

布拉格经济与商业大学工商管理学院



国际高管 MBA 课程 (ISBM)

**硕士论文 ( Master Thesis )**

**开源的价值观与文化的传递**

**探索 Apache Way 文化在中国发起和非中国发起  
的项目中的传递**

[Sharan Foga] 2019 年 6 月

论文指导老师 : Ing. Aleš Kubíček, Ph.D. & Justin McLean

# 导读

## 译者序：

这篇论文是我在 2019 年就想翻译的，当时和徐红伟起了个头（他翻译了第 4 章的前二节），当时因为种种原因和自己的懈怠没能继续完成。但是这几年来始终念念不忘想完成这个自己托付给自己的任务。

回想起开源社与 ASF 联合举办 2015 Apache 中国路演时（2015 年 10 月 24-25 日，北京），当时 ASF 只有一个源自中国的项目 Apache Kylin（源自当时的 eBay 中国研究院，2014 年 11 月进入 ASF 孵化器，并于 2015 年 11 月 18 日毕业成为 Apache 顶级项目；而系出同门的 Apache Eagle & Apache Griffin 之后也进入了 ASF 孵化器并毕业成为 ASF 顶级项目）。随后 Apache CarbonData，Apache RocketMQ 分别于 2016 年 6 月和 11 月进入 ASF 孵化器，逐渐开启了源自中国的开源项目进入 ASF 孵化器到顶级项目波澜壮阔的征程。



Brett Porter, 阿帕奇软件基金会主席



Ross Gardler, 阿帕奇软件基金会总裁



David Nalley, 阿帕奇软件基金会董事



Niclas Hedhman, 阿帕奇软件基金会副总裁

### 2015 Apache 中国路演的部分 ASF 贵宾

截至 2019 年 4 月 30 日为止（本论文的样本数据时间），源自中国的 ASF 项目已有 14 个（8 个顶级项目和 6 个孵化项目）；到了 2022 年 7 月底，ASF 源自中国的项目已然翻倍达到了 30 个（19 个顶级项目和 11 个孵化项目）。

目前 ASF 孵化器里的 34 个项目，有 11 个源自于中国（约占 1/3），特别值得注意的是，

**2021~2022 年所有 9 个新的 ASF 孵化项目都来自于中国**（2022 年 4 个，2021 年 5 个）。而 2020~2022 这三年毕业的所有 20 个项目里，源自中国的项目达到了 9 个（2020：5 个，2021：1 个，2022：3 个毕业项目）。

注：请参考导读最后之参考资料之“ASF 源自中国之项目列表”。

根据 ASF 的正式统计，中国是全世界下载 ASF 项目/代码最多的国家，而目前 ASF 全体将近一千名（997）正式成员（Member）里，可统计到或是有联系方式的中国人/华人身份者大约只有 46 名左右，而全球八千余名（8,257）ASF Committers 里，中国人所占的比例应该也是比较少的（目由于前缺乏正式统计数据，可以约略以 5% 以下估计）。

从上述的源自中国的孵化项目到顶级项目的蓬勃发展速度，以及 2021 与 2022 连续二年有中国的正式成员当选 ASF 董事会董事（2021 吴晟，2022 姜宁）来看，中国的开源专家参与到全球开源基金会治理中，这具有里程碑意义。



图片来源：A Guide to Enterprise Open Source, by Dr. Ibrahim Haddad, 2022

开源的发展分为四个阶段：消费、参与、贡献与引领，我们看到了源自中国的开源项目，正在不断增长与进步，从最初级的拿来主义的消费者，快速转型为参与者、贡献者，甚或引领者。源自中国的开源项目以及贡献者们在国际顶级的开源基金会，如 Linux 基金会、CNCF，以及 ASF，似乎正如鱼得水。

当读者看完本论文之后，您可能会拍案惊奇道：原来如此，有为者亦若是！

最后，我也为读者们整理了一份思维导图，帮助大家快速地提纲挈领。同时也提醒读者务必不要错过精彩的第二章，尤其是第一节：国家层面的文化要素。

[开源的价值与文化的传递思维导图](#)

--- 刘天栋 Ted，开源社联合创始人 | ASF 正式成员、  
ASF 孵化器 PMC & 导师、社区发展 PMC、筹款委员会 & 赞助伙伴大使

## 来源：

- 原文及译文：<https://github.com/kaiyuanshe/open-source-articles/issues/104>

## 作者、翻译及编审：

- 作者：**Sharan Foga**, Apache 软件基金会董事（2021 届、2022 届），曾亲身来到 2019 中国开源年会（COSCon'19）以本文内容发表主题演讲。
- 翻译：刘天栋.Ted，徐红伟.stronghx（第四章 4.1 节 ~ 4.2 节）
- 审阅：庄表伟、陈阳、陈亮（Apache member，Apache CarbonData PMC）
- 编辑：
- 设计：

## 许可证：

本作品采用[知识共享署名-相同方式共享 4.0 国际许可协议](#)进行许可。



## 专家点评：

### ● ASF 正式成员、开源社理事会、开源社顾问委员会（以下按姓名汉语拼音排序）

陈亮、陈阳、堵俊平、郭雪、江波、姜宁、史少锋、谭中意、王伟、王小瑞、吴晟、杨丽蕴、庄表伟、赵生宇

#### 【专家点评】（以下按提交先后次序排序）

**吴晟**，Apache 软件基金会的社区模型和开源社区构建之道，一直都是中国开展开源软件的重要参考资料。Sharon 的这篇论文的研究背景和时机，恰恰是中国开源软件真正开始拥抱世界，开始全面融合和学习的几年。大量的中国原创项目选择加入 Apache 软件基金会。通过这篇论文，我们可以更深入的学习到，这些对于我们的大事件，在 Sharon 这位 Apache 软件基金会的自身成员眼中的理解和剖析。对于我们更好的融入世界开源，合作发展，有着至关重要的意义。

---- 吴晟，Apache 软件基金会首位中国董事，Apache Member, Apache SkyWalking Creator

**谭中意**：Sharon Foga 是 Apache 基金会的董事，曾来过中国发表关于开源文化的主题演讲，介绍她对 Apache Way 的研究。现在 Ted 牵头把这篇文章翻译成为中文，我浏览了下，才明白为什么 Ted 念念不忘一直要把它翻译完成。这篇文章，从文化的角度来解读 Apache Way，并分析了中国文化是如何跟 Apache 的文化高度匹配和不那么匹配的地方。并从数据分析上给出了很多解释和总结，对于中国人理解什么是 Apache Way，如何结合 Apache Way 所代表的开源文化来加强和外部的协作，并发展自己的开源社区，从而发展开源事业非常有帮助。

--- 谭中意，ASF Member

**庄表伟**：Sharon 的论文非常有趣，也非常有启发。比如其中一个结论：**Apache 孵化器在将 Apache 文化植入中国贡献的项目中的成功程度时，作者发现我们已经采用了 "+1" 共识指标，并正在大量使用它。**这使我产生一些联想，中国网民的确已经大量使用了 "+1"，甚至 "+1024"、“+10086”也很常用。也许对于现代中国，接受外来文化，尤其是网络文化，本来就很容易吧。另外，Sharon 的研究，其实开启了一个非常重要的方向，在这个方向上，我期待很多的探索。例如：与中国开发者（Apache or 非 Apache）的深度访谈，从而获得更多的感性认识，并进而能够发现更多值得研究的文化差异、文化相似与文化融合的现象。更进一步值得研究的，则是这些异/同，对于开源项目的成功，有何影响？

--- 庄表伟，开源社理事、华为开源管理中心专家

**史少锋**：这是一篇非常有趣的论文；开始它非常细致地介绍了开源文化、开源的兴起，以及 Apache 孵化器如何诞生以及影响各个项目；更有趣的是论文介绍的几种工具和指标，例如小马系数等，用它们来分析中国贡献的项目和非中国贡献的项目，再结合中西方文化特点，你可以发现一些很有趣的现象。

作为 Apache Kylin 这个第一个中国人贡献的项目的参与者，同时也是 Apache 孵化器的导师，我有幸见证了中国开源力量的兴起及壮大；非常高兴看到这股趋势正在引起国际性的关注，也为我们中国的开发者感到自豪。希望未来中国开源项目和社区更加多样和活跃！

--- 史少锋，ASF Member，Apache Kylin PMC Chair

**王伟**：第一次看到这么系统地通数据与实证的方法研究 Apache 开源社区文化的研究论文，特别是里面所列举的来自中国发起的若干项目，以及这种中西方对比研究，极具意义。文中系统地介绍了通过八大指标来量化文化基线的方法，进而通过量化的方式来研究开源文化嵌入的作用，非常值得我们做开源研究的人学习。也希望我们能够在此基础上做出更多的开源研究工作。

--- 王伟，开源社副理事长，华东师范大学研究员

**王小瑞**：多年前和 Ted 认识，他一直很热心的帮助孵化开源项目，Ted 帮助 RocketMQ 团队与 Apache 董事会建立了桥梁，RocketMQ 今天成为流行的消息平台，离不开 Ted 的大力支持。他还积极的促进多个中国开源项目与 Apache 基金会的合作。今天很欣喜的看到他翻译的《The Transmission of Values and Culture in Open Source》中文版，本文从理论和数据的角度探讨了开源文化，特别是对中国本土开源项目的大量数据统计和理论分析。Ted 作为多年在开源领域的知名专家，对中西开源文化把握非常到位，全书翻译用词精准，是难得的开源领域佳作，特别推荐热爱开源的人士品读，可以极大提升对开源文化理解。

--- 王小瑞，ASF Member，Apache RocketMQ 创始人 | PMC Chair

**陈阳**：有幸跟作者 Sharon 在 COSCon '19 女性论坛上有过深入沟通，一起探讨关于开源生态中女性的话题，当时被 Sharon 的开源热情和能量所震撼。这次拜读 Sharon 的论文，感慨

终于有人把 Apache 之道为什么在中国能够生根发芽并发展壮大，从文化的角度分析得这么系统和透彻。我也特别有幸跟 Ted 共事，从 2015 年一起发起 Apache Roadshow 大会，到近年来 Apache 在中国的蓬勃发展的潮流中，我是近距离的旁观者，Ted 更是亲身实践者，长期贡献者和中西文化的桥梁。这一篇来自 Sharon 优秀的论文和 Ted 精心的翻译，是解读 Apache 之道的深度好文，透过现象看到本质，解答了很多年我一直在思考的问题，相信对你也一定有帮助。

--- 陈阳，开源社理事长，微软首席产品经理

**赵生宇**：个人虽然已经做了较长时间的开源研究，尤其在开源数据层面投入较多，这篇论文依然让人眼前一亮。在进行大量数据分析的同时，该论文旁征博引，对于开源尤其是 Apache 软件基金会中的文化要素也做了详尽的分析，是量化与实证研究结合的绝佳案例。也要感谢 Ted 对这篇论文的精心翻译，也正是像 Ted 一样一直践行开源精神，沟通中西文化的布道者，使得开源的理念得以更好的传播。

--- 赵生宇，开源社理事，开源博士

**堵俊平**：回想十几年前刚开始开源工作的时候，Apache 的项目和社区成为了我开源之路的领路人。那时候，我就有种感觉，在社区里开发代码，有种说不出的愉悦。直到后来，在一步步成长为项目的 committer, pmc 以至于 ASF Member 的时候，才深刻的理解：真正的参与开源，不仅意味着学习和贡献技术与代码，同时也意味着真心认同并推广开源的价值观与文化 - 这也是 Apache Way 如此吸引人的地方。非常欣喜的看到 Sharon 的论文，通过深入浅出、旁征博引，以量化和模型向读者展示了文化遗产对于开源社区的重要性。更让人感叹源于西方的开源运动与社区精神，竟然与中华文化所强调的集体主义精神不谋而合。这也完美的解释了以 “Apache Way” 为代表的开源文化为何能快速的席卷神州大地，落地生根。感谢 Ted 对这篇论文的推荐与翻译，他与开源社的小伙伴数年如一日、矢志不渝的分享与推广开源文化，为中西开源文化的交流做出了重大贡献，期待 Ted 和开源社更多的作品。

--- 堵俊平，ASF Member，华为计算开源总经理

**江波**：在本篇论文中，Sharon Foga 非常详细地为我们介绍了何为 Apache 之道以及其文化



兴起的历史渊源，对于我们更进一步理解开源文化和 the Apache Way 将有很大帮助。更有趣的是，在本篇论文完成的过程中 Sharon Foga 不仅结合了其作为 Apache 基金会董事的大量实践经验，也利用了诸如小马系数等定量的研究方法，通过理性的数据分析进行「感性」的文化研究。同时，这些指标公式和工具对于一个开源项目作者自我验证其社区成熟度，以及思考如何更好地发展社区也具有很强的指导意义。在文末，Sharon Foga 也给出了很多很有趣的课题，我也将择其一在我即将开启的 MBA 学习生涯中以深入研究。最后，感谢作者这篇极具启迪意义的论文，以及 Ted 和开源社同仁们在翻译、发布本篇译文的过程中所付出的辛勤劳动。Ted 是我进入开源世界大门的重要领路人，通过日复一日的开源布道与身体力行，让我领会到开源与开放式协作的魅力，带给我很多思考和无穷的快乐，也让我所在的 SegmentFault 公司找到新的业务增长点，借此机会向他表达我诚挚的谢意。

--- 江波，开源社理事、2022 年度执行长，SegmentFault 思否 COO，ALC Beijing Member

**郭雪：**2021 年第一次读这篇论文，当时还不知 Ted 要翻译成中文（早知道应该互通下阅读心得）。当时一直困惑自己的是开源文化是什么？为什么大家经常把开源文化归纳为“开放、协作、公开、透明”，依据是什么？开源文化到底是开源人的气质还是开源社区需要遵守的规则？所以看到这篇论文的时候非常激动，激动之处在于本文用客观的数据系统分析不同地域对于开源发展的影响以及 Apache 文化对中国贡献项目的影响。读完译文再次印证我国开源参与者多以集体为基础，同时注重群体和谐，利于形成包容型的开源社区，“独乐乐不如众乐乐”与“礼物文化”本质的相似性。推荐喜欢探究开源文化的同学阅读。

