Tab.5.5 Weighted association rules

关联规则	支持度	置信度	加权支持度	加权置信度
B_1=>A_1	0.3	0.75	0.137	0.75
A_1=>B_2	0.4	0.57	0.309	0.57
B_2=>A_1	0.4	0.67	0.309	0.67

通过加权关联规则的挖掘，将所计算的加权支持度、加权置信度与最小加权支持度、最小加权置信度比较，得到两条关联规则A_1=>B_2和B_2=>A_1。与第四章找到的关联规则对比可知，挖掘出了一条重要的关联规则A_1=>B_2，即得分率在[0,0.5]之间，项目问题严重的概率为 0.57，剔除了无用的关联规则B_1=>A_1。

在系统运用的算法中加入加权的概念，挖掘出了使用 Apriori 算法未挖掘出的关联规则，剔除了以前使用 Apriori 算法中存在的无用的关联规则，使用户得到的关联规则更可信。

6 系统界面设计

本章使用 ASP.NET 对系统的整体界面进行了设计，设计的这个系统可以作为一个单独的模块嵌套在电力系统安全性评价的各个方面，包括发电厂、调度、输电网、供电企业等安全性评价软件中使用，故具有一定的通用性。

6.1 系统运行环境

系统设计采用 B/S 结构模式，即浏览器/服务器模式。这种结构最大的优点是可以在任何地方进行操作而不用安装任何专门的软件，只要一台能上网的电脑就能使用。系统用户可以使用各自原有 PC 机通过浏览器访问系统服务器，对系统进行操作。利用 ASP.NET 技术和 ADO.NET 数据访问接口，将 B/S 模式的数据库结构与 WEB 技术密切结合。前台开发工具为 Microsoft Visual Studio 2005，后台数据库采用 Microsoft SQL Server 2000 数据库管理系统，运行所需的软、硬件环境如表 6.1 所示。

表 6.1 系统运行环境

Tab.6.1 The environment of system operating

硬件环境	奔腾 2.4、内存 512、硬盘 80G 及以上
软件环境	数据库软件：Microsoft SQL Server 2000 网络浏览器：Microsoft Internet Explorer 5.0 以上 其他：Microsoft office 2003 套件等
应用平台	Microsoft Visual Studio 2005
编程语言	Visual C#

6.2 系统登录

电力系统安全性评价辅助决策系统是是作为安全性评价的辅助工具，此处，本模块是嵌套在供电企业安全性评价中使用的，故使用的是安全性评价的登录界面。系统的登录界面如图 6.1 所示。

