

大数据助力中医药文化传承

王 钰 孙亚非 郭 盛

(沧州职业技术学院, 河北 沧州 061000)

摘要: 本文引入大数据技术来研究中医药疾病预防与预测的架构, 设计了观测器和控制器结构以实现观测和控制功能, 提升了系统的自适应能力。根据实时收集的中医药药物信息和中医药对疾病治疗的数据确定疾病的发生何传播规律, 自动输入疾病数据分析到预防诊断系统, 并对中医药疾病预测与预防系统中的所有节点以及在这些节点上运行的任务进行动态实时监控, 以实现智能高效的疾病预测。

关键词: 大数据技术, 中医药疾病预防与预测, 自适应能力, 预防诊断、动态实时监控

一、概述

中医药疾病预测是实现中医药疾病预防与预测系统的关键步骤, 预测结果直接影响中医药疾病控制的效果。国内外学者从不同角度对如何实时准确地预测疾病做了大量研究, 但这些关于中医药疾病预测的研究几乎都假定疾病信息是静态的, 而在实际中医药疾病预测中, 这些前提限制了研究的应用。本文所讨论的中医药疾病预测问题指的是短期疾病预测, 由于预测时间跨度较短, 疾病的变化呈现出越来越大的不确定性, 疾病变化的规律性变得不明显, 这是短期中医药疾病预测的难点^[1]。作为这些新概念和新方法之一, 大数据技术可以通过赋予计算系统一系列人工智能特性来降低其复杂性。开发人员不再需要考虑所有可能的系统行为, 而是给予计算系统一定程度的自由, 让系统能够对新情况做出智能响应。

本文将大数据技术的理念引入中医药疾病预防与预测系统, 设计了一个智能中医药疾病预防与预测系统模型。提出了一种自适应疾病预测方法, 并利用滑动周期机制来实现该方法, 以解决智能中医药疾病预防与预测系统中的动态疾病预测问题, 并提高系统的适应性。

(一) 基于大数据技术的中医药疾病预防与预测系统模型

中医药疾病预防与预测系统的目标功能是: 进行实时疾病预测, 有效防止疾病的发生和传播, 最终实现中医药对疾病预防的最优配置。

(二) 中医药疾病预防与预测系统模型基本框架

中医药疾病预防与预测系统模型基本框, 该模型由观测器和控制器所组成, 与 TCMBG (Model of Traditional Chinese Medicine Disease Prevention and Prediction System Based on Big Data Technology) 构成一个闭环控制结构。

该系统的目标函数能够有效地防止疾病的发生, 减少因搜集

数据而造成的制定疾病治疗方案的时间, 并最终实现中医药对疾病治疗和预防的最优方案。

作为观测和控制的模型对象, TCMBG 为中医药疾病预防与预测系统提供中医药对疾病治疗的信息。本文仅给出了 TCMBG 疾病预测系统的四个重要组成模块: 疾病监控、监控分析、中医药引导和效果评估。以下是对疾病观测器和自适应疾病预防控制器的主要介绍。

(三) 实时观测器

实时观测器用于从 TCMBG 收集疾病信息, 识别和描述当前的病人状态, 并预测病人的未来状态。实时观测器的组成和工作过程如下:

(1) 监控器: 它包括来自 TCMBG 监控模块的系统分析数据和来自其它疾病数据监控模块的实时采集数据。来自 TCMBG 的系统分析数据负责周期性地从 TCMBG 中检测信息, 及时反映疾病的动态化趋势, 为疾病预测提供数据样本。通过使其它疾病数据监控模块实时采集数据进行局部优化, 作为一个可扩展的接口, 随着数据来源的增多、数据量的增大、数据复杂度的加强, 控制器之间的数据交换将被纳入其中, 有助于实现全局优化。

(2) 日志文件: 监视器收集的历史疾病信息原始数据根据需要记录在日志文件中。

(3) 疾病预处理器: 对疾病治疗原始数据进行预处理, 融合各种疾病信息采集中的数据, 得到统一的数据格式。

(4) 疾病信息分析器: 分析实时采集到的疾病信息, 提取数据为疾病预处理器提供信息。

(5) 疾病预测器: 利用疾病预处理器和疾病信息分析器的数据预测交疾病的信息。本文将 Spark 大数据技术中的 RDD 滑动窗口和机器学习中线性回归预测方法相结合, 设计了一种自适应的中医药疾病预测方法。

