

# 开源模式下我国软件人才培养的挑战和对策

陈丽 赵斌 倪超 杨小虎

【摘要】随着软件技术尤其是互联网的蓬勃发展,开源越来越成为世界信息技术及产业发展的重要方向。然而,在日益繁荣的开源环境下,我国的开源人才无论在质或量上都存在较大的不足,开源的核心技术和话语权主要掌握在国外少数科技巨头手中,面对美国对我国进行的技术封锁和科技脱钩的危境,存在闭源断供的风险。因此,如何大力发展我国的开源教育,快速培养开源人才是高校人才培养面临的挑战。本文通过分析开源的技术发展、产业和人才现状以及高校培养人才的紧迫性,提出以开源能力评价为导向,创新人才培养模式,引导师生进军开源,以培养我国开源技术核心人才。

【关键词】开源教育 人才培养 能力评价 科技自立自强

## 一、引言

2021年3月《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中明确指出发展开源社区等创新联合体,这表明将开源作为国家战略纳入我国实现关键数字技术创新应用新的突破,将从“开源大国”走向“开源强国”之决心。<sup>[1]</sup>开源软件以开放、共享、协同、共赢的新型生产方式,成为全球信息技术发展的强大推动力,也是互联网时代人们在软件领域极为成功的大规模协作实践,已成为构建人类数字社会的重要基石。当前99%的商业软件含有开源组件,75%则直接由开源代码组成<sup>[2]</sup>,开源生态造就了众多成功的项目,许多行业创新是站在开源肩膀之上的产物。虽然开源运动发展至今,软件产品的种类越来越多,功能也越来越强大,社区的规模也越来越庞大,但相应的,支撑开源开放的工

具和技术体系也越来越难以掌握,新的形势对开源的进一步发展带来了不小的挑战。<sup>[3]</sup>

开源项目代表着信息技术发展的前沿应用方向,开源社区代表着开放的协作精神与创新的摇篮。《中国开源年度报告2020》<sup>[4]</sup>对2020年全年活跃项目进行了统计与排名(见表1、表2)。

从世界Top10和中国Top10活跃的开源软件项目来看,世界排名前10的开源项目中没有一个来自中国,并且中国第1名与全球第10名活跃度差距超1.6倍,说明了我国的开源项目在全球影响力不大。我国开源业绩比较突出的公司是阿里巴巴和百度,统计第1名与第10名的项目活跃度后,发现相差接近4倍,表明中国开源项目之间的发展极不平衡。从表2中还可以看出,国内开源项目活跃的是一般不太涉密的前端群体,缺乏核心项目。

表1 世界开源Top10项目

#	项目 (Project)	开发语言 (Language)	项目加权 活跃度 (Activity)	开发者数量 (Developer count)	评论总数 (Issue_ comment)	开放状态 的问题 (Open_ issue)	新增问题 单数量 (Open_pull)	PR下的Review 评论数 (Pull_review_ comment)	被合并的 PR数量 (Merge_pull)
1	flutter/flutter	Dart	35108.77	16758	128091	14857	7248	18778	5003
2	microsoft/vscode	TypeScript	27336.95	14162	101450	16366	1816	1778	1383
3	microsoftDocs/azure-docs	PowerShell	24021.03	9507	87663	11749	3118	885	1853
4	home-assistant/core	Python	22609.91	8215	77017	5381	7851	31039	6697
5	tensorflow/tensorflow	C++	21637.69	9681	63983	6241	2993	8073	2198
6	kubernetes/kubernetes	Go	19878.4	6111	239791	3669	6869	31345	4817
7	NixOS/nixpkgs	Nix	18571.45	3049	84523	4368	19693	0270	16309
8	pytorch/pytorch	C++	14515.73	4844	69110	4890	11776	39109	336
9	Dotnet/runtime	C#	14076.75	3731	83600	7082	7540	41215	6531
10	DefinitelyTyped/DefinitelyTyped	TypeScript	13267.93	4100	53588	538	6426	6529	5414

