

“互联网 + 竞赛” 驱动下的研究生教育模式研究

丁兆云, 王 晖, 李 沛, 张 鑫

(国防科技大学 信息系统与管理学院, 湖南 长沙 410073)

摘要: 课后大作业为提升研究生实践动手能力的重要途径之一。国内研究生教育助教机制不够健全导致研究生教师工作量大, 教师重视课堂教学而忽略大作业; 另一方面, 学生对大作业的兴趣度不高, 完成大作业的成就感不强, 很难达到提高研究生实践动手能力的效果。以“数据挖掘”研究生课程为例, 利用互联网提供的数据挖掘竞赛平台, 将竞赛平台引入到课堂教学中, 以竞赛平台为牵引, 驱动研究生的课后大作业, 重点研究了引入竞赛平台的研究生成绩评定模型与互联网竞赛团队组合模式, 最大化地发挥学生参与互联网竞赛的积极性与提高学生实践动手的能力。

关键词: 互联网 + ; 研究生大作业; 数据挖掘; 数据挖掘竞赛

中图分类号: G643 **文献标识码:** A **文章编号:** 1672 - 8874 (2017) 03 - 0018 - 05

A Research on Postgraduate Education Model on Internet Plus Competitions

DING Zhao-yun, WANG Hui, LI Pei, ZHANG Xin

(School of Information System and Management, National University of
Defense Technology, Changsha 410073, China)

Abstract: The appropriate assignment is one of the important ways to improve the practical ability of graduate students. The workload of graduate teachers is heavy for lack of necessary assistant mechanism. So, teachers pay more attention to classroom teaching and ignore assignments. Moreover, students are not interested in assignments, and the sense of achievement is not strong. In this paper, the graduate course of “Data Mining” is taken as an example. Using the data mining competition platform provided by the Internet, the competition platforms are introduced into the classroom teaching, and the competition platforms are used as the traction to drive the graduate’s assignments. Graduate student performance evaluation model and Internet competition team combination model are studied in this paper, in order to maximize the students’ enthusiasm to participate in the Internet competition and improve their practical ability.

Key words: Internet + ; undergraduates’ assignment; data mining; data mining competition

一、传统大作业存在问题

2015年的两会上, “互联网 + ” 被纳入国家战

略, 在此基础上, “互联网 + 教育” 的模式逐渐引起相关学者的重视。教师借助互联网, 以微课、慕课、翻转课堂、手机课堂等新颖的方式进行授课。研究生课程通常由课堂教学和课后大作业

实践两部分组成,课堂教学能够提高研究生的理论学术能力,课后大作业能够提高研究生的实践动手能力。

尽管“互联网+教育”的模式在一定程度上提高了研究生的课堂参与积极性,但是,该模式还不能够很好的解决研究生对课后大作业的参与积极性这个问题。目前,国内研究生对大作业参与的积极性不够高,完成的质量相对也较差。一组山西大学的问卷调查数据显示,大学生不做作业的情况普遍存在,从来不做作业的占到了39%,而每次都完成作业的只占到了17%,有1/3的大学生认为大学里没有必要布置作业,57%的学生认为大学作业是必不可少的,其他学生对作业的必要性持有可有可无的态度^[1]。在这一方面,距离国外研究生教育还存在一定差距,究其原因,主要存在以下几个问题:

1. 国内研究生教育助教机制不够健全,教师被动的布置研究生大作业。目前,国内研究生课堂主要由一名教师独立授课,课后实践大作业也有该老师布置,尽管有些高等院校安排了研究生助教,但助教大部分是兼职或者由博士研究生来完成,助教实质性参与度不高。因此,授课与研究生大作业的工作实质都由一名研究生授课老师完成,从而导致授课老师的工作量巨大。授课老师通常把大部分精力都花费在课堂教学方面,而不重视课后大作业,从而导致授课老师布置的课后大作业本身质量不高,甚至有些大作业被用来应付多届研究生,大作业题目较老,已经完全脱离时代,陈旧的题目很难提高学生参与的积极性。另一方面,授课老师也很难花费大量时间来评阅学生的大作业,从而导致一些完成质量较高的大作业容易被授课教师漏掉,大作业的评分差异性不大,从而更加影响部分花费大量时间在大作业上的学生积极性,形成恶性循环。

