

# 简析数据挖掘技术在气象业务中的应用

汤津赢<sup>1</sup> 王玮琪<sup>1</sup> 周泽文<sup>2</sup>

1 辽宁省气象信息中心 2 辽阳市气象局

DOI:10.12238/eep.v4i4.1403

**[摘要]** 随着互联网技术的发展,庞大数据中隐藏着很多价值信息,为了在大量的数据中找到其存在的规律,技术人员必须采用数据挖掘技术,从而有效预测未来的发展趋势。气象业务在国民经济建设与科学研究过程中具备十分重要的价值,应用十分广泛,可为基础的地面与高空气象提供更多的参考资料,且很多灾害资料与雷达气象资料中蕴藏着气象规律,可有效预测气象。基于此,本文就数据挖掘技术在气象业务中的应用进行浅要探析。

**[关键词]** 数据挖掘技术; 气象业务; 应用

中图分类号: Q142.9 文献标识码: A

## A Brief Analysis on the Application of Data Mining Technology in Meteorological Services

Jinying Tang<sup>1</sup>, Weiqi Wang<sup>1</sup>, Zewen Zhou<sup>2</sup>

1 Liaoning Provincial Meteorological Information Center 2 Liaoyang Meteorological Bureau

**[Abstract]** With the development of Internet technology, a lot of valuable information is hidden in huge data. In order to find the existence law in a large amount of data, technicians must use data mining technology to effectively predict future development trends. Meteorological services have very important value in the process of national economic construction and scientific research, and are widely used. They can provide more reference data for basic ground and upper air conditions. Many disaster data and radar weather data contain meteorological laws. It can effectively predict the weather. Based on this, this article makes a brief analysis on the application of data mining technology in meteorological services.

**[Key words]** Data Mining Technology; Meteorological Service; Application

在当今世界上信息系统与数据系统最为前沿的一个分析领域就是数据挖掘技术,其可以成功地应用于商业领域。为此,能够将数据挖掘技术应用于气象业务当中,从而分析各种数据,明确数据特点与模式以及改变方向,再根据各种用户的实际需要制作相应的公共气象服务产品。

### 1 数据挖掘相关理论概述

1.1 数据挖掘技术。数据挖掘技术主要应用于商业数据的统计领域,是一种新型的挖掘操作技术,主要是从海量、无序、随机和无价值的数据中挖掘出对自我需求有意义、有价值和特定的知识信息,这是一个连续查找有用知识的过程。最终被查找到的数据信息一定是客户所需要且乐于接受和易于理解的知识信息。但这些知识信息并不是在任何范围内都被需要,

只是在特定的场合被需要的用户所利用。

1.2 数据挖掘的流程。一个数据的挖掘项目会有一个数据信息被挖掘的过程,而这个过程主要由六个环节组成。第一,需求提供。对一个数据挖掘的项目而言,要从业务的角度去理解客户的需求,同时将数据挖掘的技术问题转化为一个有具体特征的知识信息,进而完成项目目标的初步计划。第二,采集分析数据。采集数据主要体现在对数据信息的理解方面,通过处理和识别等手段对采集到的数据质量进行评估和分析能够了解采集到的数据的各种属性。第三,数据预处理。在采集分析完数据后应进行数据预处理,对无用数据进行清理,记录及归纳客户所感兴趣的数据。第四,建模。在这一阶段应选择应用各种模型技术对数据进行重新整

理和分配,使之成为相应技术所需要的数据格式。第五,模型评估。对建立的模型要用科学的方法进行评估,以确保建造模型的可行性,进而完成数据挖掘的最终目标。第六,实施。建造完模型后需要继续发挥模型清理数据的作用,找到有用的知识信息,满足客户的连续需求,及时提供报告以供操作人员及客户使用。

### 2 数据挖掘技术的基本分析方法

2.1 分类和数值预测。基于问题预测理论有两种表现形式,一个是分类,另一个是数值预测。预测分类能够识别出描述内容中所确定的数据类别,然后检索出客户所需要的数据集合。预测分类是一个可以监控连续不断数据内容的过程,可以用连续的数据集合来区分和识别出有描述意义的信息。

2.2 聚类分析。聚类分析就是通过物理和数据集合的形式对研究对象进行细致分组,然后针对这些分组信息进行类似与相似性分析。聚类分析主要是通过相似性原则查找出想要的和有价值的信息,然后将其进行区分和提取。聚类分析可以把与目标数据相关的类别数据全部提取出来,所以被广泛应用在气象、统计、医学、营销、心理学等领域。

2.3 遗传算法。遗传算法是一种全新的搜索算法,主要按照生物遗传的方式进行计算。遗传算法被广泛应用在气象预报之中,能有效提升预报的准确率,并得到相关的集成性结果。但这种算法也有一定缺陷,就是搜索数据信息的能力不强,只能通过局部搜索的方式进行,搜索数据信息所需的时间较长。

2.4 关联规则法。这种方法的主要原理是对以往的数据信息进行分析研究,可以在大范围和大领域内获得客户需要的数据信息资料。此方法可以对以往的天气数据信息进行分析,能够达到精准预测气象变化的要求,提升了气象预报的准确度。此方法也有缺点,在面对众多候选集时,可能发生数据库重复扫描的问题,如果解决不好这个问题,在气象预报中就无法应用这种方法,也就预测不到相对准确的气象信息。

