ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research (ASSEHR), volume 300
2018 International Workshop on Education Reform and Social Sciences (ERSS 2018)

Practical Research on Improving Education Targeting based on Big Data Mining

Mingju Zhao[1, a], Wanyi Zhao[2, b]

[1] Jilin Institute of Chemical technology, Jilin 132022, China

[2] DongHua University, Shanghai,200051, China

[a] zmj@jlict.edu.cn, [b] 908718532@qq.com

Abstract. Promoting the traditional advantages of education, employment guidance and other work in colleges and universities and highly integrating information technology are required for the development of education, and the scientific application of education big data is indispensable for the realization of "individualized teaching". On the basis of demand analysis and data mining technology, practical exploration on improving education's targeting is carried out. Hidden rules are found in education big data. The significance of each stage is interpreted and analyzed from the application of data mining algorithm and the visualization stage of results. Taking the typical association rule mining and classification decision tree method as an example, this paper proposes an effective way to use the advantages of big data technology analysis to improve the education targeting of ideology and politics

Keywords: Big data; Data mining; the ideological education; Employment guidance; Targeting.

基于大数据挖掘提升教育靶向性的实践研究

赵明炬[1]，赵婉伊[2]

[1.] 吉林化工学院, 吉林 中国

[2.] 东华大学, 吉林 中国

摘 要：推动高校思想教育、就业指导等工作的传统优势与信息技术高度融合是教育发展所需，实现真正的“因材施教”离不开教育大数据的科学运用。以需求分析和数据挖掘技术为基础，提升教育的靶向性方面的实践探索，从教育大数据中寻找隐藏的规律，从数据挖掘算法的应用、结果的可视化呈现阶段, 诠释分析每个阶段的重要意义。以典型的关联规则挖掘和分类决策树方法为例, 提出如何利用大数据技术分析的优势, 提升思想政治教育靶向性的有效途径。

关键词：大数据；数据挖掘；思想教育；就业指导；靶向性。

1. 前言

习近平总书记在全国高校思想政治工作会议上强调，推动思想政治工作传统优势同信息技术高度融合，增强时代感和吸引力[1]。我们已经迎来了大数据融入日常生活的特殊时代，高校教育工作者要主动迎接挑战，顺应量化研究的新崛起，树立大数据思维的理念，多学科、多样化研究思想政治教育的内在机理，探索教育数据的预测功能，提升思想政治教育的实效性和针对性。数据挖掘就是从数据中寻找有价值的规律，通过数学算法和计算机技术的应用辅助决策。但是，大数据的数据量往往很大，传统的数据挖掘技术必须要通过改进算法、改革处理方案的框架等角度提高适用性。目前结合大数据的数据计算角度研究思想政治教育接受以宏观研究居多，如梁剑宏提出的以大数据思维审视思想政治教育环境的立论，而以数学和计算机学科开展的技术驱动思想政治教育的相关研究有待丰富。

Copyright © 2019, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

2. 教育大数据的核心运用需求分析

大数据最核心的应用在于预测。在高校的运行过程中，产生了大量的数据，比如各个高校使用的教务管理系统，记载了每名学生在校期间的选课、成绩等信息；学生工作部门使用的学生信息管理系统，记载学生的自然情况信息、奖惩信息、诚信记录、各类社会活动等数据；智慧校园建设使用的校园一卡通记载了学生在校园的消费情况、宿舍及图书馆等门禁系统记录、使用一卡通所产生的一切活动记录；财务系统包括了学生缴费记录；图书借阅系统记录了图书借阅记录；就业管理信息系统记录了学生的生源地、就业单位等数据；科研部门的科研管理系统记录了学生参与科研情况的所有记录。此外，高校还有一些部门自行使用的一些管理信息系统，以上各个系统产生的大量历史数据往往容易被遗忘和闲置，再次利用的概率较低。高校如果重视并利用好这些数据，将产生更多有价值的数据，对于辅助决策大有裨益。