南京大学徐萍等人在《学位与研究生研究》中总结了我国研究生助教中的五个问题^[2]:(1)认识不到位;(2)部分研究生对助教兴趣不大;(3)教师滥用研究生助教制度;(4)经费缺乏保障;(5)管理无约束。

2. 学生对大作业的兴趣度不高,完成大作业的成就感不强。授课老师布置的大作业质量本身不高,导致研究生也很难有兴趣来完成大作业,研究生通常为了拿到平时成绩敷衍了事的完成大作业,从而导致大作业的质量不高,没有真正达

到提高实践动手能力的效果。

福建师范大学教育学院王志霞等人在《比较教育研究》中总结了我国大学生在作业上的五个问题^[3]:(1)作业的教育价值未被充分认识;(2)作业的数量少、频次低;(3)作业的类型单一;(4)作业抄袭现象普遍;(5)缺乏有效的作业反馈机制。

因此,目前迫切需要研究一种新型的研究生大作业模式,适应新时代的发展。

二、“互联网+竞赛”的大作业更具有时代先进性

如何建立一种耗时少,质量高的研究生大作业模式已成为目前研究生教育的一个难题。随着互联网技术的发展,这种大作业模式已逐渐成为可能性。以“数据挖掘”这门研究生课程为例,目前互联网上已存在多种数据挖掘竞赛平台,比如国外的kaggle竞赛平台、国内的天池大数据竞赛平台、国内的数泉竞赛平台等大数据竞赛平台,另外还包括各种数据挖掘国际会议举办的竞赛,比如kddCup等,这些竞赛平台提供的题目通常都是主流技术需求,且时效性非常高。国内已存在部分高年级的研究生参加此类数据挖掘竞赛,且取得了较好的成绩。另外,现在各大企业招聘应届研究生的时候,参加竞赛获得的名次证书也起着重要作用。

将竞赛平台引入到课堂教学中,以竞赛平台为牵引,驱动研究生的课后大作业,可以让学生尽可能早地了解国内外主流技术需求,也可以让学生在课堂上学以致用,利用所学的理论解决当前的迫切问题。另外,由于这种竞赛平台提供排名机制,学生很快就能够根据排名了解自己所学知识的不足,在排名机制的驱动下,学生通常会自学新知识,以积极性的学习态度来提高自己的排名。同时,研究生就业时需要的竞赛证书也迫使学生对竞赛具有浓厚的兴趣,不仅能够学到知识,而且好的排名证书在就业时也起着重要作用。

因此,引入互联网竞赛模式,来驱动研究生的课后大作业,不仅能够提高学生的主动积极性和成就感,而且真正能够达到通过实践来巩固所学知识的效果,将知识转换为能力;同时,教师可以直接利用竞赛题目,通过竞赛提交次数来判断学生是否在大作业上花了功夫,通过排名来判

断学生的实际学习效果,从而节约了传统的批改作业时间,在一定程度上也降低了教师的工作量,起到助教的作用;进一步地,竞赛平台的题目内容超过了传统的课本题目,更加具有新颖性和时代前沿性。

三、“互联网+竞赛”的大作业面临新问题

如何在课堂上利用好竞赛平台,以最大的可能性来驱动学生的参与积极性,还存在如下几个问题:

(1) 评分的标准难以确定,研究生课程通常具有多个互联网竞赛平台,每个竞赛平台包含多个竞赛题目,使得学生所选题目分布在多个平台、多个题目上,互联网竞赛平台的差异性导致综合衡量学生不同平台、不同题目上的成绩存在困难性。同时,互联网竞赛平台自身对每个参赛选手提供的排名具有时效性,参赛选手每天可以提交自己的参赛结果,参赛选手的排名每天都存在差异性,为了激发学生力争向上的积极性,在一个好的排名基础下,更好的成绩应该偏向于每天都争取进步的学生。