收稿日期: 2021-11-03

作者简介: 陈丽,浙江大学软件学院副院长、副研究员;赵斌,浙江大学软件学院教学办副主任;倪超,浙江大学软件学院讲师;杨小虎,浙江大学软件学院研究员。

表 2 国内开源 Top10 项目

#	项目 (Project)	开发语言 (Language)	项目加权 活跃度 (Activity)	开发者数量 (Developer count)	评论总数 (Issue_ comment)	开放状态 的问题 (Open_ issue)	新增问题 单数量 (Open_pull)	PR 下的 Review 评论数 (Pull_review_ comment)	被合并的 PR 数量 (Merge_pull)
1	ant-design/ant-design	TypeScript	8090.1	3746	32882	3341	2844	3315	2416
2	PaddlePaddle/Paddle	Python	5236.1	960	15226	1817	5487	9517	4083
3	pingcap/tidb	Go	4339.4	482	53550	2727	5041	11117	4097
4	NervJS/taro	JavaScript	3649.4	1621	9595	1768	805	132	602
5	apache/shardingsphere	Java	2858.7	786	9332	1581	3234	1851	2990
6	ant-design/Ant-design-pro	TypeScript	2784.2	1362	6625	1383	228	118	118
7	alibaba/nacos	Java	2760.6	1204	7037	1233	635	764	491
8	apache/incubator-tvm	Python	2699.6	355	8255	363	2106	8526	1876
9	ElemeFE/element	Vue	2507.7	1694	5016	567	117	7	19
10	apache/incubator-echarts	TypeScript	2307.4	1183	7622	962	301	328	230

目前,据 Github 统计,我国开源贡献度已排名世界第二,但国际上开源的核心技术和话语权主要掌握在国外少数科技巨头手中。随着近些年来中国综合国力的不断增强,国际上个别发达国家感受到了中国的“威胁”,开始制约中国发展,阻碍中华民族复兴。构建国内的开源生态和培养有核心竞争力的开源人才,既能避免我们过于依赖国外某些领域开源技术,造成未来可能出现的技术“卡脖子”和产业“贫血”现象,又能快速发展我国的关键数字技术,引领前沿技术的演进方向。

二、产业变革驱动下开源软件人才培养面临的挑战

我国是拥有软件工程师最多的国家,但人才结构失衡,多年来培养的几乎都是基于 Windows 平台的人才,熟悉 Windows 平台的人才过盛,而熟悉开源软件人才,尤其掌握核心技术的人才匮乏。<sup>[5]</sup>根据 Linux 基金会和 edX 发布的开源就业报告<sup>[6]</sup>,开源人才的招募已经成为软件企业人力资源主管的第一要务,92%的招聘经理表示开源人才难觅。如何提高高校师生对进入开源技术开发紧迫性的认识,结合丰富的开源环境培养具有开源思维和开源能力的软件人才是高校面临的挑战。

1. 产业变革对软件人才提出新要求

目前,全球 97%的软件开发者和 99%的企业使用开源软件,全球 70%以上的新立项软件项目采用开源模式。红帽公司 2021 年发布的全球《企业开源现状》调查报告显示“最具创新能力的企业都在使用开源”。开源已应用到绝大部分软件产品,延伸到开源硬件、设计、产业互联网等很多领域,几乎覆盖了软件开发全域场景。

随着开源软件的盛行,不少国家也积极提倡、推广开源软件。同时也涌现了许多非常优秀的开源社区,例如最早且最著名的 Git:一个分布式的版本控制系统,由 Linus Torvalds 编写,用作 Linux 内核代码的管理,一经推出就大受欢迎,目

前已经有 140 多万开发者用户。开源中国成立于 2008 年 8 月,其目的是为中国的 IT 技术人员提供一个全面的、快捷更新的用来检索开源软件以及交流使用开源经验的平台,目前开源中国社区已收录超过两万多款开源软件。开源已成为全球数字科技创新的大趋势。经过近 40 年的实践发展,开源逐渐成长成为一种强大的技术创新模式,并从最初的软件行业走向了硬件、芯片等多个领域。开源商业模式也在逐渐成熟。

我国作为互联网技术强国,拥有庞大的应用市场和丰富的开发者群体等资源优势,需要把握好历史发展机遇,努力在全球开源图景中占据应有之位。然而,在日益繁荣的开源环境下,我国的开源人才无论在质或量上都存在较大的不足。<sup>[3]</sup>为了适应新时代发展对开源人才培养的需要,高校应特别注重发展学生的开源意识和开源能力,来应对技术和产业变革的需要。

2. 如何引导高校进入开源教育模式,为满足我国开源生态建设急需培养人才

开源可以助力我国关键数字技术“弯道超车”,因此开源生态本土化是必经之路。近年来,我国科技发展速度惊人,但在开源的商业策略、长远的战略上与国外的企业还有着较大差距,建设我国开源生态,需要在人才培养方面加强开源意识,建立开源文化,引入开源教育模式,培养开源核心技术人才。