2.1. “靶向性”与“精细化”区别

近几年，在思想政治教育领域，学界提出了精细化的概念，从管理学以及统计意义上讲，这与我们提出的思想政治教育靶向性有着根本的区别。前者着重强调的是个体行为，是一种教育理念和管理模式的变革和价值追求，基于对受教育对象的精准认知，强调的是“精”与“细”比如针对某个个体开展有针对性的思想政治教育。而靶向性指的是通过大数据分析与挖掘，找出数据之中隐藏的规律和趋势，基于各种教育运行过程中产生的历史数据，强调的是不同层次和分类的对象运行趋势进行预测，核心在于预测。从而为未来的思想政治教育工作的方向、方法指明方向。

2.2. 教育大数据的核心运用

教育靶向性的核心是在教育过程中提高目标群体教育接受的针对性。实现真正的“因材施教”离不开教育大数据。大数据时代的到来为高校开展靶向性思想政治教育工作迎来了更多的挑战。大数据的核心在于预测。可以理解成大数据是场景是问题，而数据挖掘是手段。大数据对于思想政治教育领域的改革带来的将是传统思维的颠覆、教育靶向性的标的准确性的提升[2]。除了在线学习，大数据可以用在学生服务等方面，将对教师教学、校园治理、教育决策、就业管理、学生成长等方方面面带来很大影响。会驱动整个思想政治教育领域的变革，使精准提升思想政治教育靶向性成为可能。

3. 以数据挖掘视角提升教育靶向性的实现方案

将大数据技术的优势融合到高校思想政治教育靶向性中去，在思想政治教育实施过程中运用数据挖掘技术，涉及到对数学学科和计算机学科的灵活运用。在实际运用中，教育数据挖掘向着可视化、行为预测等方向发展。尤其是实现预测，在教育接受的运行机理内融合发挥作用，在教育接受的各个阶段深入融合发挥作用。发挥思想政治教育接受传统优势和大数据有机融合，分析利弊，以科学的方法论指导实践工作。从确定教育大数据挖掘对象开始，数据挖掘过程还要经历教育大数据的准备、数据的各种数学预处理、挖掘算法的选用、结果的可视化呈现、预测与决策这几个阶段,整个过程可能需要重复多次[3]。如图 1 所示。与一般的数据挖掘相比，教育数据挖掘具有多学科、人本化、多语义等特点。从工作流程来看，是基于思想政治教育环境下发现知识的循环过程。

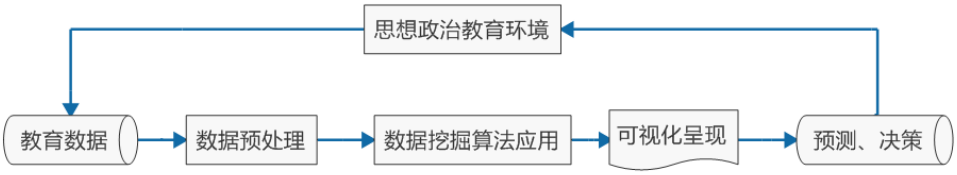

图1 教育大数据的数据挖掘

3.1. 以教育大数据为基础

随着网络化、互联网+以及物联网等热词的诞生，揭示了数字化智能时代的发展规律。伴随着中国高等教育的发展，教育现代化、信息化成为了时代必需，数字化校园、智能校园、云平台建设成为各个高校进行校园建设的必由之路。如何有效整合之前孤立运行的单个管理信息系统是目前数字化校园建设的瓶颈，比如教务管理系统、学生信息管理系统、校园一卡通、图书借阅系统、学生收费系统、科研管理系统等分散运行的数据，此外，一些职能部门还有很多独立运行辅助决策的数据库，这些都促使了教育数据的高速增长，面对日益庞大的历史教育大数据，多数被闲置和抛弃，并没有二次利用。如果我们运用数据挖掘技术，从学校的各个管理信息系统中保留的历史数据中提取对思想政治教育有关联的数据信息，实现对历史数据的二次利用，用以改善和提高思想政治教育靶向性，无疑是将原始的教育数据转变成有用知识的过程。

3.2. 选择适合的数据挖掘方案

数据挖掘的目标是采用与预测用途紧密相关的数据对象，选用科学适用的挖掘算法，从数据集中找出可能最初未知但最终可理解的知识，并可视化表达出来。因此，算法的选用在数据挖掘上起着至关重要的作用[3]。目前，在众多数据挖掘算法中，经过分析比较，适用于教育大数据处理的算法有两个，一是关联规则挖掘算法，二是建立决策分类树。两种方法各有特点和不足，在不同的应用方向上要科学应用。