--- 郭雪，中国信通院云大所开源和软件安全部

**杨丽蕴：**当接到 Ted 的邀请，对这篇含金量非常高的论文做点评时，我个人是感到荣幸之余又充满好奇的。坦白的说，我个人与 Apache 软件基金会的交集不多，更多的是通过开源社和国内众多开源精英的口中认识 Apache，为什么大家对它的文化导向、治理模式和成效如此认可、自发传播，同时也对我国开源发展产生了广泛的影响？这是我非常好奇的地方。简单的看了一下论文的目录和前几部分内容，成体系也有深度，既有文化历史来源，也有严谨的数据分析和指标建模。我想，我对 Apache 的好奇之心可以从这篇论文中得到满足和启发。同时，呕

心沥血的写和译，传播和共享知识是一件非常不容易的事情，向作者和译者致敬！

--- 杨丽蕴，中国电子技术标准化研究院软件中心研究室主任

**姜宁：**也许是因为文化对大家的影响是潜移默化的，也许是因为国内参与开源的人大多是理工科背景的原因，也许是因为国内刚刚开始大规模重视开源。在国内的开源圈很少看到大家对开源以及东西文化的差异分析和讨论。非常高兴能够接到 Ted 的邀请对 Sharon Foga 的这篇论文进行点评。Apache 软件基金会（ASF）所奉行的 Apache Way 是对健康可持续发展的开源社群建设经验的高度概括，ASF 孵化器则为众多的新项目提供了一个学习和实践 Apache Way 的场所。在我参与孵化的众多来源自中国项目的过程中，我能深刻体会到传统的中国文化与开源文化的冲突，特别是在大家处理冲突和达成共识层面还有很多需要提升的地方。Sharon Foga 的论文中对中西文化多维度的分析，以及通过邮件列表所做定量分析向我们揭示了开源文化在源自中国项目之间传递的过程，这对我们进一步分析开源文化和中国文化的差异，以及进一步在国内推广开源文化，都提供宝贵的参考建议。

--- 姜宁，Apache 软件基金会 2022 年董事，ASF Member，ALC Beijing 发起人，华为开源管理中心技术专家

## 参考资料：

- [霍夫斯泰德文化维度理论](#)
- [Review Kibble Against CHAOSS Metrics](#)
- [Extended Business Project-Open Source Culture](#): 2019 年 1 月 10 日
- Sharon Foga 演讲集（部分）
  - COSCon'19
    - [主题演讲：The Cultural Heart of Apache: The Apache Way](#)
    - [女性参与开源分会场：开源生态中的女性力量](#)
    - [开源治理分会场：What is Community Development?](#)
  - ApacheCon Asia 2021
    - [ApacheCon Asia 2021 Keynote: Sharan Foga - Welcome Diversity 拥抱多样性](#)
    - [ApacheConAsia2021 开源社区 0808：Sharan Foga - 开源共同体健康报告 101](#)
- [Apache 孵化器指南](#)
- [Apache 项目毕业指南](#)

● ASF 源自中国之项目列表 (30)

○ [ASF 源自中国之顶级项目列表](#) (19 : 按成为 TLP 时间倒推排序)

序号	项目名称	开始孵化日期	成为 TLP 日期	备注
1	<a href="#">Apache ShenYu</a>	2021/05/03	2022/07/28	
2	<a href="#">Apache Doris</a>	2018/07/18	2022/06/15	
3	<a href="#">Apache InLong</a>	2019/11/03	2022/06/15	原 TubeMQ
4	<a href="#">Apache DolphinScheduler</a>	2019/08/29	2021/04/08	
5	<a href="#">Apache ECharts</a>	2018/01/18	2020/12/16	
6	<a href="#">Apache Ozone</a>	2018/11/22	2020/10/21	
7	<a href="#">Apache IoTDB</a>	2018/11/18	2020/09/17	
8	<a href="#">Apache APISIX</a>	2019/10/17	2020/07/15	
9	<a href="#">Apache ShardingSphere</a>	2018/11/10	2020/04/16	
10	<a href="#">Apache Submarine</a>	/	2019/10/16	Hadoop spin-off
11	<a href="#">Apache Dubbo</a>	2018/02/16	2019/05/15	
12	<a href="#">Apache Skywalking</a>	2017/12/08	2019/04/17	
13	<a href="#">Apache Griffin</a>	2016/12/05	2018/11/21	
14	<a href="#">Apache ServiceComb</a>	2017/11/22	2018/10/17	
15	<a href="#">Apache HAWQ</a>	2015/09/04	2018/8/15	
16	<a href="#">Apache RocketMQ</a>	2016/11/21	2017/9/20	
17	<a href="#">Apache CarbonData</a>	2016/06/03	2017/04/19	
18	<a href="#">Apache Eagle</a>	2015/10/26	2016/12/21	
19	<a href="#">Apache Kylin</a>	2014/11/25	2015/11/18	

○ [ASF 源自中国之孵化项目列表](#) (11, 按孵化开始时间倒推排序)

序号	项目名称	开始孵化日期	成为 TLP 日期	备注
1	<a href="#">Apache Uniffle</a>	2022/06/06		
2	<a href="#">Apache DevLake</a>	2022/04/29		
3	<a href="#">Apache Kvrocks</a>	2022/04/23		
4	<a href="#">Apache HugeGraph</a>	2022/01/23		
5	<a href="#">Apache SeaTunnel</a>	2021/12/09		
6	<a href="#">Apache Linkis</a>	2021/08/02		
7	<a href="#">Apache Kyuubi</a>	2021/06/21		
8	<a href="#">Apache EventMesh</a>	2021/02/18		
9	<a href="#">Apache Pegasus</a>	2020/06/22		
10	<a href="#">Apache Teaclave</a>	2019/08/19		原 MesaTEE
11	<a href="#">Apache brpc</a>	2018/11/13		

○ [ASF 源自中国之退休项目列表](#) (1, 按退休开始时间倒推排序)

序号	项目名称	开始孵化日期	退休日期	备注
1	<a href="#">Apache Weex</a>	2016/11/30	2021/05/12	

# 目录

第 1 章 绪论	12
1.1 背景 .....	12
1.1.1 文化的定义 .....	13
1.2 研究的动机 .....	14
1.3 研究的问题 .....	14
1.4 研究的重要性 .....	15
第 2 章 参考文献	16
2.1 国家层面的文化要素 .....	16
2.2 开源文化 .....	19
2.3 ASF 文化，愿景与价值观 .....	21
2.3.1 Apache 之道的实践 .....	23
2.4 中国文化与开源 .....	24
2.5 文化的变迁 .....	27
2.6 Apache 孵化器 .....	28
2.6.1 角色和职责 .....	28
2.6.2 Apache 项目成熟度模型 .....	29

## 第 3 章 研究方法

36

3.1 方法论和假设 .....	36
3.2 研究设计 .....	37
3.3 环境 .....	38
3.4 研究工具和指标 .....	38
3.4.1 Apache Kibble .....	39
3.4.2 小马系数 Pony Factor .....	39
3.4.3 增强型的小马系数 Augmented Pony Factor .....	40
3.4.4 元小马系数 Meta Pony Factor .....	40
3.4.5 小马系数、增强型小马系数、元小马系数的范例 .....	40
3.4.6 感性分析 .....	42
3.4.7 关键句提取 .....	43
3.4.8 贡献者留存 .....	43
3.5 数据采集 .....	44
3.5.1 中国发起的孵化项目 .....	46
3.5.2 非中国发起的孵化项目 .....	46
3.5.3 绕过孵化的非中国项目 .....	46

## 第 4 章 数据分析

47

4.1 Apache HTTP 服务器项目的文化基准线 .....	47
4.1.1 标志 1 : 做为基准线之代码库的小马系数 .....	47
4.1.2 标志 2 : 做为基准线之邮件列表的小马系数 .....	48
4.1.3 标志 3 : 做为基准线之贡献者经验 .....	48
4.1.4 标志 4 : 做为基准线之代码库的贡献者留存 .....	49
4.1.5 标志 5 : 做为基准线之留存贡献者的邮件列表统计 .....	50
4.1.6 标志 6a : 2019/4/30 做为基准线之情感分析 .....	50
4.1.7 标志 6b : 2019/4/30 做为基准线之情感分析比较 .....	51
4.1.8 标志 7 : 做为基准线之随时间变化的情感 .....	51
4.1.9 标志 8 : 做为基准线之关键词句的提取 .....	52
4.2 非中国发起的孵化项目 .....	53
4.2.1 标志 1 : 非中国发起的孵化项目的代码库 PF 值 .....	53
4.2.2 标志 2 : 非中国发起的孵化项目的 Email PF 值 .....	54
4.2.3 标志 3 : 非中国发起的孵化项目的贡献者经验 .....	55
4.2.4 标志 4 : 非中国发起的孵化项目代码库的贡献者留存量 .....	56
4.2.5 标志 5 : 非中国发起的孵化项目之留存贡献者的邮件列表统计 .....	57
4.2.6 标志 6a: 2019/4/30 的非中国发起的孵化项目之情感分析 .....	57
4.2.7 标志 6b: 2019/4/30 的非中国发起的孵化项目之情感分析 <b>比较</b> .....	58
4.2.8 标志 7: 非中国发起的孵化项目之随时间推移的情感分析 .....	59
4.2.9 标志 8: 非中国发起的孵化项目之关键词提取 .....	59
4.3 中国发起的孵化项目 .....	61
4.3.1 标志 1 : 中国发起的孵化项目的代码库 PF 值 .....	61



4.3.2 标志 2 : 中国发起的孵化项目的 Email PF 值 .....	61
4.3.3 标志 3 : 中国发起的孵化项目的贡献者经验 .....	62
4.3.4 标志 4 : 中国发起的孵化项目代码库的贡献者留存量 .....	63
4.3.5 标志 5 : 中国发起的孵化项目之留存贡献者的邮件列表统计 .....	64
4.3.6 标志 6a: 2019/4/30 的中国发起的孵化项目情感分析 .....	64
4.3.7 标志 6b: 2019/4/30 的中国发起的孵化项目情感分析 <b>比较</b> .....	65
4.3.8 标志 7: 中国发起的孵化项目随时间推移的情感分析 .....	66
4.3.9 标志 8: 中国发起的孵化项目的关键词提取 .....	66
4.4 绕过孵化的非中国项目 .....	67
4.4.1 标志 1 : 绕过孵化的非中国项目之代码库的小马系数 .....	67
4.4.2 标志 2 : 绕过孵化的非中国项目之邮件列表的小马系数 .....	68
4.4.3 标志 3 : 绕过孵化的非中国项目之贡献者经验 .....	69
4.4.4 标志 4 : 绕过孵化的非中国项目之代码库的贡献者留存 .....	69
4.4.5 标志 5 : 绕过孵化的非中国项目之留存贡献者的邮件列表统计 .....	70
4.4.6 标志 6a: 2019/4/30 绕过孵化的非中国项目之情感分析 .....	71
4.4.7 标志 6b: 2019/4/30 绕过孵化的非中国项目之情感分析 <b>比较</b> .....	72
4.4.8 标志 7: 绕过孵化的非中国项目之随时间推移的情感分析 .....	72
4.4.9 标志 8: 绕过孵化的非中国项目之关键词提取 .....	73

## 第 5 章 分析结果 75

5.1 Apache HTTP 服务器项目之文化基准线 .....	76
5.1.1 小马系数 .....	76
5.1.2 贡献者留存 .....	76
5.1.3 情感分析 .....	77
5.1.4 关键语句分析 .....	77

5.2 中国发起的孵化项目 vs. 非中国发起的孵化项目 .....	77
5.2.1 小马系数 .....	78
5.2.2 贡献者留存 .....	78
5.2.3 情感分析 .....	79
5.2.4 关键语句分析 .....	80
5.3 非中国发起的非孵化项目 vs 中国发起的孵化项目 .....	80
5.3.1 小马系数 .....	80
5.3.2 贡献者留存 .....	81
5.3.3 情感分析 .....	82
5.3.4 关键语句分析 .....	83
5.4 非中国发起的孵化项目 vs 非中国发起的非孵化项目 .....	83
5.4.1 小马系数 .....	83
5.4.2 贡献者留存 .....	84
5.4.3 情感分析 .....	84
5.4.4 关键语句分析 .....	85
<b>第 6 章 结论</b> .....	<b>86</b>
6.1 对研究问题的回应 .....	88
6.1.1 我们能找到哪些文化嵌入的证据? .....	88
6.1.2 中国贡献的项目和非中国贡献的项目之间有多大的文化差异? .....	88
6.1.3 Apache 孵化器在将 Apache 文化植入中国贡献的项目中的成功程度如何? .....	89
6.2 摘要 .....	89
<b>参考资料</b> .....	<b>93</b>

## 插图列表

2.1 Hofstede 文化维度：中国 .....	17
2.2 Hofstede 文化维度：美国 .....	18
2.3 Hofstede 文化维度：德国，荷兰，英国 .....	18
2.4 Apache 项目成熟度模型 .....	30
4.1 Apache HTTP 服务器项目做为基准线之代码库的小马系数变化 1996 - 2019 年 .....	48
4.2 Apache HTTP 服务器项目做为基准线之邮件列表的小马系数变化 1996 - 2019 年 .....	48
4.3 Apache HTTP 服务器项目做为基准线之贡献者经验 .....	49
4.4 Apache HTTP 服务器项目做为基准线之代码库的贡献者留存 1996 - 2019 年 .....	49
4.5 Apache HTTP 服务器项目做为基准线之留存贡献者的邮件列表统计 1996 - 2019 年 .....	50
4.6 Apache HTTP 服务器项目做为基准线之情感分析 2019/4/30 .....	50
4.7 Apache HTTP 服务器项目 vs ASF 相对情感 .....	51
4.8 Apache HTTP 服务器项目做为基准线之情绪分析 2018/11-2019/4	52
4.9 <b>Apache HTTP 服务器项目做为基准线之关键词提取 2019/4/30</b> .	52
4.10 <b>非中国发起的孵化项目之代码库的小马系数 2012-2019</b> .....	53
4.11 <b>非中国发起的孵化项目之邮件列表的小马系数 2012-2019</b> ....	54
4.12 <b>非中国发起的孵化项目之贡献者经验</b> .....	55