(1) 如何扩大开源人才数量:加强开源意识培养,扩大进入开源社区的学生群体。调查表明<sup>[4]</sup>,如图 1 所示。我国 21%的开源参与者接触开源时间不到 1 年,接触开源时间在 1~2 年占 29%,3~5 年占 30%,6~10 年占 11%,10 年以上占 9%。同时开源社区贡献者中,年龄在 10~20 岁占 5%,20~30 岁的占 64%,30~40 岁的占 28%,40 岁以上的占 3%。说明我国开源参与者中学生参与者较多。Github 海外有 900 万学生开发者,其中高校开发者占比 23%,近五年来增

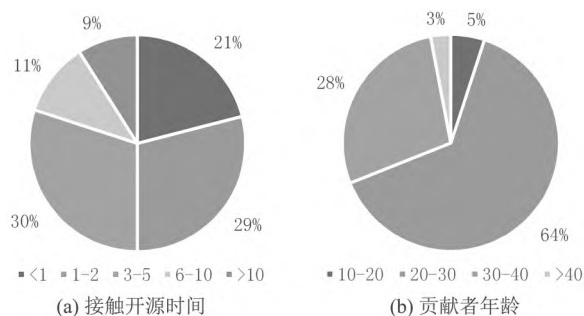


图1 开源项目参与人员情况调查结果

长超过40%。并且许多世界级的开源项目来自高校和科研机构,其中就有slurm、PyTorch、Flink、Spark、gatk、git、openstack、MESOS等知名项目。Caffe、Linux、Git、Mesos均由25~30岁的程序员创立,这也佐证了富有精力与热情的高校学生与教师是开源的重要贡献人群。Google Summer of Code社区贡献超过3800万行代码,产生25万+行代码/周,平均每位学生每周贡献超过240行代码。学生是可塑性比较高群体,更容易接受新事物跟进最新开源热度,我国高校在人才培养中要注重开源意识培养,积极引入开源文化,引导学生参与到开源工作中。

(2) 如何培养高质量开源人才:引导学生有志成为核心贡献者,引领开源技术的发展。我国开源群体大,但使用多贡献少,我国的开源参与者中一周工作时间不超过20小时占88%<sup>[4]</sup>,说明较低的职业开源开发者比例很高,显示出中国企业对于开源项目依然主要停留在用户层面。多数开源贡献者依然使用业余时间或者工作间隙进行有限的贡献和参与。缺少35~40岁以上的开发者,说明中国资深工程师参与者少。2014-2019年国内Kernel代码贡献占比2.92%,CNCF项目中,国内代码贡献占比6.19%。建立开源生态,规模化发展开源项目,达到技术崛起,需要大幅度提升资深工程师参与比例。高校学生尤其是研究生将在较短周期内步入资深工程师行列,引导其成为核心贡献者,才能达到我国引领开源技术发展的目标。

(3) 如何培养关键基础软件领域的开源人才:带领学生进入“卡脖子”技术开源社区,培养未来的核心贡献者团队,加快我国基础软件的发展速度。我国缺少顶级开源项目,从芯片、操作系统、虚拟化、云计算、人工智能、代码托管等领域来看,国内大量依赖海外开源项目,如果全部抛开,另起“炉灶”比较难,因为市场对已有成熟软件有依赖心,新软件需求长期锤炼和一定积累才能获大量用户的信赖。通过大量培养开源人才,使杰出贡献者簇涌,对国际主流技术社区后端核心层

“洗牌”,再创新,掌握基础软件的核心技术,摆脱受制于人的困境实现科技自立自强。

### 3. 如何改变人才培养模式,建立开源教育人才培养体系

传统软件人才培养模式缺乏基于开源能力的培养,与目前开源环境下的应具备的能力要求相差较远。有些企业为了尽快培育符合需求的开源技术人才,积极主动地参与到开源人才培养中。例如,红帽公司积极开展开源教育来提高参与者的开发能力<sup>[3]</sup>,希望培养更多能够参与开源运动的人才,以推动开源事业的发展。

怎样建立机制在高校推广开源教学?事实上,高校也意识到开源教育的重要性,认识到开源作为一种新的技术创新模式和商业模式,已完全改变了软件产业生态,势必要求改变人才培养模式,构建开源教育人才培养体系以快速适应产业的新发展。许多高校已经针对开源人才进行了教育方面的改革和努力,其中包括对教师进行赋能,鼓励老师将开源项目和必修课程进行结合。教育界和企业界都认识到人才是产业的基石,2020年9月“教育部—华为‘智能基座’产教融合协同育人基地”项目启动建设,统计至2021年7月,计算机系统能力高级研修班培养了725名高校教师、人工智能线上培训294名老师。而针对开源相关的专业学分课包括45所高校228门课程覆盖了29000多的学生。而相关教材涉及人工智能实践教材9本、智能计算系列丛书8本。浙江大学以开源能力评价为导向在课程教学、项目实训、实习实践、毕业能力要求等人才培养环节提出了开源能力的培养要求和教学目标,构建了开源教育人才培养体系。