3.2.1 关联规则挖掘

关联规则的挖掘是找到多个事件之间具有某种层面上的依赖性或者相关性，用于预测未来的行为模式。关联规则挖掘在众多的关联规则中找寻具有统计意义的数据上的关联关系。目前在诸多算法中，主要围绕 Agrawal 提出的 Apriori 算法展开的，它是基于支持度—置信度框架的关联规则挖掘算法，目前的关联规则挖掘通常采用支持度-置信度-兴趣度模型。首先支持度作为阀值对教育数据源搜索，对结果进行剪枝等清洗处理，Apriori 算法使用逐层搜索的迭代方法，逐次扫描搜索更小的项集，直到不能找到更小的频繁项集，其核心方法是基于频集理论的递推方法[3]。目前也有了一些改进算法以减少扫描数据量，减少了候选项集的产生，降低了算法的时间和空间复杂度。关联规则的发现是数据挖掘中一个非常活跃的研究领域，它的应用也早就影响着我们的日常生活。在淘宝、京东商场等网络购物商城中，根据用户的浏览行为推荐个性化商品就是典型例子。在教育大数据领域，采用关联规则挖掘算法，可用于发现教育大数据中不同属性之间的关联规则，通过这些规则预测未来的行为模式，比如学籍地域、学生的平均学分绩点对就业地点的影响，多运用于大学生就业指导。

3.2.2 分类决策树方法

决策树方法是数据挖掘的核心技术算法之一。它的基本思想是采用信息论中的信息增益作为决策属性分类判别能力的度量，进行决策节点属性的选择。通过将对象数据进行分类和清洗，找出与决策相关的的信息。常用于预测模型和辅助决策系统[4]。目前国际上最有影响的是由 Quinlan 研制的 ID3 决策树生成算法，它是基于奥卡姆剃刀原理的，即用尽量用较少的东西

做更多的事。决策树提供了一种展示类似在 If-then 这类规则的方法（即什么条件下会得到什么值）。所谓决策树是一个树结构，既可以是二叉树也可以是非二叉树,每一个非叶节点对应一个非类别属性,每个分支代表属性值的可能值,决策树上的的每个叶节点存放一个类别。决策的过程就是从根到达叶子节点的路径，叶子节点存放的类别作为决策的结果[5]。

属性选择度量是构造决策树的关键性内容，是一种选择分裂准则，ID3 算法使用的就是基于信息增益 Gain 的选择属性方法。期望信息越小，信息增益就越大，纯度就越高。通过计算每个属性的信息增益,选择分裂后信息增益最大的属性作为当前节点的分支属性进行分裂。采用自顶向下的贪婪搜索遍历可能的决策空间。依此改进的 C4.5 算法采用相同的基本工作流，但采用了比信息增益更好的信息增益率 GainRatio 作为属性选择的度量标准[6]。

决策树模型只需扫描几次数据库即可建立，更适用大量数据的处理。ID3 决策树算法的每个分支都对应一个易于理解的分类规则，分类速度较快,准确率较高。但也存在着不足：虽然实际数据的属性值多为连续值。比如成绩分数，但是算法最初假定为离散值，忽略差异性；对错误数据即噪声较为敏感；需多次扫描； C4.5 算法针对其不足进行了改进。用信息增益率来选择属性,克服了用信息增益选择属性时偏向选择取值多的属性的不足；在决策树构造过程中剪枝以提高抗噪声能力[5]；能够处理非离散和不完整的数据。产生的规则更易于可视化理解。但是无论 ID3 还是 C4.5，都存在训练集达到内存驻留极限值时程序无法运行的缺点。