(6) 聚合器：疾病信息分析器和疾病预测器的结果，以及来自疾病预处理器的历史疾病数据都被传递到聚合器中。聚合器还具有存储器，用于存储当前疾病预测值和过滤后的历史疾病数据值，形成数据向量网（每个向量都是结果），并将聚合的疾病预测数据发送给自适应疾病控制器。

（四）自适应疾病控制器

智能疾病自适应诱导控制器的任务是引导智能疾病预测自我组织系统的行为，只有在必要的时候，才影响系统。它的目标是对疾病进行实时动态的预测，尽量贴近实际的疾病状态，从而使提高中医药对疾病治疗的效果，并且减少疾病的发生。

二、大数据中医药疾病预测方法

使用大数据中医药疾病预测系统中的预测方法，可有效地提高中医药对疾病治疗方案的成功率。本模型引入 Spark 技术中 RDD 基于滑动窗口的预测方法来预测疾病。

基于 RDD（弹性分布式数据集）滑动窗口周期机制的线性回归疾病预测方法，是一种结合了大数据处理技术和统计预测模型的先进方法。以下是对该方法的详细阐述：

（一）基于 RDD 滑动窗口周期机制的线性回归疾病预测方法

（1）数据预处理：

① 从大数据源（如电子病历、公共卫生数据库等）中收集疾病相关数据。

② 使用 RDD 对数据进行分布式存储和处理，包括数据清洗、格式转换等步骤。

（2）滑动窗口数据划分：

① 根据指定的窗口大小和滑动步长，将数据划分为多个时间段。

② 每个时间段内的数据将用于训练线性回归模型或进行预测。

（3）线性回归模型训练与预测：

① 针对每个时间段内的数据，使用线性回归算法训练模型。

② 使用训练好的模型对下一个时间段内的疾病发病率或患病率进行预测。

（4）结果分析与可视化：

① 对预测结果进行分析，包括计算预测误差、评估模型性能等。

② 使用可视化工具（如图表、仪表盘等）展示预测结果，以便决策者更好地理解疾病趋势并制定相应的防控策略。

（二）注意事项与挑战

（1）数据质量与完整性：

确保收集到的数据是准确、完整且具有代表性的。数据质量问题可能导致模型性能下降或预测结果不准确。

（2）模型选择与调优：

根据数据的特性和预测需求选择合适的线性回归模型，并进行必要的调优操作以提高模型性能。

（3）计算资源与时间成本：

基于 RDD 的大数据处理需要足够的计算资源来支持分布式计算和存储。同时，滑动窗口机制和线性回归模型的训练也需要一定的时间成本。因此，在实际应用中需要权衡计算资源与时间成本之间的关系。

（4）隐私保护与数据安全：

在处理涉及个人隐私的敏感数据时，需要采取必要的隐私保护措施以确保数据安全。这包括数据加密、访问控制等技术手段以及遵守相关法律法规的要求。

综上所述，基于 RDD 滑动窗口周期机制的线性回归疾病预测方法是一种结合了大数据处理技术和统计预测模型的先进方法。它通过有效地划分数据时间段并利用线性回归算法进行预测，为疾病防控提供了有力的支持。然而，在实际应用中仍需注意数据质量、模型选择与调优、计算资源与时间成本以及隐私保护与安全等方面的问题。

三、小结

本文将大数据的设计思想引入中医药疾病预防和预测系统中，建立了基于大数据技术的中医药疾病预防与预测系统模型，利用该模型对疾病进行预测。结合中医药疾病预测的特点，设计了基于 RDD（弹性分布式数据集）滑动窗口周期机制的线性回归疾病预测方法来实时预测疾病信息。基于大数据技术的中医药疾病预防与预测系统模型具有良好的独立性和扩展性，并且极大地降低了管理复杂度。

本论文关于中医药疾病预防与预测问题，利用大数据的方法，进行了研究探讨。然而，工作并没有到此为止，如何更好地把提出的理论与工程实际相结合，以及如何把最新的理论应用到研究中去解决实际问题，充分考虑我国现行大数据对中医药技术研究方面客观存在的数据复杂度问题，研究更能反映实际中医药疾病预测模型也将成为今后的研究方向。

参考文献

[1] 王忠庆; 基于医疗大数据的临床检验医疗风险预警模型研究 [D]; 中国医科大学; 2021 年