(2) 如何让学生组队,既能够达到团结协作的目的,又尽可能降低作业抄袭的可能性。互联网竞赛平台通常都由一个团队协作来完成一个题目,团队成员数量一定程度上影响作业质量,团队成员过少,则竞赛成员力量过于薄弱,题目的完成质量可能得不到保障;团队成员过多,部分成员可能存在浑水摸鱼的侥幸心理,不付出力量,但跟着其他成员获得较好成绩,严重影响其他学生的参与积极性。团队中的每个成员对竞赛题目的贡献也存在差异性,如果不能很好的区分差异性,表扬付出更多努力的学生,则在一定程度上影响团队协作的积极性。

(3) 如何融合竞赛平台大作业到课堂教学中,既能够提高教学质量,又能够促进学生取得更好的成绩。互联网竞赛大作业区别于传统的大作业,其时间周期贯穿整个教学过程,如何改变教学模式,才能够使学生尽可能早的适应竞赛模式。

针对上述问题,老师需要在三个方面有所加强:第一,如何综合评定学生大作业成绩;第二,如何鼓励学生组团参赛;第三,如何改进教学模式来适应互联网竞赛。

本文针对老师需要做的三个工作,主要研究了面向互联网竞赛平台的成绩评定模型、研究生团队组合模式,最大化地发挥学生参与互联网竞赛的积极性与提高学生实践动手的能力。另外,本文所涉及的大作业通常指研究生在课堂之外完成的作业,目的是为了锻炼研究生的实践动手能力,通常适合工科类专业课的大作业。

四、“互联网+竞赛”新问题解决方案

(一) 成绩评定模型

一门研究生课程通常对应多个互联网竞赛平台、多个竞赛题目,同时,学生自身研究背景 and 兴趣存在差异性,学生根据自身的特点选择不同的竞赛题目,从而导致学生的成绩将分布在不同平台、不同题目上。但一门课程的老师需要根据学生的竞赛排名来综合评定一个班级学生的成绩,因此,需要融合多个互联网竞赛平台、多个竞赛题目的成绩来综合评定每个参与学生的成绩。以研究生课程“数据挖掘”为例,学生可以根据自己的实际情况,选择各个平台的不同竞赛题目,由于选择的题目不一样,所得到的竞赛成绩标准也不一样。

表1给出了kaggle竞赛平台成绩排名示例。由表所示可知,互联网竞赛平台一般以团队为单位来排名,不同的竞赛平台得分值具有差异性,且通常都提供了排名变化情况以及最新一次提交结果时间。

为了综合异构平台、异构题目的成绩,首先采用概率统计模型 $f(x)$ 来分析各个竞赛平台、竞赛题目中本身所有参赛选手成绩排名的概率分布。参赛选手成绩排名的概率分布需要采用假设检验方法来建模,本文假设学生成绩分布大致拟合正态分布:

$$X \sim N(\mu, \sigma^2)$$

在此基础上,以每个竞赛平台、特定题目 j 的所有参赛选手竞赛成绩数据为统计抽样对象,来估计正态分布的期望 μ_j 和方差 σ_j 。

随后,采用 z 分数规范化对参加竞赛的每个学生成绩 v_i 进行规范化:

$$v'_i = \frac{v_i - \mu}{\sigma}$$

v'_i 表示规范化后的学生成绩,以此来对参与

不同竞赛平台、不同题目的学生成绩进行综合规范化,综合来衡量参赛学生的成绩。

表1 kaggle 竞赛平台成绩排名示例

排名	注册账号	得分	上一次提交排名	提交时间
1	HangYu	0.77127	53	2016-11-18
2	Dannye	0.75047	7	2016-11-19
3	NUDT_ZYDING	0.74858	11	2016-11-29
.....