### 三、面向开源教育推进课程教学改革

课程教学是实现开源人才培养目标的重要载体,是保障规模化开源人才培养和提高培养质量的关键。通过系列课程的教与学,可以促进我国软件人才在开源技术引领、软件产业创新方面开创新局面。让学生在系列课程学习中不断接受开源项目,在课程大作中以开源软件为基础开发。这样,以任务驱动学生学习,培养其开源能力,包括开源的认知能力、开发编码能力、新技术自学能力、项目测试能力、文档写作能力、团队沟通协作能力和创新能力等。

#### 1. 精选优秀开源软件设计课程教学案例

将开源软件用于课程教学,引导学生接触开源、参与开源、分析开源、使用开源、贡献开源,营造开源教育文化。课程教育开始前,教师应精选代码规模合适、质量高的开源软件项目作为教学

案例推荐给学生,明确课程考核要求,引导学生开始阅读开源项目,促进学生自我开源意识的萌芽及形成。开源软件有别于一般教学案例:① 开源软件代表一种新的开发技术与模式,具有大规模协作的特性,学生必须学习开源的制度、方法、模式、文化等;② 开源软件是本领域最新技术的最优呈现,是最贴近产业的教学实践范例;③ 开源软件提供给学生的是跟踪学习和模仿机会,开源软件是众多开发者集体智慧的结晶,具有持续演化的特性,可以实时体会到优秀工程师的开发方法和开发技巧,从而激发学生的开发兴趣和动力,吸引学生加入。在学习过程中,学生也可推荐认为更适合本课程学习的更佳开源项目,由教师评估后决定是否列入本课程教学案例,丰富和优化案例。

2. 分层次循序推进开源能力培养,合理完成基本教学目标

以培养学生使用开源的能力为目标,分层次的完成开源教学。第一层次为阅读分析能力,根据课程特点确定教学方式,例如:将学生分成若干小组,以小组为单位阅读开源项目,通过优秀代码的阅读,也让学生意识到自己代码能力缺陷,培养学生养成良好编码习惯。这个阶段要求各小组撰写阅读综述文档或技术博客,通过小组互评,发布在开源社区结合其他开源参与者的点评,评价阅读分析阶段的成绩。第二层次为使用开源的能力,模拟或提供真实应用场景,以开源项目实践为手段,培养学生综合运用知识能力和软件复用能力。这个过程全程在一个开源应用场景下组织开展,各小组在当前开源软件的基础上开发应用软件项目,进一步加深学生对开源软件的学习。教师事先明确项目开发的基本目标,让学生开展任务驱动式学习,定期召开项目报告会,一方面同步各组的基本进度,另一方面检查督促任务进展,教师应作阶段性点评,激励先进,鞭策后进。根据项目实施的效果评价该阶段的学习成绩。

3. 设计拔尖培养通道,实现课程教学高目标

开源人才培养的最高目标是核心贡献者涌现,因此鼓励学生参与开源,激励贡献度体现,是课程教学的最高目标。在学生完成了开源软件的阅读分析和基于开源的应用场景实操后,对开源项目案例已比较熟悉,如能发现代码漏洞并加以改进,或者能对代码进行优化,体现学生对开源软件后端的维护设计能力,说明学生已进入了核心贡献者行列,对学时有限的课程教学来说是比较高的要求,是属于教学的最高目标。学生可以根据自己在前两阶段的学习成效,自主选择参加此

阶段。课时也可实行弹性制,甚至可以延长至课时结束后的 2 个月内进行考核。教师应对此阶段的学生进行指导,甚至参与其中,引导学生对开源项目进行源码贡献,对其他贡献者的源码设计进行改进维护,以软件贡献后被复用效率、产生的影响力为成效。如果能实现这个目标,就表明学生逐渐掌握了核心技术,解决复杂软件工程问题的能力已有显著提高。

4. 以 OBE 教育理念为导向,建立课程评价体系

对课程学习的各阶段,以成果导向教育理念(Outcome-Based Education, OBE) 为导引<sup>[7]</sup>,围绕课程教学目标要求,建立各项能力指标及说明见表 3。