4.13 非中国发起的孵化项目之代码库的贡献者留存 2012-2019 .....	56
4.14 非中国发起的孵化项目之留存贡献者的邮件列表统计 2012-2019 .	57
4.15 非中国发起的孵化项目之情感分析 2019/04/30 .....	58
4.16 非中国发起的孵化项目与 ASF 的情绪对比 .....	58
4.17 非中国发起的孵化项目之随时间推移的情感分析 2018/11-2019/4	59
4.18 非中国发起的孵化项目之关键词句提取 2019/04/30 .....	60
4.19 中国发起的孵化项目之代码库的小马系数 2011-2019 .....	61
4.20 中国发起的孵化项目之邮件列表的小马系数 2015-2019 .....	62
4.21 中国发起的孵化项目之贡献者经验 .....	62
4.22 中国发起的孵化项目之代码库的贡献者留存 2011-2019 .....	63
4.23 中国发起的孵化项目之留存贡献者的邮件列表统计 2012-2019 .	64
4.24 中国发起的孵化项目之情感分析 2019/04/30 .....	65
4.25 中国发起的孵化项目与 ASF 的情绪比较 .....	65
4.26 中国发起的孵化项目之随时间变化的情感分析 2018/1-2019/4 .	66
4.27 中国发起的孵化项目之关键词语提取 2019/04/30 .....	67
4.28 绕过孵化的非中国项目之代码库的小马系数 2012-2019 .....	68
4.29 绕过孵化的非中国项目之邮件列表的小马系数 2015-2019 ....	68

4.30 绕过孵化的非中国项目之贡献者经验 .....	69
4.31 绕过孵化的非中国项目之代码库的贡献者留存 2011-2019 .....	70
4.32 绕过孵化的非中国项目之留存贡献者的邮件列表统计 2012-2019	71
4.33 绕过孵化的非中国项目之情感分析 2019/4/30 .....	71
4.34 绕过孵化的非中国项目与 ASF 的情绪比较 .....	72
4.35 绕过孵化的非中国项目之随时间变化的情感分析 2018/11-2019/4	73
4.36 : 绕过孵化的非中国项目之关键词句提取 2019/04/30 .....	73

# 第 1 章 绪论

## 1.1 背景

Apache 软件基金会 ( ASF ) 是一个 501(c)(3) 慈善组织，成立于 1999 年，是在一个单一开源项目的基础上成立的基金会，该项目称为 Apache HTTP Server 项目。创建一个法律实体的关键目标之一是确保所有 ASF 项目在原项目创建者或个人志愿者参与后继续存在。所选择的实体类型，即 501(c)(3)，与其他许多基金会不同，因为它更符合 ASF 为公众利益创造软件的使命。

作为创建 Apache HTTP Server 项目的一部分而发展起来的价值观、知识和管理模式是 ASF 文化的来源，也被称为 "Apache 之道 ( Apache Way ) "。

Apache 之道依赖于一个核心原则--任人唯贤，而这一点已经嵌入到 ASF 的所有层面。从正式的管理、董事和成员的选举，一直到项目社区和对个人贡献的认可，功绩都起着至关重要的作用。

在撰写本报告时，ASF 由 350 多个项目和软件计划组成，据称这些项目都展示并接受 "Apache 之道" 作为其文化模式。

2015 年，Apache Kylin 成为第一个由中国发起和贡献的项目，成功完成了在 ASF 的孵化，毕业并成为 Apache 软件基金会的一个顶级项目。

作为一个在美国成立的全球性开源基金会，这对 ASF 来说是一个重要事件。为什么呢？因为成为顶级项目 ( TLP ) 的毕业过程的一个关键部分涉及了对 ASF 价值和文化的理解、接受和展示。

那么，当一个具有中国文化背景的项目接触到不同的环境、文化、语言和行为规范时会发生什么？在 Apache Kylin 和其他后来由中国人发起的项目中，仅仅作为一个顶级项目 ( TLP ) 毕

业，不一定能证明对另一种文化的长期接受或嵌入。

本研究的重点是探索四个已经或正在接受 ASF 孵化的中国贡献者的项目，并寻找 Apache 软件基金会孵化过程中文化嵌入的定性和定量的证据。

### 【专家点评】堵俊平

开源是一门实践科学。孵化过程是“术”，文化嵌入是“道”。从源于实践的“术”来探究“道”，原本就符合实事求是的科学精神。

## 1.1.1 文化的定义

我们接受所谓“文化”的存在，但找到一个单一的文化定义一直是个问题。也许它只是由我们周围的环境创造的？(Hofstede et al, 2015)。

文化可以影响我们的日常决定和行为，因为我们在与他人互动时，会带着我们的文化身份。这意味着我们展示的价值观会影响我们周围的人。(Allen, 2003)。

如果文化是一种无言的交流 (Guiso, Sapienza & Zingales, 2015)，那么它将依赖于观察行动中的文化行为，但无法与之互动以了解行为背后的原因。为了适应并融入，结果可能是导致简单地模仿你所看到的東西，而文化不仅仅是复制他人的行为，还需要有互动。

如果文化是在我们参与的社会群体中发展起来的東西，那么我们已经参与社会的事实就意味着一些共同的群体特征 (Seidler, 2010)。那么，不具备这些特征的人如何能够参与到不同的文化中去呢？

如果文化是关于我们如何与周围的人互动和分享信息 (Hofstede 等人, 2015)，那么就必须有一种沟通的方式，将关键的价值观和行为传递给他人。每当我们互动时，我们都会用我们的文化身份影响他人。(Allen, 2003)

本研究的重点是 12 个 ASF 项目，其中 4 个是中国人发起的，这些项目都经过了孵化，看看在项目演化和毕业到顶级项目的过程中是否发生了任何文化变化。

### 【专家点评】堵俊平

“文化”的定义千人千面，但文化的本质特征——群体性以及传染性，相信绝大多数人都会否认。抓住这两个特征，我们去评估文化在客观世界中发挥的作用也就有了依据。

## 1.2 研究的动机

这项研究的目的是记录 Apache 孵化器的过程是如何试图将被称为 "Apache 之道" 的开源文化嵌入到似乎来自不同文化的新项目中。

## 1.3 研究的问题

1. 我们能找到哪些文化嵌入的证据？
2. 中国贡献的项目和非中国贡献的项目之间有多大的文化差异？
3. Apache 孵化器在将 ASF 文化嵌入中国贡献的项目中的成功程度如何？

## 1.4 研究的重要性

本报告中的案例研究被用来探讨几个项目在孵化周期的不同阶段文化嵌入的过程。它混合使用了历史、回顾和当前的实时信息，试图捕捉文化演变的过程。

这项工作将有助于 Apache 软件基金会了解其文化嵌入的流程和 Apache 孵化器的有效性。



## 第 2 章 参考文献

### 2.1 国家层面的文化要素

文化是我们头脑中的软件，帮助我们与周围的人互动和分享信息（Hofstede 等人，2015）。世界是一个很大的地方，并非所有的人都在相同的环境中长大，这意味着我们在文化上有所差异。因此，使用 Hofstede 的计算机比喻，我们的软件版本可能有兼容性问题，因为我们不理解或误解了他人的文化行为和信号。

界定国家层面的文化差异的最著名的框架是由 Geert Hofstede 创作的。他的研究强调了六个关键领域或维度，可用于评估一种文化的基本结构。这六个维度是：

#### 1. 权力距离指数（PDI）

Power Distance Index (PDI)

权力距离指数（power distance index，缩写为 PDI）：指在家庭、公司、小区等组织机构中地位较低的成员对于权力分配不平等的接受程度。在权力距离指数高的社会，地位较低的成员更倾向于服从地位较高的成员的命令，而同样的情形在指数低的社会，则需要合理化命令

#### 2. 个人主义与集体主义（IDV）

Individualism vs. Collectivism (IDV)

与集体主义相对，指个人融入集体的程度

#### 3. 阳刚气质对柔性气质（MAS）

Masculinity vs. Femininity (MAS)

指人们（不论男女）注重成就与完成任务，更富有竞争精神，自信与野心，注重财富和社会资源的积累，而女性化社会则注重人际关系，重视合作，照顾弱者和生活的质量

#### 4. 不确定性规避指数（UAI）

Uncertainty Avoidance Index (UAI)

指社会能在多大程度上容忍未来的不确定性。对不确定性规避指数高的社会，会更努力想控制未来的不确定性，习惯照章办事，对非正规的行为诸多限制，在宗教与哲学上倾向相信绝对真理与完整理论；而规避指数低的社会则对变化更顺其自然，习惯变通办事，对非正规的行为不多加限制，在宗教与哲学上倾向相信真理是相对的与经验主义

## 5. 长期取向与短期取向 ( LTO )

### Long Term Orientation vs Short Term Orientation (LTO)

长期导向指生命中的大事发生在未来而不是正在或已经发生，好人善于适应环境，善恶视环境而定而非有普世标准，传统可以随环境改变而非神圣不可更动，重视勤俭储蓄，以及有羞耻心

## 6. 纵容与克制 ( IND )；

### Indulgence vs. Restraint (IND)

指社会成员在多大程度上意图控制自身的欲望

每个维度都描述了可能发生文化差异的一个具体面向。

在这项研究中，我们希望初步确定中国（作为发起 ASF 项目的贡献者的来源）与美国（Apache 软件基金会合法注册地）之间的文化差异。

如果我们看一下 Hofstede 关于中国的数据：

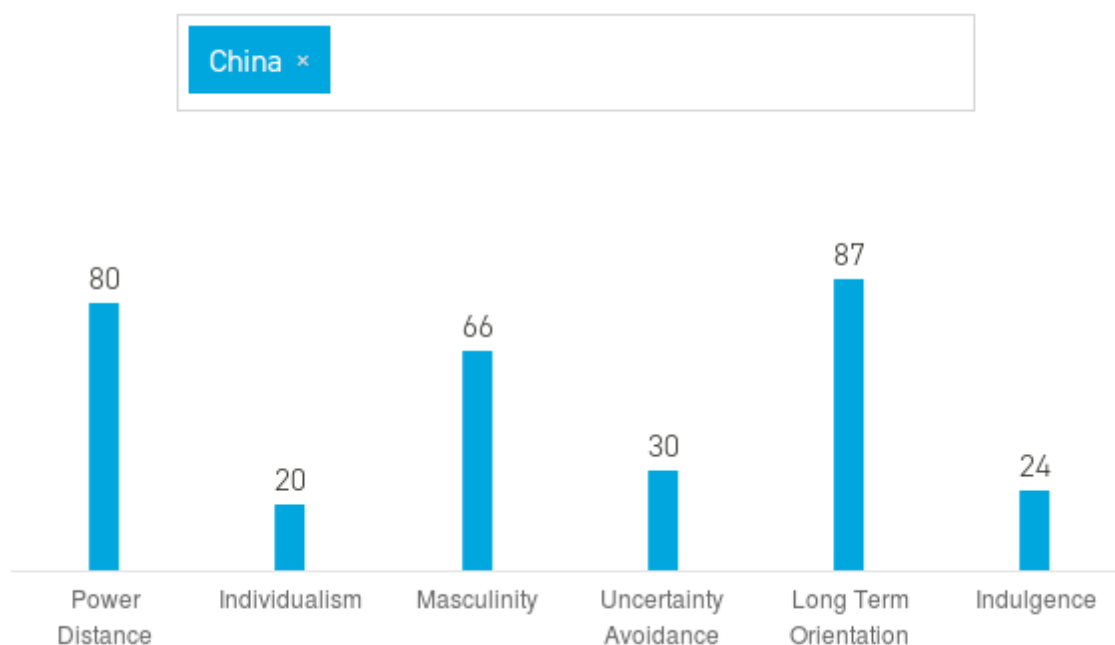


图 2.1 : Hofstede 文化维度：中国

并与美国进行比较，因为 ASF 是在美国创建的基金会。

我们看到，两个最明显的文化差异体现在权力距离（PDI）和个人主义（IDV）两个维度。除了美国之外，ASF 创始团队（Apache 小组）的成员还来自德国、英国和荷兰。

如果我们看一下这些国家的 Hofstede 文化维度，我们会看到与美国类似的模式。因此，在这项研究中，我们将重点关注这两个维度 PDI 和 IDV 这两个维度来描述 Apache 软件基金会开源背景下的文化差异。

权力距离的重点是表明社会中个人之间权力或影响的平等。它还强调了他们对这种差异的接受程度。简而言之，你在接受社会中其他人有权力影响你的生活方面有多舒服。高分强调人们普遍期待并接受权力的不平等，而低分则表示相反。

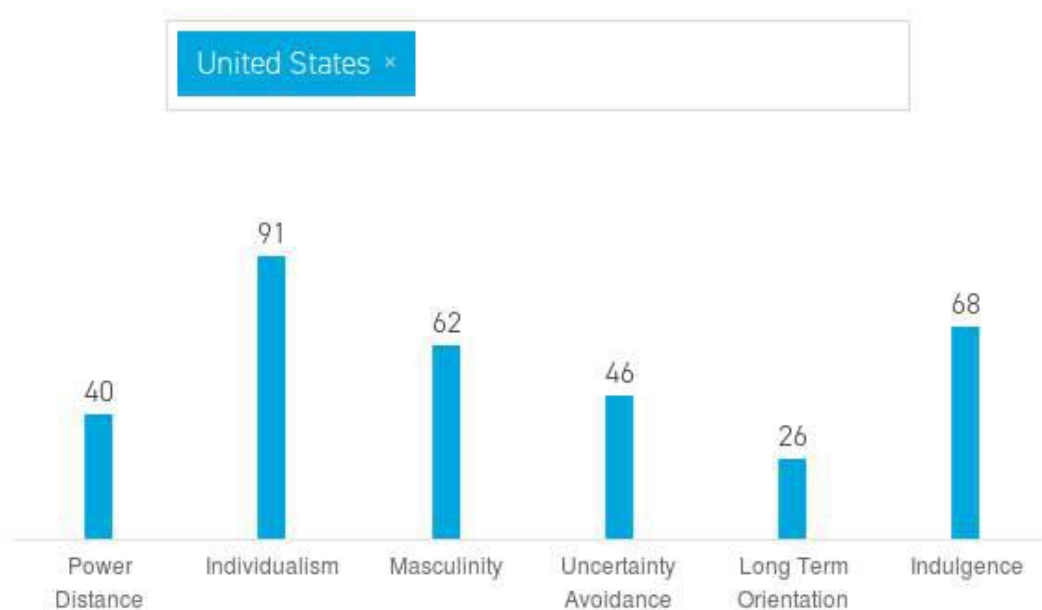


图 2.2: Hofstede 文化维度：美国



图 2.3 : Hofstede 文化维度：德国、荷兰、英国

作为一个国家，与美国的 40 分相比，中国的权力距离得分很高（80 分），表明中国是一个认为人们之间的不平等是可以预期和接受的社会。因此，我们可以看到，中国的文化期望是几乎不期望拥有影响力的权力，而美国的文化期望则相反，是期望拥有影响力的权力。

