

数据挖掘技术在气象预报研究中的应用策略研究

凌天琦

(江苏航运职业技术学院, 江苏 南通 226010)

摘要: 由于气象预测预报受到多种变化因素的制约与影响, 因此很多时候会影响气象研究人员对天气的判断, 有时气象预测与预报会出现偏差甚至错误。基于此, 准确且科学的数据挖掘方法对于研究复杂且多变的气象具有深远的意义。研究表明, 在先进的数据挖掘技术与智能计算的辅助下, 能够切实提高气象专家预测预报气象的精准度, 同时还能提高气象研究人员的预测能力。本文立足于数据挖掘技术, 对其在气象预报中的应用进行了深入探讨, 最后对数据挖掘技术应用于气象预报中在未来可能遇到的困难进行了总结, 以期能够带给气象专家更多的思考与借鉴。

关键词: 数据挖掘技术; 气象预报; 应用策略

气象预测与预报, 需要涉及的范围广、数据多且杂, 气象研究人员不仅要熟知天气学、统计学与气象学等多门学科的原理, 更要具备扎实的气象数据分析能力以及独特的抽象思维。因此, 从全世界的视角来看, 气象预测与预报极具挑战性, 当然对于社会、人类来说其重要性也不言而喻。如何借助先进的数据挖掘技术, 在智能计算的辅助下对收集好的初级气象数据信息进行分析与整合, 提升天气预报的准确率是当今时代气象研究人员与气象专家们更为关注的热点话题。对气象预报中信息处理进行深入研究, 为今后数据挖掘技术的顺利应用奠定了坚实的基础。

一、数据挖掘的概念与特点

关于数据挖掘的定义, 迄今为止专家依然没有给出确切的答案。从根本上来讲, 数据挖掘指的是从大量排列无章、繁杂琐碎的信息数据中挖掘有用信息, 发现知识的细致化过程。从整体出发, 数据挖掘可以从广义与狭义角度来进行分析。

首先从狭义角度来讲, 数据挖掘技术的出现是商业信息处理技术的一个新的变革, 它主要挖掘的对象是商业数据库。通过对库中的大量商业数据进行抽取、分析、转换以及模块化处理, 最终得到对商业用户在做重大决策的时候起到关键性主导与影响作用的数据。其次, 从广义的角度来讲, 主要针对的是普通用户, 更侧重于从大型的数据库中挖掘有用的知识信息。这种行为对于很多用户来讲具有隐秘性特点, 虽然他们事先不知道, 但是确实是存在的。例如在浏览短视频的时候, 我们有时候会惊奇地发现很多短视频与我们刚才刷到的视频题材、类型或者内容都大同小异, 这就是数据挖掘技术在起作用。它会通过分析用户的浏览习惯来有目的的推送一些相关内容, 来达到精准营销的目的, 这就是数据挖掘的真实作用所在。

从以上模糊的定义, 我们可以发现其实针对数据挖掘可以总结出以下三个主要特点:

其一, 大量数据。数据挖掘技术其最显著的特点就表现为其处理的数据数量是非常庞大且惊人的, 这是与传统的分析方法相比最突出的优势。也正是因为如此, 数据挖掘技术才会被应用于更多的行业领域中, 发挥着其独特的作用。此外, 针对于数据挖掘算法, 尤其要注意算法时间, 一旦过于复杂或者不好操作, 那么将会大大降低处理数据的数量, 且数据的规模也将受到严重限制。

其二, 未知的但是有用的规律。显而易见, 在面对如此庞大的数据信息时, 其处理技术一定是有规律可循的, 并且通过数据挖掘出来的规律或者模式其用肉眼一定是看不到的。更重要的是这些规律与模式不管从广义的商业角度还是从狭义的用户角度来讲, 其作用都是独一无二且意义重大的。

其三, 数据挖掘是一个过程。数据挖掘其更主要阐述的是一个数据通过理解、分析、整合、评估、部署等系统的过程, 它并不是一个神奇的黑盒子, 只要将数据放进去就可以自动出来结果或者呈现令人惊奇的规律。这其中, 数据挖掘人员的分析能力与理解能力将对数据挖掘的准确性与实效性产生至关重要的影响。

二、数据挖掘技术在气象预报研究中常用的方法

(一) 决策树

在数据挖掘算法中, 决策树具有比较重要的地位, 它类似于树状结构, 主要是进行分类预测的方式来实现数据源挖掘。这种树状的每一个非终结点都代表了一个测试点。不同的测试结果所选择的测试结果各不相同。在选择某个分支的时候, 是由根节点, 从上到下一点点攀升, 一直到叶子的节点停下。即使放到现在, 决策树算法仍然是使用频率较高的几大分类算法之一, 在使用的过程中操作简单、精度较高且容易理解, 具有尤为广泛的使用用途。