3.2.3 选择合适的挖掘算法，注意互相结合。

根据不同的应用场景设计，可以考虑不同的数据挖掘算法，将数据挖掘技术的优势运用到思想教育工作中去。通过学校近几年的全样本数据，采用 Apriori 关联规则挖掘、基于 C4.5 算法建立决策分类树，挖掘出思想教育及意识形态领域的若干规律。甚至可以考虑通过将两种数据挖掘算法融合在一起,形成一种基于关联规则的决策树算法。发挥运用数据分析和数据挖掘技术实现大数据核心的预测功能，在思想教育的各个阶段深入融合发挥作用，以科学的方法论指导实践工作。运用 SQL Server 等工具辅助决策；因此，在选择数据挖掘算法为思想政治教育靶向性服务的时候，首先从实际的工作需求出发，比如进行问题群体关注、大学生就业指导、学风建设等等。结合需求分析的要求，考量能够获取的历史数据资源量，尽可能多的将关联数据库纳入处理范围，综合考虑适用于哪种数据挖掘算法；其次，在数据的清洗过程中，要明晰主次，抓大放小；最后，尽可能综合多种技术和算法及客观环境的影响因素，调整影响决策的各种主客观因素，尽可能得出全面和更具说服力的结果。

3.3. 核心挑战在于结果的可视化呈现

通过以上数据的准备、分析以及演算，目的要为思想政治教育靶向性服务，因此，数据挖掘最后结果的运用是要达到可视化要求，这是所有技术手段为实际工作服务的基本要求，也是信息处理系统的基本技术要求[7]。无论采用哪种数据挖掘算法，比如关联规则或者分类决策树方法，把大量的抽象结果数据与实际工作相结合，把握规律既要抓大放小，也不能忽略数据分析中的潜在因素。高校的教育工作者要把思想政治教育贯穿到教育全过程，这个过程中需要数据分析作为依据和支撑，对于高校的决策者来说，关注的是政策以及制度的实效性，而对于数据处理的技术细节是参与数据处理以及分析环节的管理技术人员的任务，对于任何数学方法和计算机处理程序，最终形成的一定是便于管理者参与和体验的交互手段，这个手段就决定了我们进行数据挖掘的结果要具备可视化。具有友好的 UI 设计效果。所以，可视化呈现可以理解为，为决策者提供易于理解和使用的表现形式。一般体现为管理信息系统的分析决策系统模块，或者通过单次的数据挖掘运算形成的，包含数据分析过程的辅助决策建议文稿。

4. 结论

以往教育靶向性大多缺乏大数据分析的支撑，以经验主义和小规模数据分析居多，大规模发散和碎片化数据难以科学梳理，通过与传统方法的结合，尊崇教育规律。一方面尊重个人隐私，树立数据隐私的处理风险意识，依法依规加强数据的安全保护。另一方面强调数据的模糊性，而不强调个体数据的分析。通过大数据挖掘和分析提供教育问题早期预测系统、可控教育系统、可视的可量化的新手段，改变教育接受的效果，优化教育的方式方法。充分利用历史数据潜在的价值，优化育人环境，完善管理体制，占领思想教育、就业指导等大数据阵地。

致谢

感谢吉林省教育厅《基于大数据挖掘技术推进就业指导精细化研究》、吉林省教科办重点课题《大数据背景下运用数据挖掘技术推进就业指导精细化研究》、吉林省教科办《大数据挖掘提升思想政治教育靶向性实践研究》的基金。

References

[1]. Shen wenjie. Realistic path of ideological construction in colleges and universities in China [J]. Theoretical research on marxism in colleges and universities, Vol.2(2016) No.04, p.35-41.

[2]. Su ling. Education reform "targeted therapy" comes in the era of big data [J]. Liaoning education,Vol 12(2016),p.5-9.

[3]. Yi weiguo: Research and application of prediction method based on association rules and decision tree [D]. Dalian maritime university,2012. p.6.

[4]. Li haiqiong. Research and application of data mining technology in the employment assistance decision analysis system for college students in liaoning province [D]. Shenyang university of technology,2009. p.12.

[5]. Tian xin. Research review on decision tree algorithm [J]. Modern marketing (next ten-day issue), Vol.01(2017), p.36.

[6]. Zhang zhenmei. Research on association rule mining algorithm [D]. Shandong university of science and technology,2005. p.11.

[7]. Zhang haixiao. Research on the application of data mining classification technology in college teaching [D]. Guangdong university of technology,2005. p.16.