表 3 各项能力指标说明

指标名称	含义	备注
认知能力	指学生对开源案例阅读后提取信息的能力	案例阅读的时长和代码量
分析能力	指学生在思考中把开源案例的整体分解为若干部分进行研究、认识的技能	案例阅读的时长,满足新系统设计目标的程度
编码能力	指学生对开源代码的应用、读写、调试等一系列的能力	代码编写的代码量和 git 的提交量
文档能力	指学生概括和总结参与开源软件过程中用文字表达的能力	commit 的说明、代码的注释量
设计能力	指学生根据开源系统分析的结果,运用软件工程的思想和方法	设计出满足所要求的目标的新系统程度
测试能力	指学生在开源软件开发完成后纠正软件中所存在的软件故障	提交的代码的质量和、纠正软件的故障量
自学能力	指学生不需要其他任何人的帮助,主动获取所需知识,通过独自自习的方式构建开源软件方面的能力	复用开源软件代码量、独立解决 Issues 的次数
沟通能力	指学生与他人有效地进行沟通信息的能力	Issue 讨论, Discussion 讨论等有效沟通的次数
协作能力	指建立在团队的基础之上,发挥团队精神、互补互助以达到团队最大工作效率的能力	
创新能力	指学生不断提供具有开源软件中的新思想、新理论、新方法和新发明的能力	代码的复用量、贡献量、解决增加新功能的 Issue 个数

分层次对学生学习的各阶段进行评价。教学过程分为代码阅读、案例综述、开源应用(或场景应用)、开源创新等四个阶段层层递进见图 2 所示,其中前 3 个阶段中每位同学必须参加,开源创新阶段作为可选项,鼓励学生参加,尤其是鼓励前三阶段表现优秀的学生参加,并把第四阶段成绩作为课程附加分计入。另外,针对不同的课程应该有不同要求,如低年级的专业基础课程应前三阶段教学为宜,本科高年级及研究生专业课程应鼓励大部分学生参与第四阶段学习,其成绩以较小比例纳入课程评价。

可以看出,基于开源的教育与传统教育主要区别是:① 开源是课程思政的生动案例,在美国

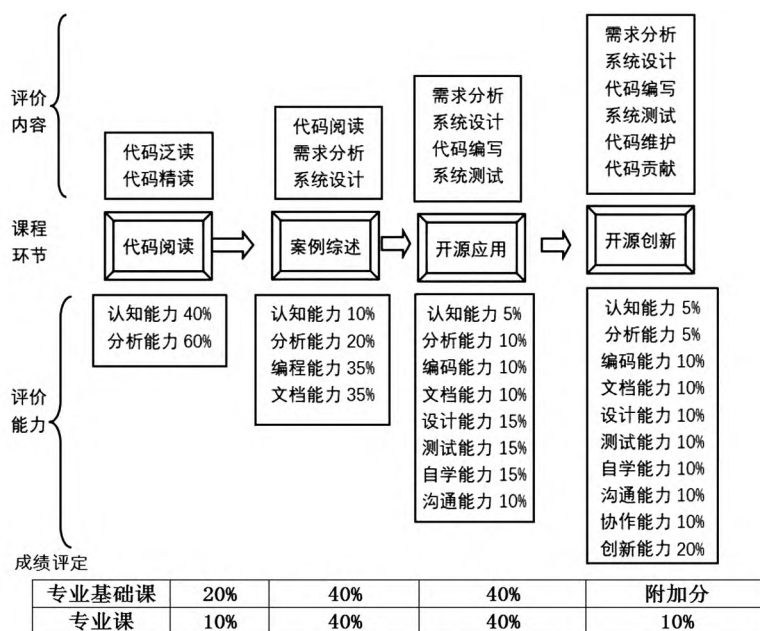


图 2 基于 OBE 教育理念课程评价模型

试图对我国进行技术封锁,科技脱钩的危境下,让学生投身开源环境,深切体会到科技自立自强的急迫性,激发奋进斗志和使命担当。② 课程知识更新实时,与产业前沿无缝接轨。③ 更注重学生的学与思,大大提高学生的课程学习获得感,激发学生的创造力。④ 软件规范、标准、协同意识大大增强,不再满足于软件功能实现和效果。⑤ 对教师的要求增加。对教师跟踪产业前沿、更新教学内容的要求提高,不能再照本宣科。

#### 四、基于 OBE 理论构建学生开源能力评价体系

在“破五唯”的要求下,学生的能力评价应该从创新型、应用型、技能型培养目标要求出发,体现服务贡献的能力。无论是课程教学还是人才培养的其他环节,如项目实训、实习实践,以及毕业时具备的专业能力要求,都应该以 OBE 教育理念为导向,紧紧围绕培养目标要求,关注学生的学习结果、能力提升和获得感。目前各高校提出的研究生毕业要求,有发表论文、申请专利、成果获奖、科研项目等指标,以此证明学生毕业时的能力和水平,在 OBE 的要求下这些评价不够科学和完备,其中学生开源能力评价是非常有效的补充。尤其是将开源能力作为毕业可选要求之一,是引导学生进军开源,培养我国的开源大军和核心贡献者的重要手段。