(二) 遗传算法

遗传算法的来源主要是模仿自然界中的生物遗传进化规律, 通过自然选择以及生物遗传机制的搜索算法。该种算法是从全局视角下衍生出来的一种随机搜索算法。在开发此种算法的时候, 达尔文的进化论以及孟德尔遗传学理论为此种算法带来了较大的启发。使用此种方法的步骤如下: 首先对需要解决的问题进行参数编码, 然后对种群进行初始化, 紧接着, 根据个体的适应度进行计算, 最后选取与适应度相符合的数值来选择合适的个体。除此之外, 还需要根据交叉以及变异的概率进行交叉操作以及变异操作。这种算法适用于其他算法难以解决的复杂性问题。

(三) 聚类分析

该种算法主要是将数据按照某种原则汇聚成不同的类比, 进而使这些数据间的相似程度比较低, 数据内部的相似度比较高。比较常见的聚类方法主要有划分、层次、密度、网格以及模型。聚类分析主要是用来发现相关性较强的数据组, 相反, 孤立点检测则是发现与对象关联性不强的对象。聚类与孤立点分析是两个相对独立的过程。具体步骤方法就是将所有的对象进行聚类, 然后根据评估对象聚集、相似的程度进行距离。对于基原型的聚类, 可以用对象到它中心的距离进行衡量。

(四) 人工神经网络

人工神经网络是现阶段数据挖掘技术之中出现的一种全新的分类方法, 它借鉴了生物的神经元的特点。神经网络是一种非线性的自适应系统, 它由大量的基础神经元单位构建, 主要的功能有处理、激活、输出、链接、传递、工作等不同的方面。其模型的建立是由最开始的输入数据决定的。人工神经网络的工作特点, 首先要输入大量的蕴含着系统过去特征的数据, 然后通过内部的学习、加工、训练, 得到最后的模型。模型内部蕴含了较强

的关联性数据。

三、数据挖掘技术在气象预报研究中的应用

(一) 数据挖掘技术在气象预报中的应用

经常收看天气预报的人应该清楚,对天气的预测与预报通常会分为三个部分,短时、短期以及中期预报。由于对不同的组成部分,其预测的内容与时间不尽相同,因此三个部分所用到的预测方法也是有所差别的。首先针对于短时气象预报,气象人员可以采用的是现代化的探测手段。而到了短期,相对于短时,其数据量在不断扩充与增长,这时候的预测手法建议结合天气学、统计学的相关理论知识,加之先进的数据挖掘技术,来达到诊断分析气象的目的。最后针对于中期气象预测与预报,需要运用到的知识更为复杂且专业,通常需要气象研究人员熟练掌握与动力学、数值预报与天气学相关的知识与技能,进而达到气象预测的目的。

基于以上针对不同的时期所采用的不同的分析方法,可以总结出由于气象预测主要针对的主体是庞大的数据信息,在数据的基础上,通过整理、分析与总结得到最为有价值的气象信息,进而为提高气象预测预报的准确性奠定坚实的基础。

(二) 数据挖掘技术在气候预测中的应用

气候预测主要指的是气象人员根据气象数据对天气未来的变化进行长期预测,例如气温变化、降雨量、冷暖变化以及旱涝情况等。同样,进行气候预测也面临着大量的资料数据。气象研究人员可以通过统计预报等方法来总结得出最终的结论,这一方法从数据挖掘的整体方向来看其应该是重要的组成部分。举一个很简单的例子,奇艺普分析法是数据挖掘中重要的组成方法之一,李永华便是运用这种方法对样本进行了分量重建,通过构建BP神经网络多步预测模型,进而对重庆沙坪坝站夏季的降水量进行了预测与分析,从结果来看,其预测取得了一定的突破与良效。