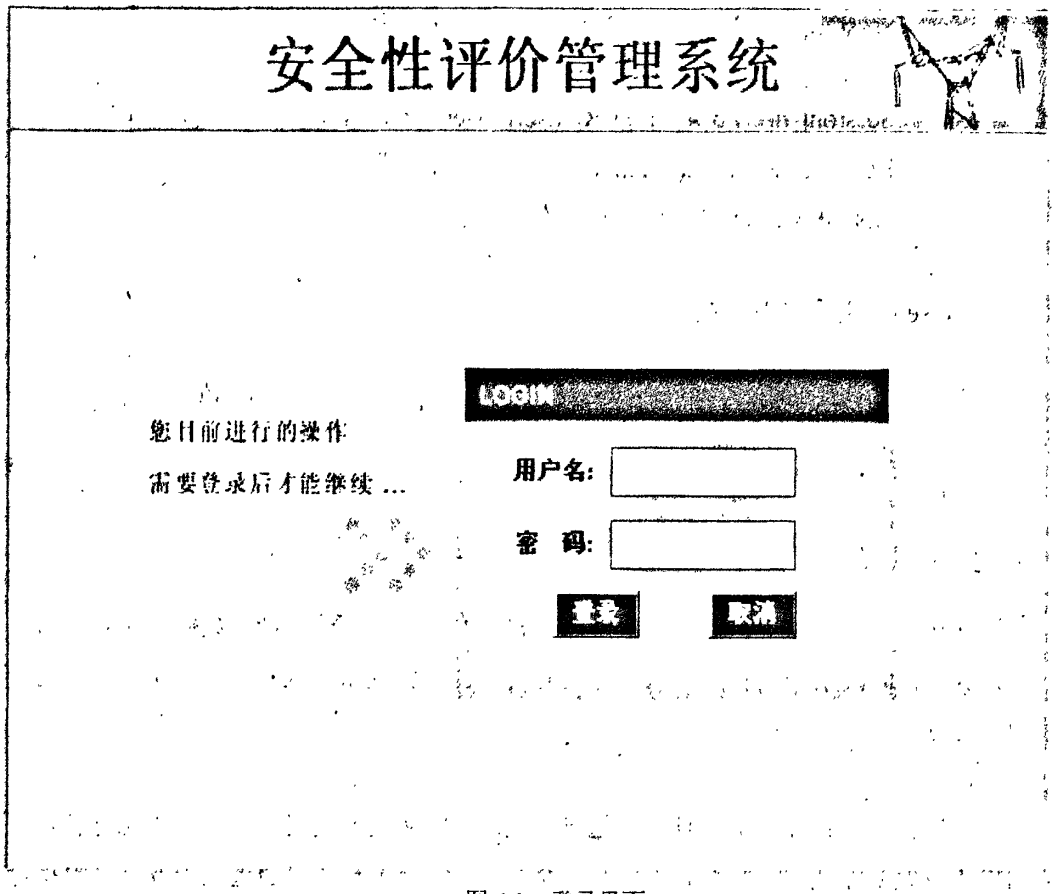


图 6.1 登录界面

Fig.6.1 Login page

主要功能如下：

（1）用户登录。输入用户名和密码，系统会自动与数据库中用户表的数据进行比较。如果输入信息正确，则跳转到系统主操作界面；如果输入信息不正确，则给出错误提示，请用户重新确认个人信息。

（2）用户信息记录。用户登录后，使用 Session 对象自动记录用户信息，包括用户编号、姓名、权限、部门等。用户在页面进行操作时可以自动识别用户的身份。程序段如下：

```
int flag = DBA.GetRowCount(strSql);//在数据库中逐条比对用户
if (flag > 0)//如果用户存在
{
    DataRow DR = DBA.GetRecord(strSql);
    Session["userID"] = DR["userID"].ToString();//记录用户编号
```

```
        Session["userName"] = DR["userName"].ToString();//记录用户姓名
        Session["depID"] = DR["depID"].ToString();
        Session["userPower"] = DR["userPower"].ToString();//记录用户权限
        Response.Redirect("Main.aspx");//跳转至主界面
        this.tbPwd.Text = "";//密码栏设置为空
    }
    else//如果用户不存在
    {
        lblError.Text = "用户名或密码不正确，请重新输入：";//提示错误
        tbPwd.Text = "";
    }
```

6.3 辅助决策

辅助决策功能包括数据仓库、数据挖掘和危险点汇总三个功能。数据仓库对任何用户都开放，辅助决策功能和危险点汇总仅向具有管理员权限的人员开放。我们对辅助决策的功能分别介绍如下：

（1）数据仓库

安全性评价数据库中的查评数据通过预处理（包括数据提取、数据清理、数据转换、数据集成、数据加载和汇总）后放入数据仓库中，系统用户可以通过某个或某几个指标对数据进行查询。数据仓库主界面如图 6.2 所示。

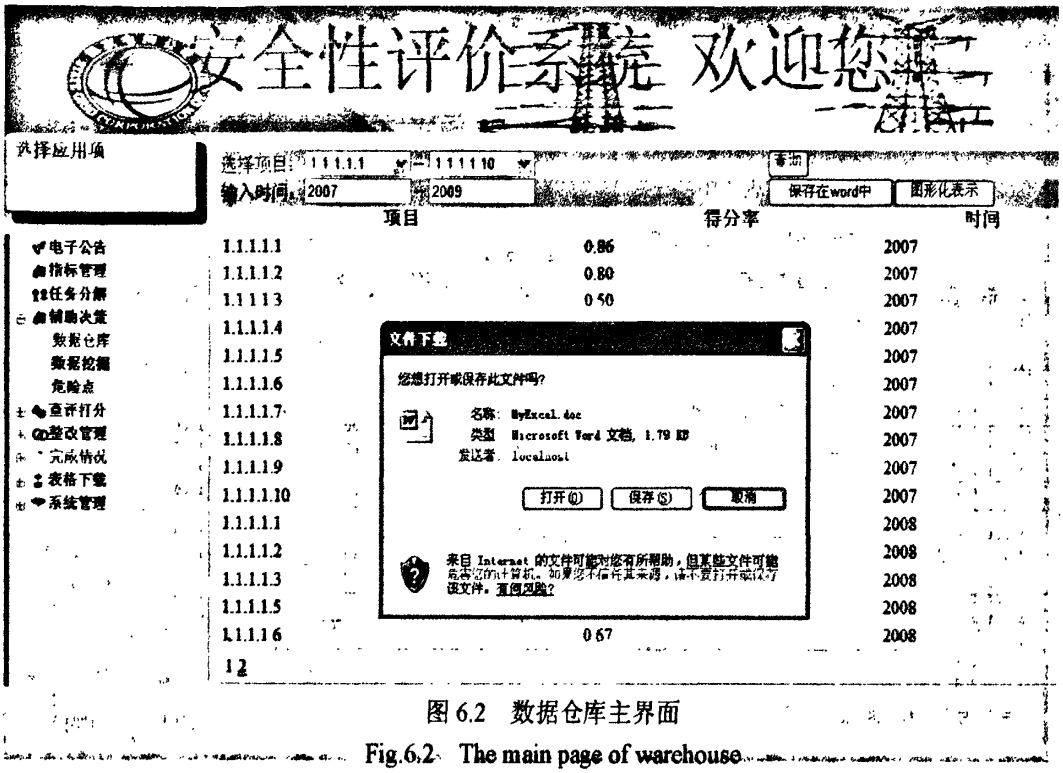


图 6.2 数据仓库主界面

Fig.6.2 The main page of warehouse

由图 6.2，可以通过对项目、项目得分率、时间进行综合查询，查询结果可以以表格的形式保存在 word 文档中，查询到数据也可以以柱状图的方式展现给用户，通过直观的对比，用户可以很清楚的了解项目目前的情况。