针对大量学生群体的开源能力评价,仅靠教师人工收集特征数据进行分析判断,不但工作量大而且成绩评定合理性上难免会有失偏颇,所以构建开源能力评价系统非常有必要。浙江大学在

多次实践的基础上,建立了学生开源能力评价模型。

##### 1. 开源能力指标体系

为了全面地评价学生的开源能力,我们从开发者(如学生)特征和社区特征两大维度共计 14 个特征对开发者和社区的属性进行全面刻画。开源能力评价模型采用特征工程,特征的名称及其说明见表 4。针对这些特征,从多个开源平台(如 Github, GitLab, Gitee, Bitbucket 等)大规模抽取相关信息,并进行人工标注,完成开发者贡献评分,构建适用于训练开发者开源能力评价模型的大型数据集。

##### 2. 开源能力评价模型

开源能力评价模型目的是构建一个“端到端”的开发者开源贡献评价模型,体现以能力培养为导向的 OBE 教育理念,其整体框架见图 3。该模型分为离线训练和在线评测两个阶段,其中离线训练阶段完成评估模型的构建,在线评测阶段将对一个特定的开发者根据其提供的开源账号信息,完成其开源贡献能力的评分,具体如图 3 所示。

##### 阶段一:离线训练

离线训练阶段涉及若干内容,包括训练数据的预处理,训练数据的划分,模型的设计和模型训练的配置等。首先对训练数据进行预处理操作,包括空值数据过滤,异常值处理,数据变化和归一化等操作。其次,对提纯后的数据进行训练集和数据集的划分,本模型采用 8:1:1 的划分策略,即将数据集中 80% 的数据用作训练数据,10% 的

表 4 开发者特征及开源社区特征说明

开发者特征			社区特征	
特征名称	说明	考查项	特征名称	说明
Commit-Num	代码提交数:代码提交次数越多,说明其在开源项目中提交的次数就越多,贡献越多	认知能力、编码能力、测试能力、文档能力、沟通能力、协作能力、自学能力、创新能力、设计能力	NumRepo	开源项目的个数:反映了一个社区自身的受欢迎程度和其重要性
PRNum	Pull Request 提交数: PR 请求的数量越多,说明开发者将代码贡献到社区的次数越多	认知能力、分析能力、测试能力、文档能力、沟通能力、协作能力	Repo-Type	开源项目的类型:不同类型的开源项目其自身难以程度也不一样,比如系统底层的、容器类的项目总体而言要比应用类的难一些
PRNum-Confirm	被合并的 PR 数:开发者提交的 PR 请求被合并到项目中,说明其贡献的代码得到了社区的认可。认可的次数越多,其贡献越大	认知能力、分析能力、测试能力、文档能力、沟通能力、协作能力、自学能力	Sponsor-Num	赞助商的数量:该数量反映了企业界有多少企业加入了该社区,这也反映了社区在企业界的认可度
Issue-Num	参与的 Issue 讨论数量:开发者参与到 Issue 讨论的数量越多,其在开源项目就更加活跃,对社区的帮助也越大	分析能力、文档能力、沟通能力	Age	社区成立的年份:历史越悠久的社区,其影响力越大
IssueFix-edNum	解决了 Issue 的数量:当解决了开源项目中 Issue,相当于解决了项目中存在的问题或者完成了新功能的添加,这直接反映了开发者的贡献大小	认知能力、编码能力、文档能力、沟通能力、协作能力、创新能力	Ratio-QstAns	用户提出问题和解答问题的比例:用户在加入社区过程中会遇到问题,所反映的问题如果能够得到快速的解答,用户的学习门槛会降低,其对社区的满意程度会增加,从而会吸引更多的人加入社区,形成良性循环
Follow-erNum	粉丝数:开发者的粉丝数量直接反映了其在开源社区中受欢迎的程度	沟通能力、协作能力、自学能力、创新能力	Reward	有无对于用户参与开源的激励措施:奖励措施会吸引更多的人参与社区,从而壮大社区,增加开源项目,增加社区影响力
RepoSta-rNum	所参与的开源项目获赞的数量:开源项目获得赞的次数,反映项目在开源平台的受欢迎程度,同时也反映了其在企业界的重要程度,如果开发者参与并贡献了广受欢迎的项目,其贡献程度越大	认知能力、分析能力、自学能力	Com-merce	有无商业化部分:商业化会给社区带来盈利,维持社区的可持续发展,增加社区的运营投入,吸引更多的开发者加入,提升影响力,开发更加优秀的产品