如果我们回到之前提到的 ASF 的文化宗旨，即任人唯贤，让每个人都有潜在的影响力。

赋予每个人潜在影响的权力，那么我们可以看到，这一价值观与美国的得分更为一致。

一个社会可以是个人主义的，人们专注于独立，照顾自己，很少或不依赖他人；或是集体主义的社会，即有强烈的团体关注，团体的需求超过任何个人需求。

作为一个国家，与美国的 91 分相比，中国的个人主义得分较低（20 分），表明中国是一个以群体文化为基础的集体主义社会，而不是个人主义。因此，我们可以看到，中国的文化期望是以团体为基础，有共同的目标，而不是专注于任何一个人。然而，美国的文化期望是个人比群体更重要。

回到 ASF 组织项目的方式，这是围绕着社区进行的，正是社区确保了项目的持久性。在此基础上，我们可以看到，这更符合集体主义思想，因此，中国人的得分也更高。

这似乎突出了一个悖论，尽管 "Apache 之道" 创立于美国，但它并不完全符合美国文化，而是包含了美国和中国文化的各个面向。

### 【专家点评】

**陈阳：**在这里拍案叫绝，中国的社区的开源文化其实是比较集体主义，而不是推崇个人主义。在开源社的社区运营中也观察到这一现象，中国的社区成员更愿意贡献到这个组织中，一起设定共同目标，打造共同的平台，大家追求在这个组织中的被认同感。在中国开源社区鲜有打造个人品牌的案例。

**堵俊平：**这里很好的解释了 Apache Way 中的 “社区大于代码”（Community Over Code）为何在中国这么容易被接受。中国文化中重视集体的一面，让社区的贡献者更容易摆脱单打独斗，去追求社区的集体认同，这是我们的一大优势。但另一方面，我们传统文化中的中规中矩、缺乏张扬的个性，对于年轻一代开发者的吸引、对于大破大立的技术创新，还是有劣势的。如何扬长避短，将中国的开源项目持续推向世界顶峰，是中国全体开源从业者需要集体深思的问题。

**姜宁：**在费孝通的《乡土中国》中，对中国 “差序格局” 和西洋的 “团体格局” 有着非常精辟的论述。中国人在处理团体和个人之间的关系的时候更讲究的是 “收放自如” 的自我主义。中国人强调的是熟人社会的协作，把陌生人变成 “自己人” 是人脉建设的重要组成部分。这和基于法律规则形成的弱关系群体社会来说就有很大的差别。开源社群正好是由一群未曾谋面基于网络连接在一起协作的人群，在 “差序格局” 下，如何让这些网络上的陌生人连接在一起，建立信任，相互协作是一个很有挑战的研究问题。

## 2.2 开源文化

开源文化真的存在吗？”“开源”这个词被用来描述政治、新闻、生物、音乐和教育（Ceraso & Pruchnic, 2006）。那么，为什么开发软件并将其免费提供会突然让它成为一种文化？开源文化是否可能只是另一个受欢迎的流行语，它背后几乎没有任何实质内容。

与其说是一种独特的文化，我们是否能说开源软件的开发只是一种利用自愿提供时间和精力的人的劳动能力的方式（Ceraso & Pruchnic, 2011）。如果开源纯粹集中在编写和生产代码上，那么仅仅这一点还不足以将开发过程定义为一种文化。

是不是因为人们习惯于在工作环境中被组织和管理？在参与开源的过程中，人们把“工作文化”这个术语转换成“开源文化”来描述他们没有报酬的工作环境？如果这仅仅是关于生产恰好是开源的软件，那么，这并不是什么新鲜事，因为软件开发从 1960 年代就开始了，从来没有人称它为文化，为什么开源会有什么不同呢？

1984 年，Richard Stallman 创立了自由软件基金会（FSF）。它的第一个目标是创建 GNU 许可证。一些人认为这一行动是出于政治动机（Kelty, 2004），这种对知识产权的关注促进了需要为之奋斗的意识形态（Carillo & Okoli, 2009）。

GNU 许可证为自由和开源软件提供了一个法律定义和保护。此外，自由软件基金会（FSF）的口号“如言论般的自由，而非如啤酒般的免费（Free as in Speech, not as in Beer）”在自由和开源软件与发表意见的权利之间提供了明确的联系。

Stallman 本人明确区分了开源和自由软件，他强调开源是一种开发方法，而自由软件是一种社会运动。（Zhou, Y. 2011）这里我们有一个明确的区分，只有自由软件才被描述为社会性的，而社会互动是传播文化的一个关键方面。

1991 年，Linus Torvalds 创建了 Linux 内核，当时人们认为它的动机是对现有商业市场的专利软件挑战（Kelby, 2004）。有一个开源的软件替代现有的操作系统是很重要的，也是革命

性的。这一事件影响了社会。我们再次有了一场社会运动，但这次是竞争性的，是针对现有软件公司的。

【专家点评】庄表伟

事实上，Linux 本人的动机，并不是要去挑战，参见《Just for Fun》

在 90 年代中期，世界上最流行的网络服务器软件（HTTP 守护程序）是免费和开源的。它是由伊利诺伊大学国家超级计算应用中心（NCSA）的 Rob McCool 开发的。当 Rob 离开 NCSA 后，HTTP 守护程序的开发和维护就停止了。

由于许多人仍然在积极使用它，所以一小群用户没有放弃这个项目，而是决定一起工作，提供修复和扩展。这个小组被称为 "Apache 小组"，由于他们的合作，在不到一年的时间里，他们发布了他们自己版本的 HTTP 守护程序，并将其称为 Apache HTTP 服务器，或者更常见的 "Apache Web 服务器"。我们再一次看到社会互动被用于创造工作和协作方式的基础。

虽然这些开源运动都是由不同的事件触发的，但有一个共同的因素，即社会联系是一个重要的因素。因此，触发因素和环境会对由此产生的行为、价值观和交流机制产生影响。而这些正是通常用来描述和定义文化的。

【专家点评】堵俊平

开源项目的成功，离不开很多偶发因素，这其中“人”的因素格外重要。与其临溪羡鱼，不如退而结网。我们在膜拜这些开源前辈和大神的同时，将开源优秀的文化传承下来会显得更有意义。

## 2.3 ASF 文化，愿景与价值观

如前所述，许多开源项目是作为基金会或非营利实体创建的。虽然有些基金会专注于为他们的股东或成员们提供利益，但非营利组织的使命是专注于为公众提供利益。为了招募新的贡献者，为了帮助实现他们的使命，开源项目需要提供一些独特的东西来吸引贡献者并鼓励他们参与。文化可以作为吸引和留存贡献者的一种方式。

ASF 有自己独特的文化，叫做 "Apache 之道 ( Apache Way )"。它是一种直接由于建立 Apache HTTP 服务器项目而产生的文化，并随着时间的推移而不断发展。

当 ASF 在 1999 年成立的时候，一个关键的目标是确保所有的 ASF 项目在个别志愿者参与之后还能继续存在。这种情况与 NCSA HTTP 守护程序直接相关，在该项目中，最初的开发者离开了该项目，而 Apache 小组则一起努力恢复、维护并成功地发展了该项目。

ASF 是基于任人唯贤的模式，即承认个人的贡献。每项贡献都能获得 "功绩"，随着个人功绩的积累，他们会得到奖励，被允许承担更多责任。另一个关键概念是，功绩永远不会过期，这意味着贡献者将永远保持对他们所做工作的认可。

任人唯贤也被称为 "干活的民主"，因为做得多的人可以获得更多的责任。这意味着影响项目的权力在于那些积极贡献的人。

#### 【专家点评】

堵俊平

任人唯贤、知易行难。Apache Way 创造性的把同行评议与公开透明结合了起来，让 "任人唯贤" 真正有了可操作性。具备现实意义的 "理想" 自然也就成为了众人的信仰。

姜宁

面向公众的开发需要建立起一套公平的体制，任人唯贤中的贤就是让大家的功绩 ( Merit ) 能够被众人所看到，让干活的人安心干活，而不用绞尽脑汁去想如何 "包装"，如何 "呈现"。因为在权力分配的时候，大家是靠贡献说话，而不是靠距离权力的远近。这种体制可以极大激励积极贡献的志愿者，从而推动社群建设。

ASF 准则指出：

**"ASF 是由个体成员构成的"**

对于那些刚接触 ASF 文化的人来说，这个概念可能会让他们感到困惑，因为对个人的提及并



不是在提倡个人主义。相反，它意味着每个贡献者不仅是 ASF 的一部分，而且他们也因此得到认可。

需要注意的一个关键点是，只承认个人功绩而非公司功绩，这意味着公司不能获得功绩和影响的权力。这有助于保持 ASF 的厂商中立环境。

#### 【专家点评】江波

开源社成立于 2014 年，由几家公司共同发起成立，并在 2017 年转型成为完全由个人成员组成的开源社区联合体。开源社的定位、厂商中立的原则以及社区的治理模式都很大程度上参考了 ASF 的运作，并在 Apache Way 的指引下链接了数万名开源人，集聚了上千名社区成员、志愿者和海内外数百位讲师。

ASF 还提倡：

#### **"社区重于代码"**

这意味着维护项目社区的繁荣比生产软件更重要。对于那些目标是为公众利益生产可用软件的项目来说，宣传软件开发不是主要的优先事项，因为维持项目的是社区，所以拥有一个繁荣和活跃的社区有助于确保项目在未来的生存。

### 2.3.1 Apache 之道的实践

“假设”可以被描述为深植于组织中的无意识的行为和价值观。作为组织文化的一部分，这些行为发生得如此自然，以至于他们甚至没有意识到它们。在许多情况下，只有那些与文化无关的外部人员才能识别或认识到这些“假设”的展现。

所有的开源项目都是不同的，正是这些固有的假设将形成每个项目的文化核心。以下是观察到的 Apache Way 的一些潜在的固有假设。

- 共识

- 决策是通过协商达成共识作出的。
- 达成共识不是透过投票，而是透过表明一种意见、支持，或者被用作在意见分歧后创造一条前进道路的方法。
- 任何属于 ASF 社区的人都有影响的力量。
- 从社区中获得共识比投票更好，并减少了 "他们" 对抗 "我们" 的行为。

【专家点评】

堵俊平：共识是决策民主化、科学化的体现。在开源领域，“共识”机制真正释放了社区的活力，让每个人都能平等的参与到社区建设中来。

- 社区自我纠错

- 允许每个项目社区以独立的方式发展与维护自我。
- 如果发生负面的事情，允许社区自我纠正。
- ASF 董事会的干预是最后的解决方法。

- 社区重于代码

- 拥有一个协作的社区比拥有最好的软件和一个不正常的社区更重要。
- 即使双方观点相反，仍可进行尊重的互动和协作。

- 不同的身份和角色

- 对不同观点的理解。
- 商业与雇员的角度。
- 社区与个人的角度。
- 在 ASF 中的不同角色，需要澄清在特定情况下戴的是哪一顶 "帽子"（如董事会成员、PMC 主席、成员、提交者、个人）。

- 作为真相之源的邮件列表
  - 如果它不在邮件列表中，就说明它没有发生！
  - 激进的透明度 - 讨论是公开的，包括整个社区。
  - 决策需要公开进行。
- 功绩
  - 每个人都有能力获得功绩。
  - 表彰功绩。
  - 功绩永不过期。

#### 【专家点评】

江波：社区的核心是共识。如今有各种各样的组织和平台被称之为「社区」，不仅仅限于开源软件社区，还包括如 WeChat / Slack 社群，UGC 内容平台，如：小红书、大众点评、即刻，以及技术领域的 SegmentFault 思否、Stack Overflow 等等。在我们看来，一个社区的三要素为平台、用户和共识，其中共识是核心，以一个内容型社区为例，用户是否具备普遍共识是区分一个普通 BBS 和 Community 的核心要素。

在 SegmentFault 思否的社区治理中，我们同样借鉴和学习了很多 Apache Way 的精华之处，比如用户成长体系的搭建，和开源软件社区类似，SegmentFault 的用户们也可以通过贡献获取更高的社区声望，从 User 到 Contributor、Maintainer 和众审中心成员，有权投票参与社区的开放治理中。而这些投票、有关社区治理的讨论，全部 Open Communications 的原则，可被公开访问和检索。

## 2.4 中国文化与开源

东亚文化倾向于少言寡语，不公开展示情绪，这与西方或欧洲的文化有很大不同，后者对表达自己的想法毫无顾忌（Caldwell-Harris 等人，2013）。

中国文化重视维护群体和谐。任何负面情绪的表现，如愤怒或批评，都是对这种和谐的破坏。(Caldwell-Harris 等人，2013) 不和谐会影响群体有效合作的能力。在一个集体主义的社会里，群体的需求大于个人的需求，展示破坏性的情绪可能会被避免，因为它可能会把注意力集中在个人身上而不是群体身上。

在美国文化中，直接甚至是过度表达并不罕见。但在中国文化中，这往往是相反的，真正的技能实际上是关于如何婉转表达。(Caldwell-Harris 等人，2013)。

导致透过 Stallman、Torvalds 或 ASF 启动开源的触发因素和事件，在中国并没有以同样的方式或在同样的水平上发生。人们认为，对开源和 Linux 的兴趣是在 1990 年代初由爱好开源的学生和热心人士带到中国的。(Murray, 2006)。

除了个人之外，中国政府也扮演了一个赞助者的角色，帮助推广开源 (Pan & Bonk, 2007)。他们的主要目的之一是帮助减少对专利软件的依赖，并利用开源作为催化剂来启动其内部软件产业的发展 (Zhou, 2011)。