例如：用户要查询 2007、2008、2009 年对指标 1.1.1.1.1－1.1.1.1.10 的得分率(selfRate)，这个查询就包括时间维度、指标维度等。查询结果如图 6.3 所示。

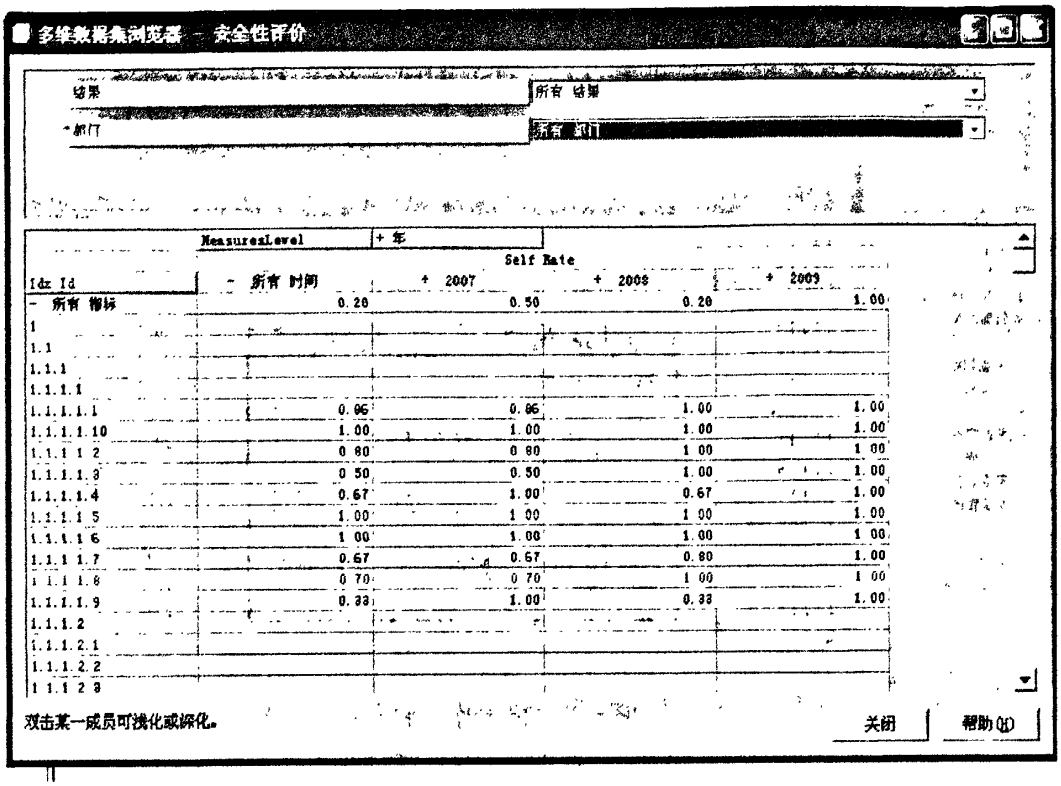


图 6.3　查询结果

Fig.6.3　Query result

从图 6.3 中可以看出，对于项目 1.1.1.1.9，所有时间查评的综合得分率为 0.33，2007、2008、2009 年的得分率分别为 1.00、0.33、1.00；对于项目 1.1.1.1.10，在所有时间内查评的综合得分率为 1.00，2007、2008、2009 年的得分率均为 1.00。那么管理者可以通过数据仓库查询到的结果对项目进行初步判断，通过项目的得分率，管理员可以初步判断项目 1.1.1.1.9 存在危险点，项目 1.1.1.1.10 安全。

查询结果保存 word 中的算法：

```
protected void Button2_Click(object sender, EventArgs e)
{
    Response.ClearContent();
    Response.AddHeader("content-disposition",
    "attachment;filename=MyExcel.doc");//保存在文件MyExcel.doc中
    HttpContext.Current.Response.Charset = "UTF-8";//页面编码为UTF-8
    HttpContext.Current.Response.ContentEncoding = System.Text.Encoding.Default;
    Response.ContentType = "application/word";文本类型为word
```

```
        StringWriter sw = new StringWriter();
        IHtmlTextWriter htw = new IHtmlTextWriter(sw);
        GridView1.AllowPaging = false;//不允许翻页
        this.BindGridView1();//在页面中绑定查询的数据
        GridView1.RenderControl(htw);
        Response.Write(sw.ToString());
        Response.End();
        GridView1.AllowPaging = true;//允许翻页
    }
```

同样的，查询的数据还可以以柱状图的形式直观的展示给用户，如图 6.4 所示。图中给出的是得分率 selfRate 在 2007 和 2008 年对比的情况。这种表示方式使用户对于查评的数据的了解更加直观和方便。