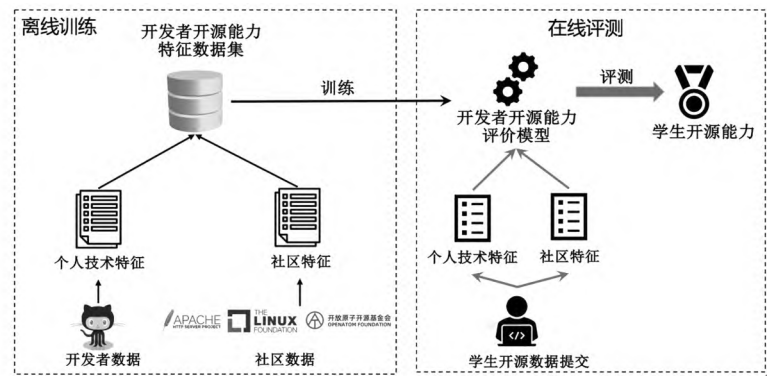


图 3 开源能力评价模型框架图

数据用作训练阶段的验证数据,剩下的 10% 数据用作模型的测试数据,并保证训练数据、验证数据和测试数据中的数据分布与整体数据的分布相同。然后,构建一个深度学习模型,该模型根据训练数据完成模型的构建,并在训练的过程中以小批量 256 个实例进行迭代更新,并时刻检验模型是否已经达到较好的评测能力,从而尽早结束训练。最后在测试数据集上判断模型的泛化能力。

总体而言,“端到端”的开发者开源贡献评价模型是一个回归模型,其目的是建立输入特征(如开发者特征、社区特征)与目标数值(如开源贡献能力评分)之间的映射关系。对于一组实例,其中每一个实例都是由开发者特征  $F_{\text{developer}}$  和社区特征  $F_{\text{community}}$  组成,即  $\langle F_{\text{developer}}, F_{\text{community}} \rangle$ ,模型的目标是

学习一组参数  $w$ ,使得损失函数最小,形式化如下:

$$\begin{aligned} w^* &= \operatorname{argmin}_w L(w) \\ &= \operatorname{argmin}_w L\left[\sqrt{\frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^m w_{ij} F_{ij}\right)^2}\right] \end{aligned}$$

其中,  $n$  表示训练实例的个数,  $m$  表示特征的个数,并且  $F = \langle F_{\text{developer}}, F_{\text{community}} \rangle$ 。

阶段二:在线评测

该阶段在评测学生开源能力时,首先需要提供学生的账号信息(如 GitHub 上的用户名,或者作者 Committer 的邮箱等),然后根据其账号信息,系统会自动地从开源平台抽取与之相关的开发者特征和社区特征,并根据训练阶段的预处理操作,对当前数据进行必要的处理。然后,将处理

后的数据输入到训练好的开发者开源能力评价模型中进行评测,从而得到学生的开源能力评估结果。

### 3. 评价结果运用

从评价模型得出的评估结果可以运用于课程教学和其他培养环节中,浙江大学将评估结果运用于招生选拔、课程教学、项目实训、毕业基本要求等环节,有效激励了学生对开源的投入,取得了开源人才培养的初步成效。不同环节要求不同见表5。

表5 各环节评价结果运用参考

	课程	硕士招生	项目实训	毕业要求
评价成绩	不同课程要求不同	≥60	≥75	≥90

在课程教学中,根据教学目标和课程安排运用结果,课程考核一般是图1中第三、四阶段运用该模型评价结果,其他阶段教师评定为主。第三阶段要求低,可设定低值,也可以教师另设基本标准评定。第四阶段可设定40为阈值。

浙江大学将评估结果运用到硕士招生选拔,设定阈值为60,作为学生实践能力的佐证材料,在面试过程中作为考查项供专家参考。项目实训为4学分,是本科高年级和硕士研究生必修的实践环节,开源能力评价达到75分可获学分,当然学生也可以选择其他实践项目获此学分。开源能力评价达到90分可作为满足硕士研究生毕业要求的可选条件项。在结果运用过程中,浙江大学根据师生对评估结果的反馈不断优化离线训练,完善评价模型。

### 五、结束语

开源模式从以个人兴趣驱动的自由与共享的软件开发模式,发展为汇聚全球优秀人才、技术迭代迅速、绑定市场应用、掌控产业生态等的新的商业模式,从而成为国际科技巨头争夺软件产业生态主导权的新赛道,对我国的产业安全造成了较