2.5 向量机方法。向量机方法具有精确度高、运行速度快等实际特点,对较小规模的数据信息处理具有一定优势。对于复杂气象条件下的气象预报工作,通过这种方法能获得更多准确度较高的数据信息,所以向量机方法被广泛在气象预报工作中应用。但向量机方法也存在一些缺陷和不足,在处理数据规模较大和种类较多的问题过程中,其处理速度和反映能力变差,无法在全领域应用。

### 3 数据挖掘技术在气象业务中的应用分析

3.1 在数据储存方面的应用。面对海量的气象数据,数据挖掘任务在于:拟合模型、确立数据关系。正是由于数据量大,数据存储是突出的问题。为了解决这个问题,中国气象局建立了全国综合气象信息共享平台(CIMISS),CIMISS数据包括全国

常规站点建站以来的地面资料、高空资料、海洋资料、农气资料、服务资料等14个大类147种资料,在全国建立31个分中心。为了压缩数据存储量以及便于处理,对数据采用分布式存储和分布式处理,同时系统进行气象要素的统计,例如做了1981-2010年日、旬、月、年值等不同类别的统计,为数据价值挖掘和分析提供了便利。基于CIMISS, 国省气象业务部门进行了许多研究,开发了国家级气象业务内网、公服中心一体化产品加工系统、公服中心气象数据服务系统等,西藏气象局建设实时业务集约化监控和应用平台。为便于支持社会对气象数据的需求,国家气象局开发了中国气象数据网。为进一步提高气象数据价值挖掘的效率,从国家到地市局积极研究气象大数据云平台,开展多源气象数据融合研究及应用。

3.2 在预测预报方面应用。现如今气象预测预报的方法主要有:数值预报方法、天气学预报方法、统计预报方法,国内外研究学者将数据挖掘技术应用到气象预测中,取得了丰硕的成果。GuoZ等运用关联规则挖掘技术建立了高原环境物理量场及其MCS东移传播之间的模型,实验证明这种模型的使用,提高了长江流域暴雨灾害天气的预测预报准确率。徐会明等决策树算法生成对雷电趋势预测取得较好的效果。汪春秀等根据浙北地区5-9月的气象资料建立SVM模型,并对汛期降水量进行预测。很多研究成果表明将数据挖掘技术应用到气象预测预报的领域是可行的,对降水、气温、雷电、农业自然灾害等方面预测能力提高。结合西藏特殊的地理条件,基于历史数据,运用人工智能、机器学习等方法,发展高原预报预测模式,研究气候变化和对人们生产生活产生的影响。

3.3 在气象服务方面应用。所有气象资料都是为气象服务这个功能所准备,气象服务可以概括为政府决策服务、公众气象服务,专业气象服务等。近年来,各种气象服务资料的发布模式由单一的电视及广播逐渐向互联网转型,人们对服务产品要求也越来越精细化。数据挖掘技术在气象服务方面研究也在推进,例如利用大数

据分析不同的气象条件下,人们对各种商品的购买量规律,使商家根据结果做出商品生产销售的策略,利用气象大数据分析,为城市建设、农业生产、防灾救灾等提供决策服务,利用问卷调查,python网络爬虫等来获取公众对气象服务的需求资料,使气象部门可以制作更有针对性、更让人喜闻乐见的服务产品。数据挖掘技术在气象服务方面的应用直接展示了数据之美,服务形式和效果更加进步。

3.4 数据挖掘技术在气象业务中的应用展望。首先基于分布式存储和计算的原理,构建更优化的国省气象大数据平台,实现对多源气象数据资料的高效管理和加工运用,打破行业信息壁垒,整个诸如林业、水利、地震等部门的数据。同时通过建立国省大数据平台,实现数据的互联互通、互为备份、互为服务,为进一步挖掘气象数据价值做好准备。其次基于数据挖掘技术,提高气象数据的基础支撑能力,为诸如MICAPS、CIPAS、一体化业务平台提供直接支撑和创新方式,同时建立高效便捷的气象预报预测和服务产品制作模式。最后基于数据挖掘、多源数据分析、智能算法等,为精细精准预警预报、防灾减灾、工农业服务决策研判以及跨行业的数据挖掘、共享提供数据、算法、软硬件支持,为构建智慧气象发挥积极作用。

### 4 结语

气象数据大部分都是动态改变的,借助数据挖掘技术能够聚类分析与定量检测气象数据,从而有利于准确地预测气象预报。在气象业务领域,借助数据挖掘技术重点分析数据处理需要与气象业务需要,从而实现数据挖掘方法的优化,最终提高气象服务能力。

### [参考文献]

[1]郭伍民.数据挖掘技术在气象信息存储中的应用[J].信息记录材料,2019,20(06):179-180.

[2]原久淞,吴杨,崔福涛,等.数据挖掘技术在气象信息存储过程中的应用[J].南方农机,2017,48(10):109.

[3]张德龙,张立,苏和.数据挖掘技术在气象信息存储中的应用[J].内蒙古气象,2008,(06):34-35.