开源软件的有限可用性和缺乏成熟的开源社区让中国的学术机构发挥了关键作用。(Pan & Bonk, 2007)。他们鼓励他们的学生专注于为开源项目做贡献或自己创建新的开源项目。

1999 年是一个重要的年份，它见证了第一家中国本土的 Linux 公司 -- Xteam 的成立。(Murray, 2006)。这引起了其他 Linux 软件供应商的注意，如 Turbolinux (拓林思) 以其产品的中文版本进入中国的开源市场。

在政府的支持下，2001 年，中国科学院软件研究所成功地发布和分发了他们自己的开源操作系统 -- 红旗 Linux。(Pan & Bonk, 2007)。红旗 Linux 最终在 2014 年停产。

在 20 世纪 90 年代末和 21 世纪初，随着一些开源产品的崭露头角和持续推出，开源在中国

的商业和学术界都被带到了前沿。

中国的开源用户，最初是作为开源软件的消费者或终端用户 ( Zhou. 2011 )，使用必要的工具来满足他们的需求。现在情况似乎正在发生变化，人们期望对帮助过他们的开源项目做出一些贡献。

在中国的开源社区中，回馈的趋势似乎越来越多。截至 2019 年 4 月 30 日，ASF 有 8 个源自中国的顶级项目：

- Apache CarbonData
- Apache Eagle
- Apache Griffin
- Apache HAWQ
- Apache Kylin
- Apache RocketMQ
- Apache ServiceComb
- Apache Skywalking

以及 6 个孵化项目：

- Apache brpc
- Apache Doris（注：已毕业成为 TLP）
- Apache Dubbo（注：已毕业成为 TLP）
- Apache ECharts（注：已毕业成为 TLP）
- Apache ShardingSphere（注：已毕业成为 TLP）
- Apache Weex（注：已退役进入 Attic）

译者注：截至 2022 年 8 月，源自中国的 ASF 顶级项目，已经达到了 19 个，而孵化项目则达到了 11 个，总共 30 个项目。不包含 1 个退役项目。时隔三年，已然翻了一番

【专家点评】堵俊平

中国的开源发展近些年一直在提速，仅从项目角度而言，下一步的发展需要从量到质。

## 2.5 文化的变迁

Hofstede 的研究确定了国家文化的某些面向。这种方法意味着文化被视为可以在国家层面上进行整体打包和参考的东西 ( Rose, 2007 )。

技术已经改变了我们的互动方式，意味着我们的社会文化互动也发生了变化。我们生活在一个不再受国界限制的虚拟和电子世界里 ( Rose , 2007 )。我们可以从任何地方接触文化并与之互动。

如果文化不能被传播 ( Rose , 2007 )，那么它如何能够长期生存？更重要的是，它如何能够代代相传，并且仍然表现出相同的社会行为和特征？

答案可能就在我们使用的术语中。也许这不是关于文化的传播，而是文化所创造的关系。我们已经知道，文化是互动和社交性的，所以，也许文化只是一系列的关系，有共同的看待问题和做事情的方式 ( Rose, 2007 )。

如果文化是我们需要内化的东西，那么我们就会在我们做的每一件事和与他人的每一次互动中无意识地展示它。(Rose, 2007) 这意味着我们实际上无法控制何时以及如何将我们的文化传递给他人。因此，通过简单的互动，我们就在传递着什么。

发起 "Apache 小组" 的那群人今天仍在参与 ASF 的工作。这意味着直至今日，ASF 与其文化的源头仍有直接和存续的联系。我们能够对文化价值和相关行为的理解直接向创始团队要求澄清，有助于确保文化不会随着时间的推移而过分偏离其原始形式。简单地说就是导师辅导。

指导一个人是相当直接的，但当你需要指导整个项目及其社区，甚至是 50 个不同的项目时，该怎么办呢？ASF 已经提出了这个问题，它的回答是 Apache 孵化器。

## 2.6 Apache 孵化器

随着基金会旗下的开源项目的数量开始增长，ASF 作为一个组织被创建出来，为其项目提供管理和法律框架。Apache 孵化器创建于 2002 年，其作用是专门指导和辅导希望加入 Apache 软件基金会的项目：

"接受新项目加入基金会，提供指导和支持，帮助每个新项目产生自己的协作社区，教育新的开发者了解基金会成员定义的协同开发的理念和准则，并在这些项目的社区达到成熟时，向董事会提议将其提升为独立的项目管理委员会（Project Management Committee - PMC）地位"

孵化器有责任对潜在的项目进行教育，使其了解 ASF 期待其项目应有的行为方式。

为了支持进入孵化期的新项目，ASF 已经制定了一个正式的管理程序、角色和责任以及项目成熟度模型。尽管项目成熟度模型是一个有用的框架和工具，但是它是有弹性的，可以帮助项目制定基准，了解他们需要关注哪些方面。

### 2.6.1 角色和职责

治理的最高层次是 ASF 董事会。正是他们批准建立一个新的顶级项目，任命项目主席并批准项目管理委员会（PMC）成员。为了减少被董事会拒绝的风险，Apache 孵化过程对每个项目进行评估，以确保其达到被批准毕业所需的最低标准。

向 ASF 董事会报告的是孵化器项目管理委员会（Incubator PMC - IPMC）。它是由对帮助或指导潜在的新 ASF 项目感兴趣的人组成的。IPMC 的作用是接受新的项目进行孵化，然后与项目及其导师一起工作，争取毕业成为顶级项目。

接受一个新项目进入孵化器的流程是从创建一个孵化提案开始。在该项目被接受之前，它被称为 "项目建议书"。每个项目建议需要一个 "领路人（Champion）"。这通常是指已经深入参与 ASF 的人，他们可以看到将项目引入孵化的价值。

领路人与拟议中的项目团队一起工作，创建提案。提案书详细介绍了项目的现状并评估了项目对未来的期望。它着眼于该项目为什么想成为 ASF 的一部分，并讨论它是否适合。

每个项目提案还需要数位导师，以帮助新人了解 ASF 的政策和指导方针，以及确保该社区发展 ASF 所倡导的做法和文化。导师必须是孵化器项目管理委员会（IPMC）的成员，一般需要 3 名以上的导师。

导师们在项目邮件列表中活跃地提供建议，并在出现争议时可以充当调解人。导师的作用对于帮助项目社区融入 ASF 文化至关重要。

## 2.6.2 Apache 项目成熟度模型

ASF 已经发展出一套 “项目成熟度模型” 作为协作过程的一部分。它是一个可弹性选用的框架模型，可以用来协助评估毕业的过程，也为项目提供了评估领域的透明度。

在孵化期间，该项目将定期向 ASF 董事会报告。在提交定期报告之前，由项目导师审查并可对任何领域提出意见，然后批准该报告的提交。

这份报告可以关注以下任何领域的状况：



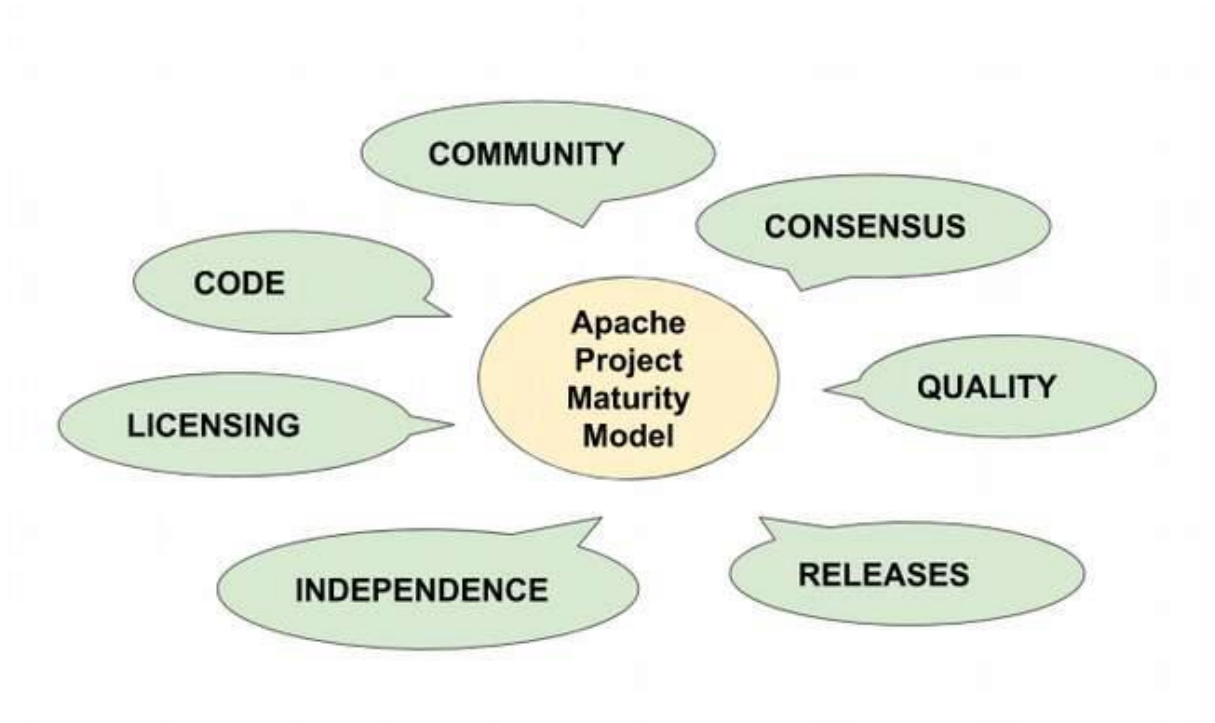


图 2.4：Apache 项目成熟度模型

每个领域都可以进一步细分为孵化项目需要确保在孵化期间完成的具体任务。

### Apache 项目成熟度模型细节级别

#### 代码 CODE

代号 描述

CD10 本项目为免费向公众分发而生产开源软件。

CD20 本项目的代码容易被公开发发现和取得。

CD30 本项目的代码能使用公开的标准工具，以可重复生产的方式来构建。

CD40 本项目的代码的完整历史能透过源代码版本管理系统取得，能让所有已发布的版本都被重新创建。

CD50 本项目的代码能透过源代码管理系统明确每一行代码的出处，及其提交者的可信的

认证信息。当第三方贡献被提交时，提交信息提供了关于代码出处的可靠信息。

## 许可证及版权 LICENSE & COPYRIGHT

代号 描述

LC10 本项目的代码发布须基于 Apache 许可证 v2.0。

LC20 本项目的代码所必须采用和依赖的库，不得产生比 Apache 许可证更多的限制。

LC30 LC20 所提到的库也可和开源软件一样的方式取得和使用。

LC40 提交者受到个人贡献者协议 (Individual Contributor Agreement - ICLA) 的约束，该协议定义了提交者可以提交哪些代码，以及需要辨识非自身的代码。

LC50 清楚定义和记录本项目生产的一切事物的版权的所有权。

## 版本发布 RELEASES

代号 描述

RE10 发布的版本须包含以标准和开放的长期可读的归档 (archive) 格式分发的源代码。

RE20 发布的版本须由本项目之项目管理委员会批准 (参考 CS10)，以形成基金会的正式举措。

RE30 发布的版本须经过正式签署及/或分发，并伴随着可靠、能验证该下载项目归档 (archives) 的项目摘要。

RE40 便利二进制文件 (Convenience binaries) 可以和源代码一起分发，但不能被视为 Apache 发布版 – 而只是为用户之便利使用而提供，不带任何保证。

RE50 版本发布流程需被完整记录，且可被新用户独立地重复制作相同而完整的版本。

## 质量 QUALITY

代号 描述

- QU10 对项目的质量保持开放且诚实的态度。只要能清楚的传达，不同软件模块有不同层次的质量和成熟是理所当然而且可以接受的。
- QU20 生产高安全性的软件是项目的优先考量。
- QU30 本项目提供完整记录、安全且私密的渠道以报告并回应安全问题。
- QU40 本项目对向后兼容优先考量，也记录不兼容的改动，并提供工具与文档协助用户顺利转移到的新功能 (或特性)。
- QU50 本项目对记录在案的 Bug Report 努力且即时地回应。

## 社区 COMMUNITY

代号 描述

- CO10 本项目有一个广为人知的官网，指向所有根据成熟度模型而运营项目所需的信息。
- CO20 本社区欢迎所有怀有诚心，尊重他人并能为项目增加价值的贡献者。
- CO30 贡献不仅只是源代码，还包含了文档、建设性的 Bug 报告、建设性的讨论、营销和其他一切能对项目增加价值的行动。
- CO40 本社区努力推动精英治理，并以赋予更多权利和责任给能对项目增加价值的贡献者为目标。
- CO50 清楚记录如何一视同仁地赋予贡献者更多权利，如提交的权限或决策权力。
- CO60 本社区根据有决策权成员 (参考 CS10) 的共识来运营。独裁者 (无论是否仁慈) 在

项目中不受欢迎。

CO70 本项目社区致力于即时地回答用户的问题。

## 独立性 INDEPENDENCE

代号 描述

IN10 本项目独立于任何公司或组织的影响。

IN20 贡献者只代表个人，而非任何公司或组织的代表。

## 共识 CONSENSUS

代号 描述

CS10 本项目维护一个由有决策权的贡献者组成的项目管理委员会 (PMC) 的公开名单。

CS20 由 PMC 成员依共识形成决策，并记录于该项目的主要沟通渠道 (如邮件列表)。项目社区的意见会被考量。

CS30 当讨论不足以形成决策时，有文档记录的投票规则得以用来建立共识。

CS40 根据上述 CS30 定义的 Apache 投票规则，在 ASF 项目里，否决只适用于代码提交，同时需要提出合理的技术解释。

CS50 所有重要讨论，异步发生并以文字形式记录于该项目的主要沟通渠道 (如邮件列表)。凡是在线下，面对面或私聊讨论而影响该项目者，皆须记录于该渠道。

进入孵化期的新项目在头三个月内，需要每月向 ASF 董事会报告。此后则改为每三个月一次。报告一般包括关于项目已达成的成就的信息，以及他们为实现毕业而计划做的事情。