大影响,开源科技自立自强、生态建设本土化迫在眉睫。2020年,开放原子开源基金会作为我国第一个开源基金会成立<sup>[8]</sup>,2021年6月2日华为正式将其著名的鸿蒙操作系统的所有源代码和知识产权捐赠给了开放原子开源基金会,这不仅补齐了基金会在我国开源生态的重要版图,也标志着本土化开源生态建设已经起步。开源本土化需要大量贡献者支撑才能真正实现,高校学生是开源开发者的重要来源,是未来我国开源生态建设的主力军。开源教育是产教融合构建创新人才培养体系的重要组成部分,是开源事业可持续发展的重要保证。目前国内高校虽在不同程度地推进开源教育,但需要进一步加快推动和繁荣发展,我们分享浙江大学的经验,供兄弟高校参考和交流提高。

### 参 考 文 献

- [1] 江亚蔓. 中国工程院院士倪光南: 拥抱开源构建科技命运共同体[EB/OL]. [2021-08-25]. [http://www.liangjiang.gov.cn/Content/2021-08/25/content\\_10216472.htm](http://www.liangjiang.gov.cn/Content/2021-08/25/content_10216472.htm).
- [2] 王晓冬. 我国开源软件产业面临的突出风险及对策研究[J]. 信息安全研究, 2021, 7(10): 973-976.
- [3] 梅宏, 周明辉. 开源对软件人才培养带来的挑战[J]. 计算机教育, 2017(1): 2-5.
- [4] 王伟, 赵生宇. 2020 中国开源年度报告[EB/OL]. [2021-08-19]. <https://kaiyuanshe.cn/document/china-os-report-2020/>.
- [5] 郑金武. 中国工程院院士倪光南: 开源软件人才需求将日趋旺盛[N]. 中国科学报, 2013-06-06(1).
- [6] JASON PERLOW. The 2021 Open Source Jobs Report[EB/OL]. [2021-09-20]. <https://www.linuxfoundation.org/tools/the-2021-open-source-jobs-report/>.
- [7] 张男星, 张炼, 王新凤, 等. 理解 OBE: 起源、核心与实践边界[J]. 高等工程教育研究, 2020(3): 109-115.
- [8] 马越. 中国开源的底气[J]. 软件和集成电路, 2021(6): 60-61.

## On Challenges and Solutions of Software Talent Cultivation in Colleges and Universities under the Open Source Mode

Chen Li, Zhao Bin, Ni Chao, Yang Xiaohu

**Abstract:** With the vigorous development of software technology, especially the internet, open source has increasingly become an important direction for the development of information technology in the world and industry. However, in the increasingly prosperous open source environment, open source talents in China are insufficient in terms of quality or quantity. The core technology and voice of open source are mainly in the hands of a few foreign technology giants. There is a risk of closed source and stopping supply since the technology blockade and technology decoupling from American. Therefore, how to vigorously develop open source education and quickly train open source talents in China is a

challenge for Chinese universities. This paper firstly analyzes the development of open source technology, the status of industry and talent, and the urgency of cultivating talents in universities, and then proposes an innovative talent training model based on the evaluation of open source capabilities, and finally guides teachers and students to join open source to cultivate the core talents of open source in China.

**Key words:** open source education; talent cultivation; capability evaluation; science and technology self-independent (责任编辑 黄小青)

---

(上接第 80 页)

### Thoughts and Explorations on the Cultivation of Transportation Talents under the Strategy of Country with Strong Transportation Network

*Jiang Wei, Sha Aimin, Xiao Jingjing, Wang Xuhao*

**Abstract:** Transportation is a basic, leading, strategic, and service industry for national economic and social development, and is the foundation for guaranteeing the country's political stability, economic & social development and national defense security. The major strategic demand of building a "transportation power" has brought great challenges and opportunities to the transportation industry, and the demand for talents is unprecedented. The development of new infrastructure based on big data, internet of things, and artificial intelligence technology has brought new impetus into the high-quality development of the transportation industry, and at the same time puts forward new demands for talent training. The paper analyzes the requirements of the development direction of "Transport plus" on the interdisciplinary knowledge background, international vision, and country needs of talent training. Based on the above requirements, it points out the current problems in the training of transportation professionals in universities. As the same time, it puts forward development ideas from three aspects: expanding professional connotation and keeping pace with the times on cultivating compound talents, innovating talent training mode and improving internationalization ability as well as systematically building the ideological and political theories teaching in all courses in transportation. This paper introduces the practice and exploration of future transportation College of Chang'an University, international engineering direction and ideological and political theories teaching in all courses, which provides a reference for the reform of talent training under the strategic of country with strong transportation network.

**Key words:** transportation specialty; talent training; interdisciplinarity; internationalization; ideological and political theories teaching in courses (责任编辑 黄小青)