(三) 数据挖掘技术在气象灾害预测中的应用

每年,我国都会发生各种各样的气象灾害,例如洪水、台风等,这些气象灾害对人民群众、社会稳定带来的消极影响必须引起气象部门的高度重视。不得不说,精准且及时的气象灾害预测与预报能够极大地降低灾害的发生率,不仅有利于人民群众安居乐业,更能对社会稳定、国家长治久安产生积极的意义。通常,对气象灾害的预测用到的方法主要是以定量遥感技术为主,辅助气象研究专家丰富的动力学知识,进而可以实现对气象灾害的临近预报。然而,由于气象灾害出现的次数不多,往往都是以个案的形式出现,因此,为气象人员对气象灾害的预测与预报带来了一定的难度。但是考虑到这项研究的重要性,势必需要气象研究人员掌握先进的数据挖掘技术,有目的、有针对性地众多大气与海洋因子中挖掘出干旱因子、台风因子等,进而为相关部门部署灾害预防与治理提供先机,进一步保障人民群众与社会的稳定、长远发展。

四、数据挖掘技术在气象预报研究中的应用策略

(一) 构建数据信息库

要想使得数据挖掘技术在气象预报中的应用更明显,更广泛,首先构建数据信息库是首要任务,以便充分发挥数据信息库的强大功能,对原始数据进行分类,最终实现通过先进的数据挖掘技术来对数据进行有机整合的目标。针对于气象预报中的数据信息库,它主要的职责是通过从全世界的气象数据信息中提炼与挖掘有用的天气预报数据,加之气象专家的细致分析,最终得到对广大人民群众都较为有利且具有参考价值的气象预报信息。数据信息库能够为气象工作人员准确且客观地对气象进行分析提供丰富

的数据支撑,同时还能简化气象专家的工作流程,为他们的日常工作带来极大的便利,最重要的是能够满足广大气象专家的工作需要。

(二) 完善气象决策服务过程

在气象预测中,预报的准确性是非常重要的,预测结果不管是对生活生产还是国家发展都具有不可估量的价值与作用。因此,提高气象预报的准确性是如今越来越多专家与相关人员较为关注的话题。基于此,必须完善有利于提高气象预报准确性的决策服务模式,凸显出气象决策服务的科学性、实效性与技术性。通常来说,气象决策服务的过程一般包括以下两个部分。第一部分为提出问题。这一过程主要是实时了解当地的天气变化规律,为最终的气象决策奠定良好的数据基础。第二部分为数据预处理。为了切实凸显大数据的科学性与完整性,在对大数据进行挖掘的过程中,需要气象工作者对数据进行进一步的优化处理,进而满足不同用户对气象数据的多元化需求。

(三) 加强各部门之间的协调工作

要想确保气象预测预报工作能够顺利展开,并且真正确保工作落实到位,保障气象预报信息的准确性,就需要加强各部门之间的配合与联系,保障各个部门之间的信息无障碍交流,最终共同有效开展气象预报工作。具体做法为,可以在相关部门内部建设信息交流平台,便于不同部门之间能够平等的互换与气象有关的信息内容,为今后数据挖掘技术的合理运用提供较为真实、客观的信息支撑。同时,加强各个部门之前的频繁沟通交流,有利于管理者从宏观角度把握各个部门的工作实际情况,从而能够及时向气象部门提供有用的信息,最终减少由于信息互通有误而给气象预报的数据挖掘工作造成的负面影响。另外,为了调动每个部门人员配合数据挖掘工作人员工作的积极性与主动性,需要制定出与此相关的各项规章制度来严格约束部门人员的行为,规范工作流程,以确保数据挖掘技术切实能够与气象预报工作有效融合在一起,为气象预报工作的准确性与及时性提供更坚实的人力保障与制度保障。

五、关于数据挖掘技术在气象预报研究中应用的展望

传统气象预报的方法在短期之内已经无法在准确性、实用性、及时性方面取得较为显著的飞跃。随着信息技术的飞速发展,大数据挖掘技术在气象预报中的应用优势一定会凸显并且将成为今后气象预报工作前进的主要方向。但是,与之相关的问题依旧不能忽视,依托于计算机的数值播报方法其准确率与运行效率方面还有待于提升。另外,和国外发达国家相比,虽然国内针对数据挖掘技术应用于气象预报研究中的应用课题信息一直在不断更新,且已经有了突破性的进展,但是不管是在理论还是实践层面依然表现出来不足之处,有待于深究的细节仍然很多,如怎样加快学习速度,减少计算机学习时间等。希望有更多的专家能够为这一课题的研究提供新思路与新方法。

参考文献:

- [1] 贾志明,王东峰,程智.数据挖掘技术在气象预报研究中的应用[J].黑龙江科学,2020,11(8):34-35.
- [2] 王换换,王建玲.分析气象预报研究中数据挖掘技术的应用[J].饮食科学,2017(8).
- [3] 崔志强.数据挖掘技术在气象预报研究中的运用[J].科技创新导报,2017,14(25):110,112.