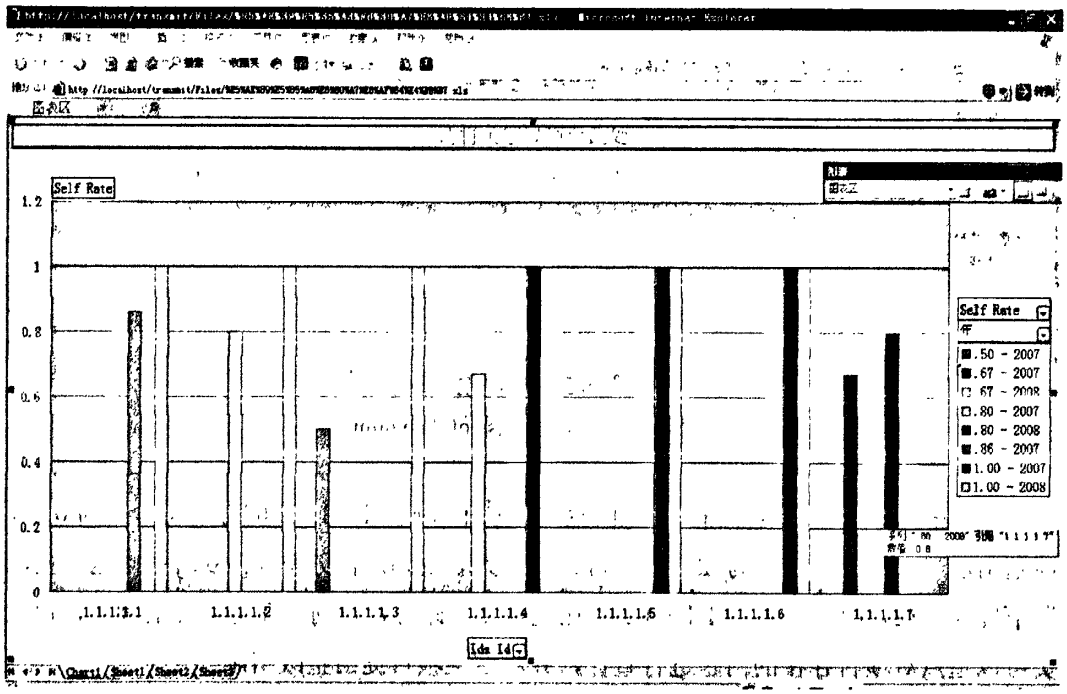


图 6.4　直观图

Fig.6.4　Illustrative diagram

如图 6.4，横坐标代表查评项目（idxID），纵坐标代表查评得分率（SelfRate）。用户可以很清楚的对 2007 年和 2008 年同一项目的得分率进行比较。项目 1.1.1.1.1 在 2007 年查评的得分率为 0.86，在 2008 年查评的得分率为 1.00。

（2）数据挖掘

根据安全性评价结果和相应的安全生产状态，运用 Apriori 算法对数据库中的数据进行挖掘，发现查评指标存在的安全隐患。在数据挖掘界面中，管理员可以任意选择要决策的项目，根据自己的经验来选择最小加权支持度和最小加权置信度。

例如，管理员用户要对项目 1.1.1.1.2 进行决策，输入最小加权支持度 0.2，最小加权置信度 0.5，通过 Apriori 算法计算，就可以知道此项是否存在危险点。界面如图 6.5 所示。