Apache 项目成熟度模型可以作为一个检查表来突出任何潜在的差距。尽管该模型是可选的，但它提供了明确的指导，为项目需要实现的目标提供了明确的指导。

每个项目社区都是独特的，因此可能有不同的挑战需要克服，这可能需要时间，所以只要项目是活跃的，并取得进展，孵化的时间一般没有限制。在项目活动非常少的情况下，ASF 董事会将要求项目管理委员会中至少有 3 名成员确认他们仍然积极参与项目，以保持孵化。

项目也可以从 ASF 的孵化中退出。这样做的原因是有多种可能，其中最主要的原因是缺乏文化适应性。

在任何项目可以被考虑毕业之前，它必须证明其

- 有一个围绕着项目的活跃社区
- 能够很好的适应 ASF
- 理解并实践 Apache 之道的文化
- 有一个现有的项目代码库
- 了解孵化过程和孵化器对项目的期望

除了法律和基础设施方面的要求外，申请毕业的项目还非常需要证明以下几点：

- 一个活跃和多样化发展的社区
- 采用基于功绩的社区互动方式
- 不依赖任何一个贡献者，至少有 3 个独立的提交者
- 没有单一个公司或实体能决定该项目的成功
- ASF 风格的共识已机制被采纳并成为标准做法
- 容忍和解决社区内冲突的能力
- 能够在公开场合创建和执行一个由社区努力推动的版本发布

如果一个项目能够满足这些要求，并且领路人（Champion）和导师（Mentor）也支持它，

那么一个项目就可以向 ASF 董事会提出决议，毕业并正式成为 ASF 的顶级项目。

## 第 3 章 研究方法

### 3.1 方法论和假设

本文将分析 ASF 文化（"Apache 之道"）的主要元素，并试图提供证据表明这种文化可以成功地传递给新项目的程度。它将测试和探讨以下假设：

1. Apache HTTP 服务器项目作为 ASF 文化的原始来源，我们可以从中提取数据并用于创建文化模型
2. 这种文化模式可以作为一个基准线，与其他 ASF 项目进行比较，看它们是否表现出类似的文化特征
3. 这个文化模型可以作为一个基准线，用来比较那些经历过或未经历 ASF 孵化器的项目，看看它们是否表现出类似的文化特征
4. 与其他非中国相关的项目相比，由中国贡献者贡献的项目可能会经历更显著的文化变化

基于上述假设，本文将重点讨论以下内容：

1. 利用公开的信息定义 ASF 文化模式的主要概念和价值
2. 界定 Apache 孵化器的角色和流程
3. 从 ASF 的 Apache HTTP 服务器项目的代码库和邮件列表中挖掘可用的公共数据，以创建一个文化基准线。
4. 从 ASF 代码库和邮件列表中挖掘一系列 Apache 项目的公共数据，这些项目是在 Apache HTTP 服务器项目之后创建的，以寻找数据指标来验证或反驳 Apache HTTP 服务器文化基准线中发现的示范要素。
5. 从 ASF 代码库和邮件列表中挖掘出一系列 Apache 项目的公共数据，这些数据是：
  - 未经历 ASF 孵化器的项目（即没有正式嵌入 ASF 文化）
  - 正在孵化中或已经从 Apache 孵化器毕业的由非中国贡献者贡献的项目（正式经历 ASF 文化嵌入）
  - 正在孵化中的或已经从 Apache 孵化器毕业的由中国贡献者贡献的项目（正式经历 ASF 文化嵌入）

由此产生的分析结果将被评估和相互比较，并用于回答本研究报告所提出的问题。

## 3.2 研究设计

这项研究的目的是为了更好地了解 Apache 软件基金会的开源文化，并分析一些具有不同文化特征的项目，以了解在项目演化和毕业成为顶级项目的过程中是否会发生文化变化。毕业并成为顶级项目的一个关键部分涉及到 Apache 行为和文化价值观的展示。

研究的重点是经历 Apache 软件基金会孵化的 12 个 Apache 项目，其中 4 个是中国人发起的，看看在孵化过程中是否发生了文化变化。

为了有一些比较性的数据，这 12 个项目还包括：

- 四个经过孵化的 Apache 项目是由非中国贡献者发起的
- 四个没有经过孵化的 Apache 项目是由非中国人发起的

研究的时间段为 2014 年 1 月 1 日至 2019 年 4 月 30 日。

## 3.3 环境

所有的 ASF 项目都有公开归档的邮件列表。它们是 ASF 创建时传承下来的通信媒介，现在构成了任何 ASF 项目的一个组成部分。

考虑到社会和通信媒介的技术变化，这种沟通方式似乎是非常规的，但 ASF 治理的一个重要部分是确保社区讨论和决定有一个记录。

邮件列表允许异步交流，这样处于不同时区的人，或那些兼职做项目的人，或在一般办公时间之外的人，都有时间和机会加入对话。



## 3.4 研究工具和指标

本研究将使用以下工具、公式和指标进行。

### 3.4.1 Apache Kibble

Apache Kibble 是一套对软件项目的数据和活动进行收集、汇总和可视化的工具。

以下 Kibble 指标将被使用：

### 3.4.2 小马系数 Pony Factor

小马系数（PF）是根据个人贡献者的贡献来衡量一个项目的多样性。它可以被定义为："贡献者的最低数量，他们的总贡献占到了任何被衡量的东西（例如，编写的代码行数，发送的信息数量等）的大多数。

在数学上的表示为：

$$\sum_{i=1}^P C_n \geq K.V$$

- P 即是小马系数
- $C_n$  是贡献者 n 的贡献数量，按贡献数量降序排列
- K 是总贡献的百分比
- V 是所做贡献的总量

较高的 "小马系数" 意味着，由于项目有很好的容忍度，即使一个或多个核心贡献者离开，可以继续生存下去。

注：小马系数包括所有贡献者的贡献，无论他们是否仍然活跃。

### 3.4.3 增强的小马系数 Augmented Pony Factor

增强的小马系数（APF）是对标准小马系数计算的调整，其中省略了不再活跃的贡献者的贡献。

$$\sum_{i=1}^P C_n \geq K.V - \text{Contributions from non-active contributors}$$

（增强的小马系数 = 小马系数 - 非活跃贡献者的贡献）

注：增强的小马系数将不作为评估的一部分，但为了完整起见，这里包括了对它的描述。

### 3.4.4 元小马系数 Meta Pony Factor

元小马系数的计算是一项正在进行中的工作。它试图根据与贡献相关的电子邮件地址来衡量贡献者的归属。如果进一步发展，这可以帮助识别不同的组织的贡献。

### 3.4.5 小马系数、增强型小马系数、元小马系数的范例

下面的例子可以帮助解释小马系数、增强的小马系数和元小马系数的使用。

一个开源项目有 35 位活跃的贡献者：

- 为 X 公司工作的贡献者 A 做出了 19% 的贡献
- 为 X 公司工作的贡献者 B 做出了 15% 的贡献
- 为 Y 公司工作的贡献者 C 做出了 22% 的贡献，但是不再活跃
- 为 Z 公司工作的贡献者 D 做出了 12% 的贡献
- 没有公司从属关系的贡献者 E 做出了 9% 的贡献
- 所有其他贡献都是由其余的 30 位贡献者所提供

小马系数将是 3，因为贡献者 A、B 和 C 的贡献量占项目的 56%。

增强的小马系数将是 4，因为贡献者 C 不再活跃，所以他们贡献的 22% 将不包括在计算中。这意味着贡献者 A、B、D 和 E 的贡献量占了项目的 55%。

元小马系数是 4，因为我们有 4 个可见的归属机构（X, Y, Z 公司 and 无隶属关系者）与小马系数的贡献相关。

### 与 ASF 的相关性

ASF 项目是由个人贡献者组成的。其中一些人通过代码对软件做出贡献，一些人通过非代码或知识（如文档、测试、营销等）做出贡献。小马系数可以衡量一个项目在贡献者之间的工作分工方面的多样性。小马系数越高，在一个或多个核心贡献者离开项目的情况下，继续生存的容忍度越强。

小马系数可用于代码库、电子邮件和问题跟踪器。代码库的小马系数将侧重于编码贡献，而电子邮件的小马系数将包括非编码贡献，如社区参与和通过邮件列表的参与。

Apache HTTP 服务器项目是直接由于一个核心贡献者离开 NCSA 项目（造成该项目无人继续维护）而创建的，所以 ASF 文化的一个关键部分是确保项目在主要贡献者离开时也能生存。

小马系数也可以帮助说明：

- 一个项目的成长
- 正在接受新的贡献者
- 功绩奖励由提交者的状态来决定

### 3.4.6 感性分析

感性分析是一种文本挖掘的形式，用于根据一个人所写的内容来预测他的情绪状态（Pang &

Lee, 2008)。这意味着它可以作为一个指标来衡量人们对某些想法的意见和反应。

数据是以文本的形式收集的，一个算法被用来识别与情感相关的关键词。任何沟通形式都可以与几种情绪相联系，因此使用权重来强调情绪的强度。

对比和相关的情绪分析是衡量所表达的情绪在整个群体或组织中统一处于相同或相似的强度水平的代表性。Apache Kibble 包括一个功能，即比较整个组织的情绪，以突出它的代表性。

所有的 ASF 项目都有公共邮件列表作为主要的沟通方式。社区中发生的一切都将反映在邮件列表里所记录的文本互动中。这意味着，对邮件列表进行有意识的感性分析可以显示出一个社区的情绪状态。

以下情绪集将被用于分析：

- 心情好 = 集合(['信任', '喜悦', '自信', '积极'])
- 心情不好 = 集合(['悲伤', '愤怒', '厌恶', '恐惧', '消极'])
- 情绪中性 = 集合(['预期', '惊喜', '暂时', '分析性', '中性'])

### 3.4.7 关键句提取

关键词句提取 (KPE - Key Phrase Extraction) 是一种提取关键词句或单词的方法。可以概括文本或文件的主要观点或主题。它已被成功地用于期刊和在线内容的索引。但在本文中，它将被用于提取任何可能表明文化思想或语言的文本。

在这项研究中，我们将使用 Apache Kibble 的关键词句提取功能，从我们选择的 12 个项目的邮件列表中提取关键词句，看看是否有文化思想或语言的迹象。

如果一种文化蕴含在一个社区中，那么所使用的语言就应该是这种文化的标志。被称为 "文化语言学" 的领域试图研究文化与所用语言之间的关系 (Sharifian, 2017)。我们期待对 ASF 文

化有重大意义的词汇和词句在本研究报告中能呈现出来。

### 3.4.8 贡献者留存

贡献者留存与一个项目在吸引和留存贡献者方面的成功程度有关。它可以被分解成以下几个方面：

- 活跃的贡献者
  - 在这个项目中，有多少贡献者是活跃的？
  - 贡献者是否定期做出贡献并在较长的时间内保持活跃？
- 留存的贡献者
  - 贡献者已经贡献了多长时间了？
  - 贡献者被留存的时间越长，项目在留住他们方面就越成功
- 已经离开的贡献者
  - 有多少贡献者离开了
  - 贡献者的离职率
- 回归的过去的贡献者
  - 有多少贡献者在过去做出了贡献，并回到了社区，重新加入了社区
  - 过去的贡献者的回归率

贡献者留存率是一个非常重要的指标，因为除了活跃或不活跃之外，它还显示了一个项目是否有良好的活跃贡献者组合，是否接受新的贡献者或主要依赖有经验的贡献者。

(译者注：留存率计算方式

$[(E-N)/S] \times 100\% = CRR$  ( 客户留存率 - Customer Retention Rate )

- S = 这个时间段开始时，既有的客户数量

- E = 这个时间段结束时，客户的总数
- N = 这段期间内新增的客户数量 (N)

范例：所得出的结果是以百分比呈现。举例来说，有一家企业在这段期间开始时有 100 个客户 (S)，结束时有 100 个客户，这段期间内新增了 10 个客户 (N)。那么，他们的客户留存率就是  $90 : [(100-10)/100] \times 100 = 90。$

## 3.5 数据采集

所有的 ASF 项目都有公开归档的邮件列表。它们是创建 ASF 时传承下来的通信媒介，现在形成了任何 ASF 项目的一个组成部分。它是一个项目的核心，是人们互动、交流、协作、争论、同意和不同意的地方。这意味着它是一个挖掘数据以寻找文化指标的合适场所。

用于创建文化基准线的数据将从 Apache HTTP 服务器项目中提取。将使用以下数据：

- Apache HTTP 服务器项目之邮件列表归档 1996 年 - 2019 年 4 月
- Apache HTTP 服务器项目之代码库 1996 年 - 2019 年 4 月
- 选定的 Apache 项目列表之邮件列表归档 2014 年 - 2019 年 4 月
- 选定的 Apache 项目列表之代码库 2014 年 - 2019 年 4 月

以下 12 个项目被选中进行比较，因为：

- 它们都是在过去 5 年（从 2014 - 2019 年）加入 ASF 的
- 他们都有 Apache Kibble 中的可用数据
- 在 Apache HTTP 服务器项目创建时，它们都还不存在
- 他们所展示的文化应该是在 ASF 成立后创造的

本案例研究中包括的项目列表

项目名称	2019 年 4 月 30 日的状态	纳入数据的理由
Apache Kylin	顶级项目 (TLP)	中国发起的孵化项目

Apache Skywalking	顶级项目 (TLP)	中国发起的孵化项目
Apache Dubbo	孵化项目 (Incubating)	中国发起的孵化项目
Apache Weex	孵化项目 (Incubating)	中国发起的孵化项目
Apache Netbeans	顶级项目 (TLP)	非中国发起的孵化项目
Apache Beam	顶级项目 (TLP)	非中国发起的孵化项目
Apache Kudu	顶级项目 (TLP)	非中国发起的孵化项目
Apache Kibble	顶级项目 (TLP)	非中国发起的非孵化项目
Apache Whimsical	顶级项目 (TLP)	非中国发起的非孵化项目
Apache Bahir	顶级项目 (TLP)	非中国发起的非孵化项目
Apache ORC	顶级项目 (TLP)	非中国发起的非孵化项目

这些项目可以细分为以下三个子组：

### 3.5.1 中国发起的孵化项目

项目名称	开始孵化日期	孵化毕业日期	孵化时长	顶级项目时长
Apache Kylin	2014-11-25	2015-11-18	12 个月	3
Apache Skywalking	2017-12-08	2019-04-17	16 个月	0
Apache Dubbo	2018-02-16	-	14 个月	-
Apache Weex	2016-11-30	-	29 个月	-

### 3.5.2 非中国发起的孵化项目

项目名称	开始孵化日期	孵化毕业日期	孵化时长	顶级项目时长
Apache Netbeans	2016-10-01	2019-04-17	30 个月	0
Apache Fineract	2015-12-15	2017-04-19	16 个月	2

Apache Beam	2016-10-01	2016-12-21	10 个月	2
Apache Kudu	2015-12-03	2016-07-20	7 个月	2

### 3.5.3 非中国发起的非孵化项目

项目名称	顶级项目起始日期	顶级项目时长
Apache Kibble	18-10-2017	1
Apache Whimsical	20-05-2015	3
Apache Bahir	18-06-2016	2
Apache ORC	22-04-2015	4

注：截至 4 月 30 日，年份已四舍五入至上一整年 2019.