另外,互联网竞赛平台自身对每个参赛选手提供的排名具有时效性,参赛选手每天可以提交自己的参赛结果,参赛选手的成绩排名每天都根据其提交结果而变化。为了激发学生力争向上的积极性,在一个好的排名基础下,更好的成绩应该偏向于每天都争取进步的学生,因此,本文在互联网竞赛的全生命周期范围内跟踪学生每天的成绩排名状况,采用成绩变化率来计算学生成绩的进步值,本文以“数据挖掘”这门研究生课程为例,竞赛平台主要包括国外的kaggle竞赛平台、国内的天池大数据竞赛平台、国内的数泉竞赛平台等,在竞赛的初赛、决赛等各个阶段,每天利用网络爬虫技术获取参赛选手的成绩。

(二) 竞赛团队组合模式

互联网竞赛中团队成员数量将一定程度上影响互联网竞赛的成绩,团队成员过少,则竞赛成员力量过于薄弱,题目的完成质量将不能够得到保障;团队成员过多,势必部分成员存在浑水摸鱼的侥幸心理,不付出力量,但跟着其他成员获得较好成绩,严重影响其他学生的参与积极性。本文采集各竞赛平台不同题目排名靠前的参赛团队数据,利用大数据分析技术,提取这些参赛团队频繁的队员数量,以此队员数量来指导学员的组队模式。

另外,为了最大化激发团队每个参与成员积极性,采用随机投票模型,类似PageRank算法的随机游走思想,每个参赛成员赋予一定的投票额度,根据其他组员在竞赛的投入度来决定投票份额,最后根据每个参赛成员手中新获得的投票额度来决定其投入度,以此来综合评价组内每个成员在竞赛中付出努力的程度,综合评价组内成员成绩,来激发每个成员的参赛积极性。

(三) 针对互联网竞赛的教学模式

互联网竞赛大作业区别于传统的大作业,其

时间周期贯穿整个教学过程。为了使学生尽可能早的适应竞赛模式,本文以研究生“数据挖掘”这门课程为案例,首先建立适应互联网竞赛的新型教学大纲,该教学大纲以知识点递进式为原则,首先讲解数据挖掘的基本算法,比如Apriori频繁项挖掘算法、贝叶斯分类算法、决策树分类算法、逻辑斯特分类算法、基于划分的距离算法、层次聚类算法等。这些基础算法约需要使用16个学时,在16个学时基础上,学生就对数据挖掘有个初步认识,即可以利用已学的知识来参加数据挖掘互联网竞赛。学生在参加竞赛过程中,将真正面临需要解决的实际问题,思考使用哪种数据挖掘算法最佳,此时学生将由被动的听课转变为主动的思考如何解决问题,在这个阶段,学生将发现数据挖掘基础算法在实际问题中可能还存在部分缺陷。针对学生竞赛中的问题,在后期的教学过程中,将进一步的讲解数据挖掘的高级算法,比如频繁序列挖掘、频繁结构挖掘、支持向量机分类算法、神经网络分类算法、谱聚类算法、EM聚类算法等。这些高级算法约使用12个学时,在这12个学时过程中,学生一边参与竞赛,一边学习新的算法,真正能够达到理论与实践的结合。通过这12个学时的理论学习和实践,部分学生将能够在互联网数据挖掘竞赛中获取优秀的成绩。随后,将这些优秀的竞赛成果引入到课堂,鼓励学生介绍有效的数据挖掘算法和其参赛经验,在已获得优秀成绩学生的经验介绍基础上,其他学生将能够进一步针对自己参赛题目,调整数据挖掘算法,使自己的竞赛成绩得到进一步提升。以此来激发每个学生的参赛积极性和课堂参与性,尽可能地让每个学生都能够学有所获。

五、世界一流大学大作业对比

卡内基·梅隆大学的斯克斯曼指出^[4],为了提高美国人的国际竞争力,作业应被视为建构学术规范和学生性格的一种方式,应被当作教育改革的一个严肃的承诺、一个重要的象征、一种关键的方法。哥伦比亚大学作业的倡导者康纳宣称^[5],布置作业在教会学生如何合理组织时间的同时,还提高了他们的责任心、独立性,以及对事物的判断力。他主张通过拓展和创造指向性的作业,提高学生的创造力,让学生学会“如何学习”。