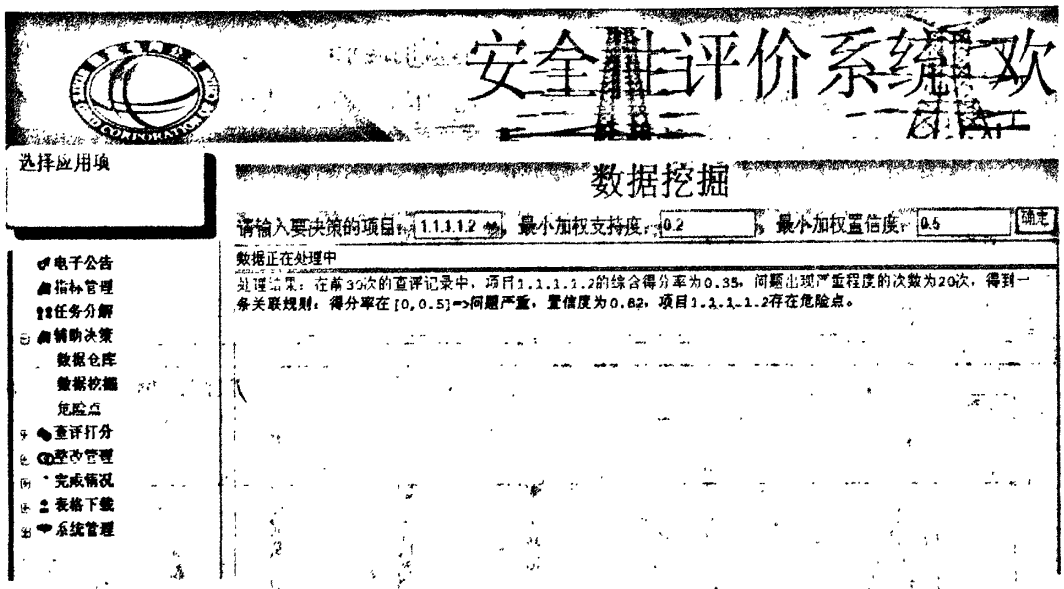


图 6.5　数据挖掘过程

Fig.6.5　The process of data mining

从图中 6.5 可以看出，在前 30 次的查评记录中，项目 1.1.1.1.2 的综合得分率为 0.35，问题出现严重程度的次数为 20 次，得到一条关联规则：得分率在[0,0.5]=>问题严重，置信度为 0.82，故认为项目 1.1.1.1.2 存在危险点。针对查找到存在危险点的项目，管理员可以重点对项目存在的问题进行查评，并对这些项目提出整改措施。

Apriori 算法找出频繁项集程序：

```
namespace testApriori1
{
    public class TestApriori
    {
        static void Main(string[] args)
        {
```

```
            Console.Write(DateTime.Now);//系统当前时间
            ArrayList D = GetEventsFromDB();//事务数据集
            ArrayList I = GetItems1FromDB();//初始项目集合
            float s = 0.2;//支持度
            List L = new List();//所有频繁项集
            L = Apriori(D, I, s);
            for (int i = 0; i < L.Count; i++)
              {
                Console.WriteLine(L[i].Items);//输出频繁项集
                Console.WriteLine(L[i].Sup);//输出各项集支持度
              }
              Console.Write(DateTime.Now);
              Console.Read();
          }
      }
  }
```