#### 【专家点评】庄表伟

这一章让人读了觉得不过瘾，我非常希望了解：数据指标与文化特征之间，是如何建立关联的？以及为何能够建立这样的关联假设？

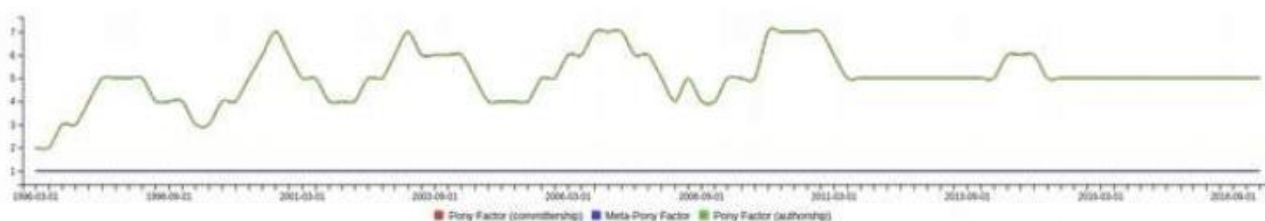


## 第 4 章 数据分析

### 4.1 Apache HTTP 服务器项目的文化基准线

#### 4.1.1 标志 1：做为基准线之代码库的小马系数 (Pony Factor)

下图展示了从 1996 到 2019 Apache HTTP 服务器项目代码库的小马系数 (Pony Factor - 以下简称 PF) 的变化。



**图 4.1：Apache HTTP 服务器项目做为基准线之代码库的小马系数变化 1996 - 2019 年**

1996 年项目刚开始的时候，PF 值很低（2）。那时 Apache 组织刚成立，所以核心维护者的人数很少。

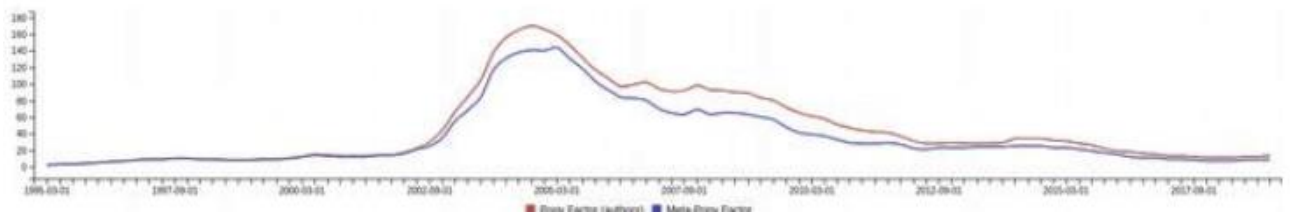
从 1996 年开始 PF 值有一段增长期，这是因为参与项目的人数增长了。之后的时间里 PF 值有升有降。2019 年初的数字依然很高并保持不变。

贡献者的 PF 值并不是清晰可见的，它和原创作者 PF 值的轮廓相似。

Meta PF 值始终为 1，很可能是因为所有的贡献都和 ASF 有关，通过提交者的 email 地址（xxx@apache.org）可以看出。

#### 4.1.2 标志 2：做为基准线之邮件列表的小马系数

下图展示了从 1995 到 2019 年 Apache HTTP 服务器项目之邮件列表的 PF 值。



**图 4.2：Apache HTTP 服务器项目做为基准线之邮件列表的小马系数变化 1996 - 2019 年**

和 Apache HTTP 服务器代码库的基准线类似，1995 年 Apache HTTP 服务器邮件列表的 PF 值很低（2），这也反映出当时邮件列表上活跃的人数有限。在 2005 年邮件列表上贡献者的人数逐渐增长到 170。然后降低到 11，现在又增长到 14。

Meta PF 值增长是因为邮件的贡献更灵活，而且新组织的加入很容易统计。它和 PF 值的曲线相似，现在降低到并稳定在 9。

### 4.1.3 标志 3：做为基准线之贡献者经验

下图展示了贡献者对代码库贡献时间长度的分布：

Apache HTTP 服务器项目已经有 20 年了，它也是 ASF 历史最悠久的项目。

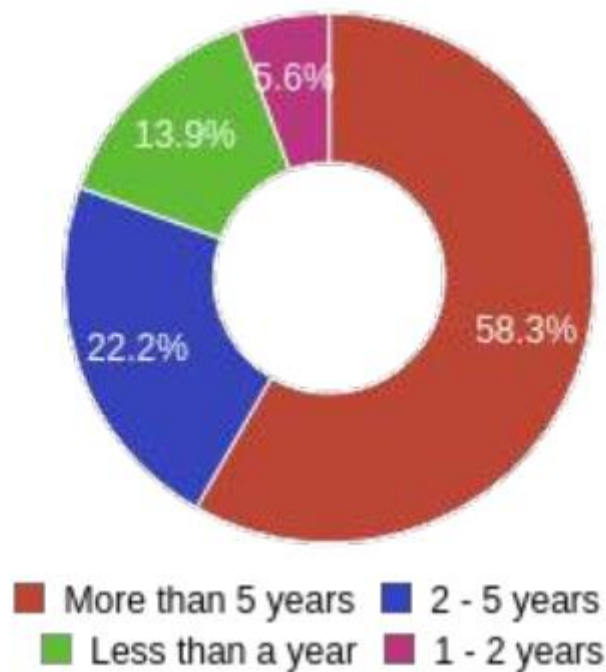


图 4.3 : Apache HTTP 服务器项目做为基准线之贡献者经验

- 超过 58% 的贡献者贡献的时长已经超过 5 年
- 只有 6% 的贡献者参与的时间少于 1 年

#### 4.1.4 标志 4 : 做为基准线之代码库的贡献者留存

下图展示了持续留在社区里对 Apache HTTP 服务器代码库做贡献的人员的数量。

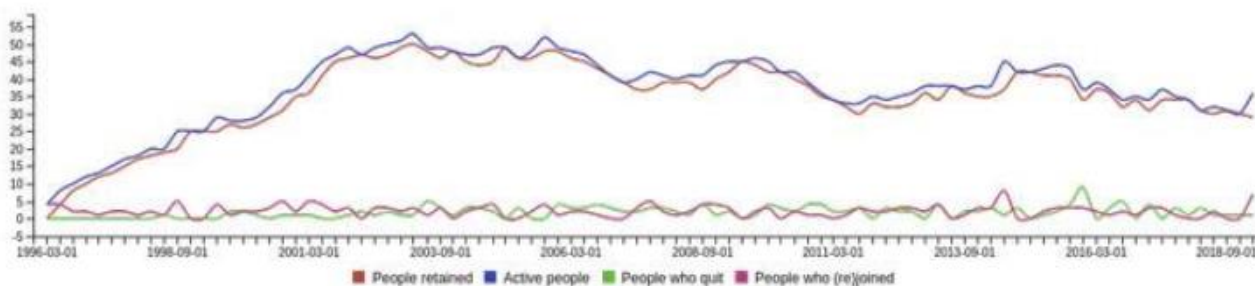


图 4.4 : Apache HTTP 服务器项目做为基准线之代码库的贡献者留存 1996 - 2019 年

活跃人数和留存人数的曲线相似。这是因为持续活跃的贡献者留在了社区里。

从图中也可以看出流失的人员数和重新加入的人员数之间的关系，因为在整个项目的生命周期里它保持在一个稳定的水平。

在 2019 年，贡献者的留存量和重新加入的人员数量都开始增长。

4.1.5 标志 5：做为基准线之留存贡献者的邮件列表统计

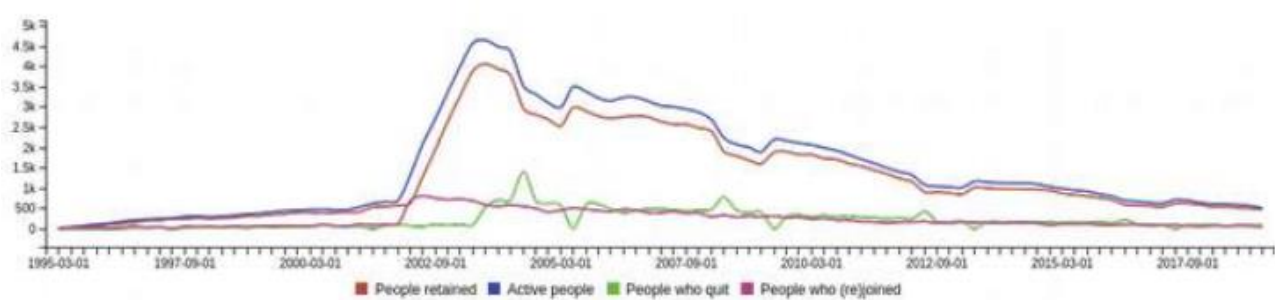
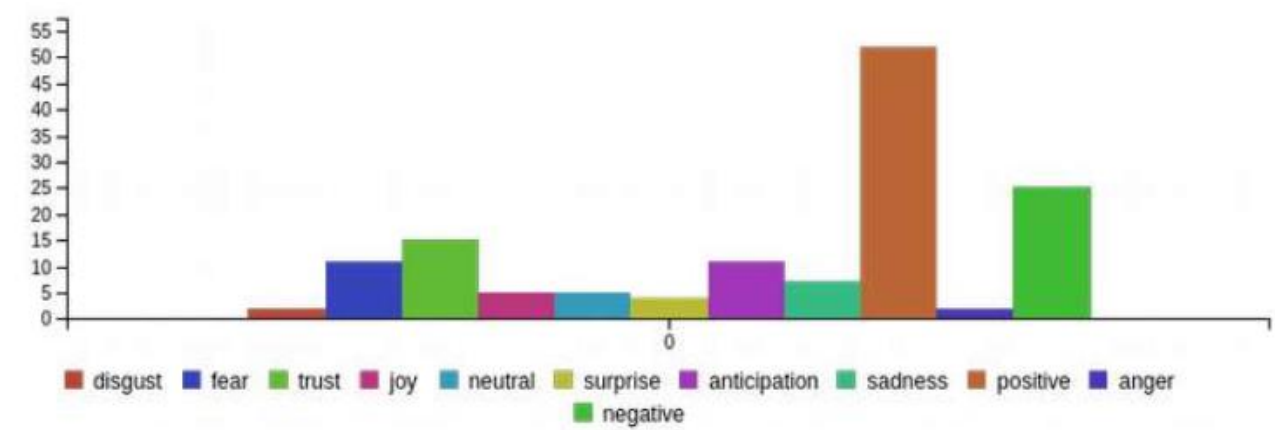


图 4.5：Apache HTTP 服务器项目做为基准线之留存贡献者的邮件列表统计 1996- 2019 年

在项目最开始的时候，活跃人数和留存人数增长缓慢，直到 2001 年左右发生了显著的增长。非常有意思的是，在活跃人数的顶峰时期，离开项目的人数也达到最高峰。之后的活跃人数和留在社区里的人数保持了稳定下降的趋势。

4.1.6 标志 6a：2019/4/30 做为基准线的情感分析

下图展示了 Apache HTTP 服务器项目在 2019 年 4 月 30 日的情感分析。



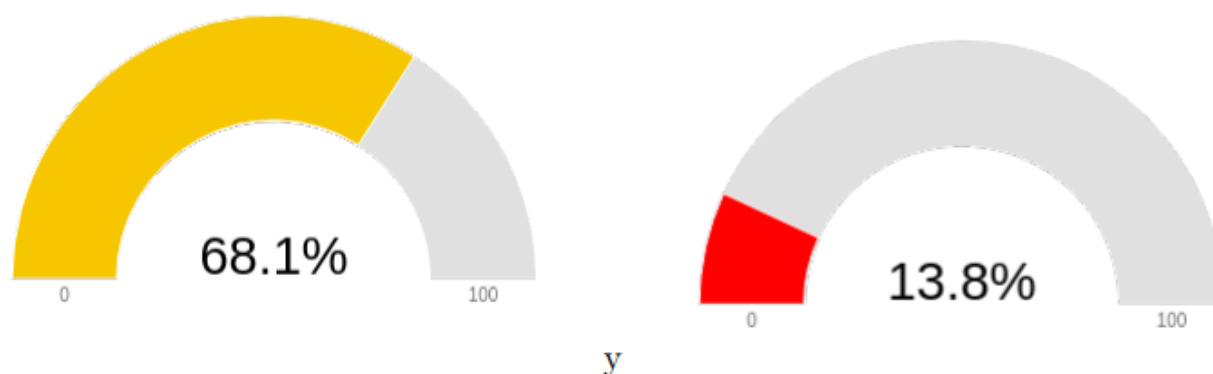
**图 4-6a : Apache HTTP 服务器项目做为基准线的情感分析 2019 年 4 月 30 日**

最高的情绪值是积极 ( 52 ), 消极 ( 25 ) 排第二位, 后面是信任 ( 15 ), 害怕 ( 11 ) 和期望 ( 11 )。

最低的 5 个情绪是中立 ( 5 ), 快乐 ( 5 ), 惊喜 ( 4 ), 生气 ( 2 ) 和厌恶 ( 2 )。

#### 4.1.7 标志 6b : 2019/4/30 做为基准线之的情感分析比较

下图展示了 Apache HTTP 服务器在 2019 年 4 月 30 日的情感分析比较。



**图 4.7 : Apache HTTP 服务器 vs ASF 相对情感**

上图展示了 Apache HTTP 服务器的相对情感数据。他们沟通的相对情感是积极的 ( 68.1% )。

当把 Apache HTTP 服务器项目的情感强度指数 ( 13.8% ) 和 Apache Kibble 中的其他 ASF 项目对比时 ( 目前为 63 ), 这意味着 Apache HTTP 服务器项目的情感强程度上比其他 ASF 项目要低。