美国课堂是师生积极互动的过程,学生要想在课堂上与知识渊博的教师和有备而来的同学进行思想碰撞,就一定要在上课前做足功课,花大量时间完成课前的准备性作业。同样,学生要在课外巩固、深化课堂所学知识,也要花大量时间完成课后作业。例如,哈佛大学的本科生每周上课时间约12小时^[6],但在课外学习、阅读、写作、准备家庭作业所花费的时间大约30个小时。杜克大学教育学教授Cooper指出^[7],作为作业的重要特性,作业数量是影响作业效用的一个关键因素。美国大学的课程不仅作业多,而且通常贯穿课程的全程。例如,美国斯坦福大学“计算机分子生物学”课程有7项作业,从期初到期末,均匀分布。同样,哥伦比亚大学“操作系统”课程也有6项作业,每项作业间隔时间为两周^[8]。

与本文非常相似的为美国在线作业系统,主要包括CAPA、Homework Service、OWL、Tychos、WebCT、WWWAssign、Blackboard、WebAssign、Mastering physics和Webwork等多种软件,在美国大学里已经有成千上万的注册用户。在线作业系统所提供的服务包括:能通过连接浏览器或因特网来进行评估;能通过特定用户名以及密码登入;能告知学生作业内容,并快速收集;能自动评价学生作业;能记录学生的得分情况,方便教师的随时查看。相对于人类有限的精力,智能型指导系统能够在学生完成作业甚至在训练过程中得到反馈,这种方式更直接、准确^[9]。

六、总结

本文以“互联网+”为基础,融合互联网竞赛平台,驱动研究生的课后大作业,针对竞赛平台的异构差异性,研究面向互联网竞赛平台的成绩评定模型,为合理的成绩评定提供理论指导;针对竞赛平台的团队协作性,研究成绩最大化的研究生团队组合模式,使得学生既能够取得好成绩,又能够规避学生之间的抄袭现象;针对竞赛平台的时代新颖性,研究新型的教学模式,使得

教学质量与竞赛成绩共同促进,形成良好的教学与大作业互促模式。

作者在2016年的研究生“数据挖掘”课程教学中,引入了互联网竞赛模式,在课堂教学中取得了良好的成绩:1组学生在天池大数据竞赛中的广州白云机场流量预测初赛中,连续两周排名第一;2组学生在天池新人挑战赛中均进入前十名;8组学生分别在kaggle、数泉竞赛平台中均进入前10%。

参考文献:

- [1] 任美霖. 大学作业情况调查研究——以山西大学教育学专业为例[J]. 经营管理者, 2013(1): 290.
- [2] 徐萍, 张菊芳. 高校研究生助教制度的现状、问题与对策[J]. 学位与研究生教育, 2009(3): 57-58.
- [3] 王志霞, 叶信治. 美国大学作业的特点及其对我国大学改进作业的启示[J]. 比较教育研究, 2015(6): 71-76.
- [4] Gill P B, Schlossman L S. A Nation at Rest: The American Way of Homework[J]. Educational Evaluation and Policy Analysis, 203(3): 319-337.
- [5] Connors N A. Homework: A New Direction [M]. Columbus: National Middle School Association, 1991: 10.
- [6] Light R J. Making the Most of College: Students Speak Their Minds[M]. Cambridge: Harvard University Press, 2001: 54.
- [7] Roderique T W. Homework: A Survey of Policies in the United States [J]. Journal of Learning Disabilities, 1994(8): 81-87.
- [8] Brender J R. Effects of Homework Completion on Test Scores in First and Second-Semester Spanish Courses at a University with Liberal Admissions [J]. The Education Resources Information Center, 1996(10): 1-10.
- [9] Bonham S W, Deardorff D L, Beichner R J. A Comparison of Student Performance Using Web and Paper-Based Homework in College-Level Physics [J]. Journal of Research in Science Teaching, 2003(10): 1050-1071.

(责任编辑: 胡志刚)