（3）危险点汇总

在辅助决策的过程中，系统会自动对查找出的危险点进行记录，便于管理员查看与下载。如图 6.6 所示。

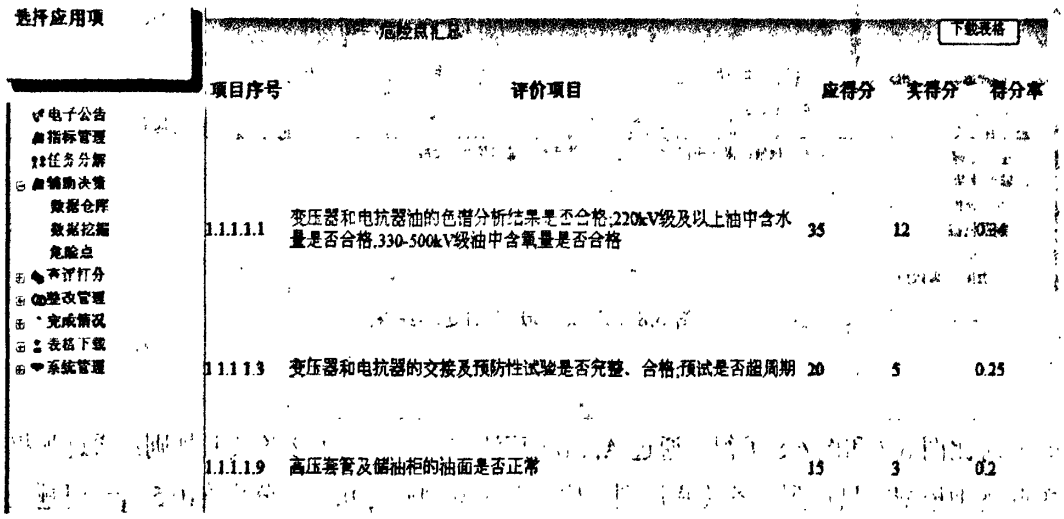


图 6.6 危险点汇总

Fig.6.6 Summary dangerous points

6.4 Apriori 算法和加权关联规则比较

本节通过界面数据对两者进行比较。图 6.7 是利用 Apriori 算法挖掘的结果，图 6.8 是利用加权关联规则挖掘的结果。

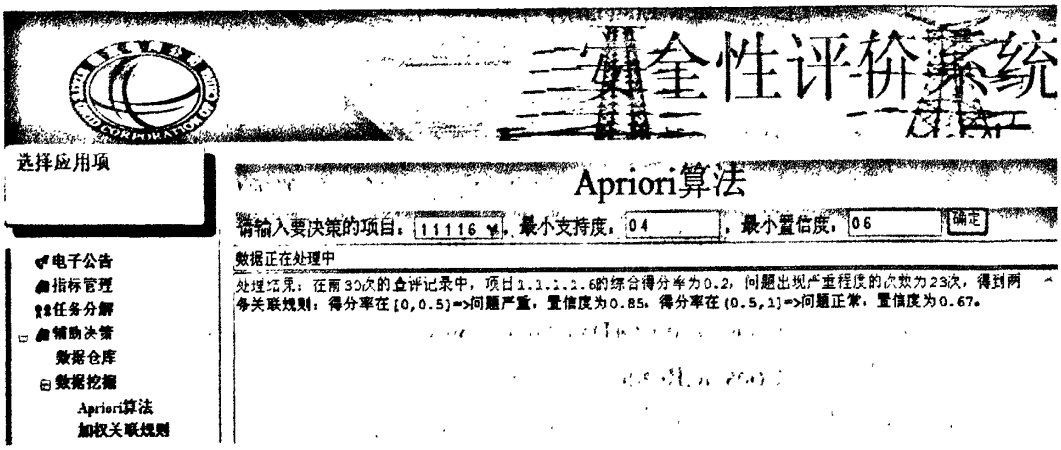


图 6.7 Apriori 算法处理结果

Fig.6.7 The results of Apriori algorithm

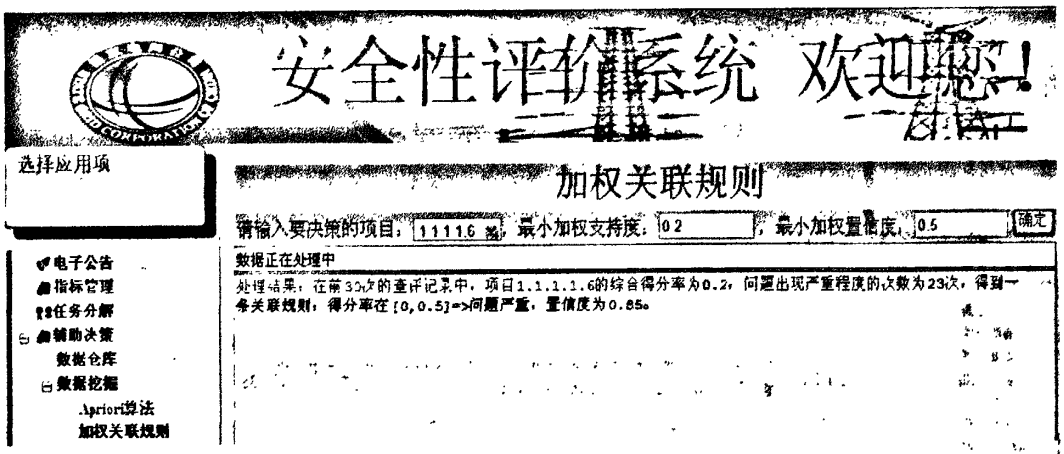


图 6.8 加权关联规则处理结果

Fig.6.8 The results of weighted association rules

对比图 6.7 和图 6.8 可得，通过 Apriori 算法挖掘，得到了 2 条关联规则；通过加权关联规则挖掘，只得到一条关联规则。比较两关联规则可知：得分率在(0.5,1]=>问题正常，置信度为 0.67，这条关联规则是用户不感兴趣的过则，通过加权关联规则的挖掘，这条关联规则被剔除。因此，在系统中引入加权的概念，能剔除无用的关联规则，挖掘出的关联规则更可信。

7 结论和展望

7.1 结论

随着计算机的普及和网络技术的迅速发展，人们进入了数据大爆炸的时代，传统的数据库仅仅用来储存数据，数据库中的数据不能直观的展现给用户，更不能对数据进行更多的分析处理。如何从大量的数据中获取人们需要或者感兴趣的信息成为了人们研究的关键。

本文开发了基于 B/S 模式的电力系统安全性评价辅助决策系统，辅助决策系统针对电力系统存在的危险点，提出了一种查找方法：数据挖掘方法。通过建立一个智能化的数据挖掘工具，从电力系统安全性评价数据库的大量历史数据中获取电力系统中存在的危险点从而指导管理者进行决策。数据挖掘工具的核心是：采用 Apriori 算法查找关联规则从而判断影响电网安全性的危险点。通过危险点查找，确认发生事故的可能性及其严重程度，管理者对危险点提出相应的整改和控制措施，达到预防、控制隐患和事故。

下面我们对文章的特点做个总结：

（1）文章对电力系统安全性评价体系做了全面深入的研究，分析了体系存在的优点及不足。

（2）建立了安全性评价数据仓库，它具有强大的数据查询和汇总功能，并根据用户需要提供历史数据的生成及下载功能。

（3）利用当前流行的数据挖掘技术对安全性评价历史数据进行分析，对挖掘出的危险点供决策者辅助决策之用。

（4）辅助决策系统采用流行的 B/S 结构设计，这样降低了软件的维护成本，同时系统具有很好的独立性，即可以做为单独的模块使用，可以用于调度安全性评价、发电厂安全性评价等多个方面。

（5）对数据挖掘中的 Apriori 算法做了改进，提出了加权的概念，使挖掘出的危险点更可信。

7.2 展望

电力系统安全性评价辅助决策系统实现了对电力系统危险点的预测，但是系统运用的算法还存在不足之处，就是最小支持度确定的问题。对于管理者来讲，如何选择合适的最小支持度成了主要问题。针对此问题，系统可以进一步对最小支持度选定进行研究，选择适当的算法，结合用户实际情况，估计或预测符合电力系统安全性评价规律的最小支持度。

另外，辅助决策系统中只给出了两种影响危险点的因素，即得分率和问题的严重程度，如果能从安全性评价中找到第三个影响危险点的因素，那么将大大增加关联规则的可信程度。

以上两个问题结合起来研究，将大大提高辅助决策系统决策准确性，对于我国电力系统安全性评价工作的开展具有很现实的意义。

参考文献

[1] 蔡树人. 再谈电力生产企业安全性评价[J]. 电力安全技术，1999，(2)：2-4.

[2] 胡涛，黄健，鄢威. 一种电力生产安全性评价方法研究及应用[J]. 电力系统保护与控制，2009，37(16)：46-49.

[3] 干金萍，樊凤林，刘发旺，等. 安全性评价在电力企业中的应用研究[J]. 华北电力技术，2005，(5)：23-26.

[4] 刘俭. 发供电企业安全性评价方法研究及实践[J]. 中国电力企业管理，2001，(10)：21-22.

[5] 袁忠军. 浅谈电力企业的安全性评价工作[J]. 广西电力，2004，(03)：55-57.

[6] 河南省电力公司安监处. 安全性评价工作初见成效[J]. 水利电力劳动保护，2000，(2)：6-7.

[7] 刘学东，张兴平. 电力企业安全性评价动态管理特性解析[J]. 太原理工大学学报，2006，37：148-150.

[8] 杨国鸿. 火电厂安全性评价与危险点预控分析[J]. 宁夏电力，2008，(04)：58-60.

[9] 马丽英，李凤荣. 电力企业安全性评价改进[J]. 电力学报，2009，24(20)：165-166.

[10] 戴庆辉，宋卫霞. 供电企业危险点预控管理系统[J]. 电力系统自动化，2007，31(1)：100-103.

[11] 李秋丹，迟忠先，金妮，等. 基于数据仓库的地区电网调度决策支持系统[J]. 电力系统自动化，2004，28(12)：86-89.

[12] 李裕辉. 危险点分析与控制的方法与原则[J]. 电力安全技术，2006，8(8)：2-5.

[13] 周浩，张富强，韩祯祥，等. 电力市场金融风险评估和辅助决策系统设计与实现[J]. 电网技术，2005，29(9)：56-62.

[14] 余毅. 数据仓库技术在电力辅助决策系统中的应用与研究[D]. 四川：电子科技大学，2005.

[15] 李惠汾. 供电企业管理辅助决策系统的开发与运用[J]. 电力学报，2007，22(3)：260-263.

[16] 陈堂，赵祖康. 供电网调度辅助决策专家系统[J]. 电力系统自动化，1991，(02).

[17] 袁林. 基于数据仓库的辅助决策系统的设计与实现[J]. 电力系统自动化，2001，21(25)：51-56.

[18] 李秋丹，迟忠先，孙瑞超. 基于联机分析与数据挖掘的电力负荷分析系统[J]. 计算机集成制造系统，2004，10(专刊)：191-194.

[19] 陈星莺，张晓花，瞿峰，等. 数据挖掘在电力系统中的应用综述[J]. 电力科学与技术学报，2007，22(03)：51-56.

[20] L. H. Fink,K. Carlsen,R. Fischl,etc.Power system security assessment[J].Decision and Control.1984,The 23rd IEEE Conference on,1984,23(1):478-480.

[21] 沈香武. 电力系统安全性评价的设计和实施[J]. 电力信息化，2005，3(2)：46-49.

[22] 刘学东. 火力发电企业安全性评价体系的建设及应用[D]. 北京：华北电力大学，2006.

[23] 勒江红. 工业企业电气安全评价方法及应用研究[D]. 北京：首都经济贸易大学，2002.

[24] 揭兴松. 发供电企业安全性评价[J]. 湖北电力，2001，25(1)：36-37.

[25] 潘守翡. 供电企业安全性评价智能管理系统的设计与实现[D]. 四川：西华大学，2010.

[26] 邢大鹏. 电网调度安全性评价管理系统的设计和研究[D]. 四川：西华大学，2008.

[27] Pan shou fei，Lei Xia，Ji er yun.The Application of Data Warehouse Technology In Security Assessment Management System[C].The Asia-Pacific Power and Energy Engineering Conference, Chengdu, 2010.

[28] 潘守翡，雷霞，余光亮，等．供电企业安全性评价自动化管理系统的设计和实现[J]．四川电力技术，2009，32(2)：11-12；36.

[29] 国家电网公司．供电企业安全性评价标准[S]．北京：中国电力出版社，2002.

[30] 王博．浅谈数据仓库在电信企业中的应用[J]．信息与电脑，2010，(3)：136-139.

[31] 张英朝，邓苏，张维明．数据仓库元数据管理研究[J]．计算机工程，2003，39(1)：8-10.

[32] 朱建平．数据挖掘的统计方法及实践[M]．北京：中国统计出版社，2005.

[33] 胡侃，夏绍玮．基于大型数据仓库的数据采掘：研究综述[J]．软件学报，1998，9(1)：53-63.

[34] 戴小廷，陈荣思．数据挖掘在电力系统中的应用[J]．重庆科技学院学报(自然科学版)，2009，11(3)：141-144.

[35] 高杰，李绍军，钱锋．数据挖掘中关联规则算法的研究及应用[J]．东南大学学报(自然科学版)，2006，36(增刊)：128-131.

[36] 罗可，吴杰．关联规则衡量标准的研究[J]．控制与决策，2003，18(3)：277-280.

[37] 钱铁云，王元珍，冯小年．结合类频率的关联中文文本分类[J]．中文信息学报，2004，18(6)：30-36.

[38] 程政，雷霞，廖翔，等．数据挖掘在电网安全性评价中的应用[J]．电气技术，2008，(08)：97-99.

[39] 杨苹，吴捷．火电厂锅炉常见故障的数据挖掘诊断方法[J]．仪器仪表学报，2005，26(7)：696-701.

[40] 朱六璋，袁林，黄太贵．短期符合预测的使用数据挖掘模型[J]．电力系统自动化，2004，28(3)：49-52.

[41] 张增敏，谢嘉，李长河．数据挖掘在变电站设备缺陷预测管理中的应用[J]．哈尔滨工业大学学报，2009，41(1)：272-274.

[42] 田晓军，闫红艳，陈树勇．数据挖掘及其在电力事故分析中的应用[J]．电力设备监测与故障诊断，电力系统及其自动化专业第二十一届学术年会论文集，2284-2287.

[43] 雷霞，潘守翡，张弛，等．安全性评价管理系统及其辅助决策系统的应用研究[J]．电力系统保护与控制，2010，38(5)：88-91.

[44] 薛小平，张思东，王小平，等．RFID 网络的数据清理技术[J]．计算机工程，2008，34(7)：92-94.

[45] 张小星，程其云，周淥，等．基于数据挖掘的电力负荷脏数据动态智能清洗[J]．电力系统自动化，2005，29(8)：60-64.

[46] 张国江，邱家驹，李继红．基于人工神经网络的电力负荷坏数据辨识与调整[J]．中国电机工程学报，2001，21(8)：104-107.

[47] Geiger,M.,Cranor, L.F.Scrubbing Stubbron Data:An Evaluation of counter-Forensic Privacy Tools[J].Security& Privacy,IEEE,2006,4(5):16-25.

[48] 杨红菊，梁吉业．布尔加权关联规则的几种开采算法及比较[J]．电脑开发与应用，2004，17(4)：12-13.

[49] 陈华月，余刚，朱征宇．基于加权关联规则的用户关注项目推荐算法[J]．计算机工程，2006，32(6)：86-88.

[50] 邱凤翔，徐治皋，司风琪，等．基于主元分析的关联规则挖掘及在电厂中的应用[J]．电力系统自动化，2009，33(16)：81-84.

[51] 王涛伟，任一波．基于加权关联规则的个性化推荐研究[J]．计算机应用与软件，2008，25(8)：242-244.

[52] 陈晓云，胡运发．规则加权的文本关联分类[J]．中文信息学报，2006，19(4)：52-59.

[53] 申向东，张海潮．基于加权关联规则的入侵检测研究[J]．现代电子技术，2004，(18)：74-76.

[54] 欧阳为民，郑诚，蔡庆生．数据库中加权关联规则的发现[J]．软件学报，2001，12(5)：612-613.

[55] 张文献，陆建江．加权布尔型关联规则的研究[J]．计算机工程，2003，29(9)：55-57.

攻读硕士学位期间学术论文及情况

1 程政，雷霞，廖翔，等. 数据挖掘在电网安全性评价中的应用[J]. 电气技术，2010年，(8)：97-99. 主办单位：中国电工技术学会。（本硕士论文第四章）

2 柏小丽，雷霞，程政，等. 配电侧购电风险规避模型综述[J]. 电力需求侧管理，2010，12(2)：19-22. 主办单位：国家电网公司电力需求侧管理指导中心。

致　谢

在我论文完成之际，衷心感谢所有关心、帮助过我的老师、同学和亲人们。

首先感谢我的导师雷霞教授，在论文的选题、研究和撰写等方面，雷老师始终给予我很大的关心和支持。三年来在雷老师的谆谆教诲和指导之下我才得以顺利完成学业。雷老师渊博的知识、严谨求实的治学态度、一丝不苟的工作作风都非常值得我学习。在此谨向雷老师表示最衷心的感谢和最崇高的敬意。

感谢我的父母在我求学期间对我生活上的照顾和支持，使我能够顺利完成学业，感谢他们在我学习期间给我的鼓励。

感谢成都天亚电气设备有限责任公司经德华总经理在课题的实际应用给予的指导及做事方法、态度等方面对我的教诲，将使我终身受益。

感谢实验室的全体同学，三年以来我们大家在生活上互相帮助，在学习上互相交流，在不同的方面给了我很多帮助。

感谢各位评委百忙之中抽出宝贵时间审阅本论文！