#### 4.1.8 标志 7 : 做为基准线之随时间变化的情感

下图展示了 Apache HTTP 服务器项目从 2018 年 11 月到 2019 年 4 月的情感变化。

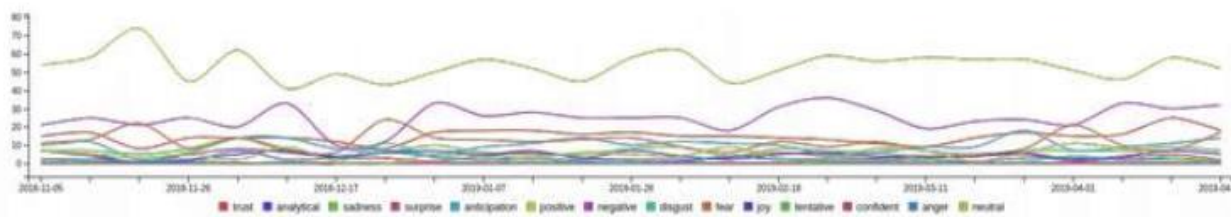


图 4.8 : Apache HTTP 服务器项目做为基准线之情绪分析 2018 年 11 月 - 2019 年 4 月

随时间变化的分析结果显示了项目社区里积极的沟通风格已经建立了，尽管数值有升有降，它依然是最强的情感表现。

#### 4.1.9 标志 8 : 做为基准线之关键词句的提取

下图显示了截至 2019 年 4 月 30 日，Apache HTTP 服务器项目中最常用的关键词句提取分析。

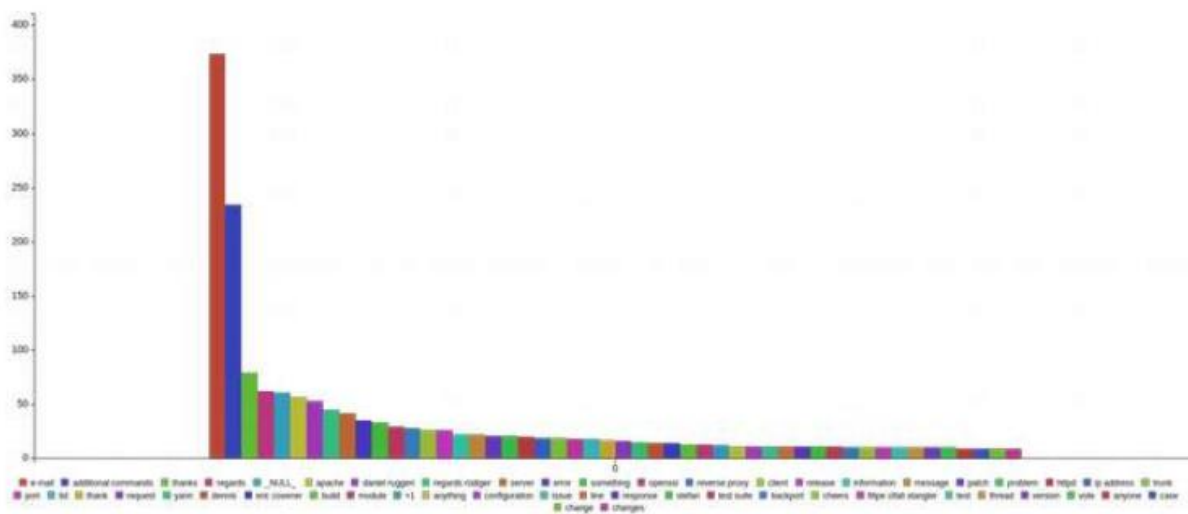


图 4.9 : Apache HTTP 服务器项目做为基准线之关键词提取 2019 年 4 月 30 日

这些词句可以被划分为以下几种类型：

- 通用或标准（例如 邮件，谢谢，问候，姓名等）
- 技术讨论（如 HTTPd，server，openssl，release，client，trunk，ip address，reverse proxy，NULL，virtual host）
- 协作（如 error，test suite，problem，thank，test，additional commands，something，

request, backport )

- 文化 ( +1, apache, vote, thread )

有些词看起来和 ASF 价值观一致，例如：

- Openness ( 开放 )：在公开场合讨论技术
- Consensus ( 共识 )：+1 = 达成共识
- Collaboration ( 协作 )：礼貌和合作的交流

注：关键词句提取 ( KPE ) 看起来和积极交流的情感分析一致。消极词汇较少出现在关键词句里。

## 4.2 非中国发起的孵化项目

### 4.2.1 标志 1：非中国发起的孵化项目之代码库的小马系数

下图展示了非中国发起的项目的代码库 PF 值，这些项目已经通过了 Apache 的孵化过程。

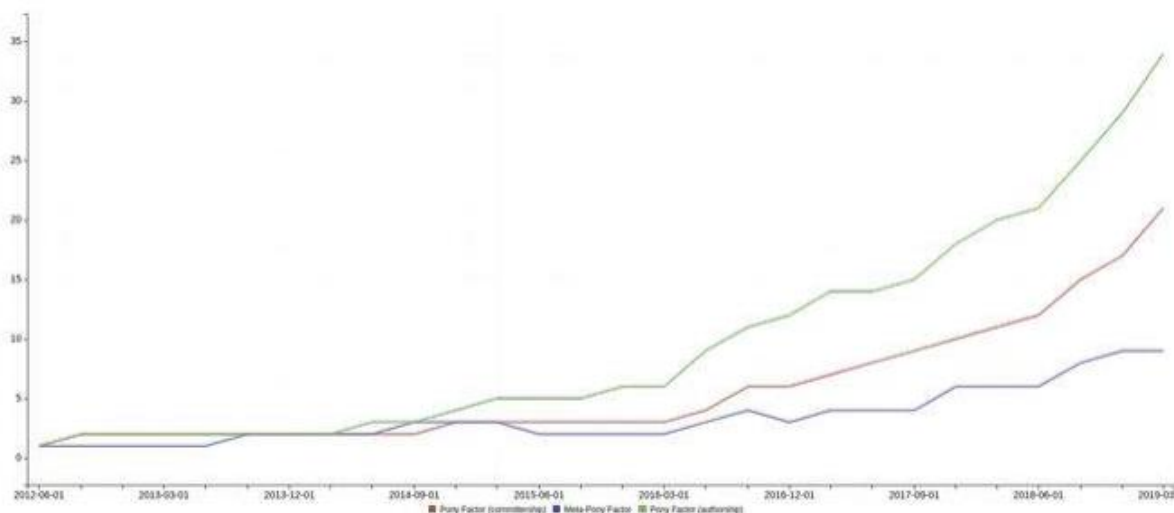


图 4.10：非中国发起的孵化项目之代码库的小马系数 2012-2019

和 Apache HTTP 服务器类似，在项目的初始阶段 PF 值较低 ( 1 )，随着时间逐渐增长到现在的水平 ( 34 )。注意，数据的实际变化情况和上图中的 Apache HTTP 服务器的曲线不一样，

上图中的曲线要平缓的多，没有剧烈的改变。

提交者的 PF 值的曲线和作者的 PF 值曲线的趋势相同，但是值不同（一个是 21，一个是 34）。

注意，Meta PF 值不是一直为 1，而是渐增的，目前是 9。这很有趣，因为它是第一个显示代码库贡献者的多个从属关系的度量结果。这可能是较新项目采用了其他版本控制系统（如 GitHub）的结果，它们没有采用传统的 Apache Subversion。

#### 4.2.2 标志 2：非中国发起的孵化项目之邮件列表的小马系数

下图展示了非中国发起的孵化项目的 PF 值。

注：邮件列表的统计数据涵盖了 2016~2019。

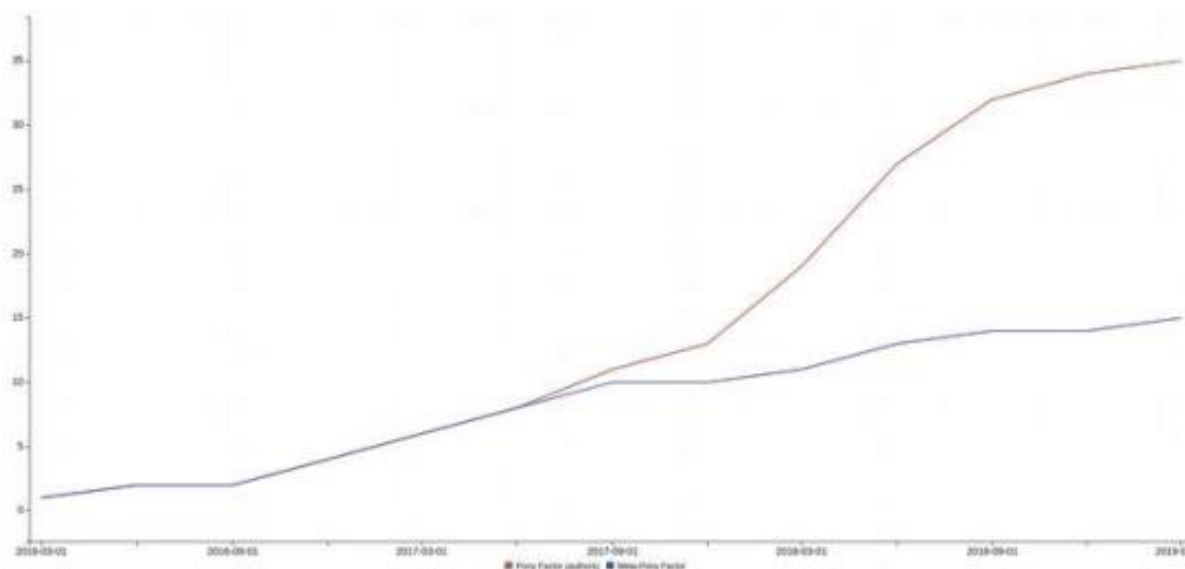


图 4.11：非中国发起的孵化项目之邮件列表的小马系数 2012-2019

与上述的代码库类似，非中国发起的孵化项目的邮件列表 PF 值开始很低（1）。邮件列表的活跃贡献者数量逐渐上升到目前的峰值（35）。

这条曲线的趋势和 Apache HTTP 服务器项目的曲线不同，它在增长的时候，Apache HTTP



服务器项目的曲线是下降的。

### 4.2.3 标志 3：非中国发起的孵化项目的贡献者经验

下图展示了贡献者对代码库贡献时长的分布。

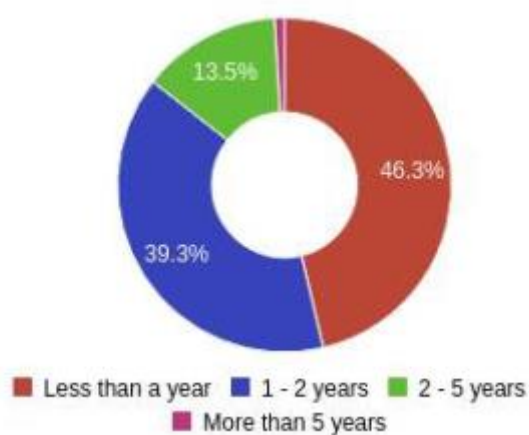


图 4.12：非中国发起的孵化项目的贡献者经验

- 超过 46% 的贡献者所贡献的时间少于 1 年。
- 超过 39% 的贡献者贡献了 1-2 年。
- 超过 13% 的贡献者贡献了 2-5 年。
- 不到 1% 的贡献者贡献超过了 5 年。

注：超过 85% 的贡献者时间不超过 2 年，这和 Apache HTTP 服务器相应的数据差别很大。

### 4.2.4 标志 4：非中国发起的孵化项目之代码库的贡献者留存

下图展示了留在社区里并对代码库做贡献的人员的数量。

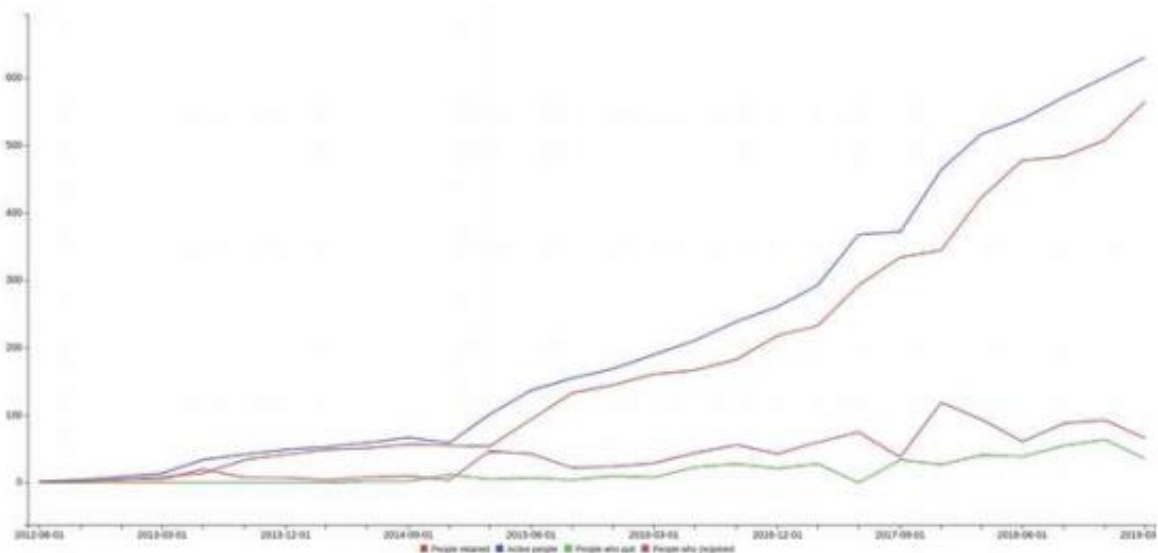


图 4.13：非中国发起的孵化项目之代码库的贡献者留存 2012-2019

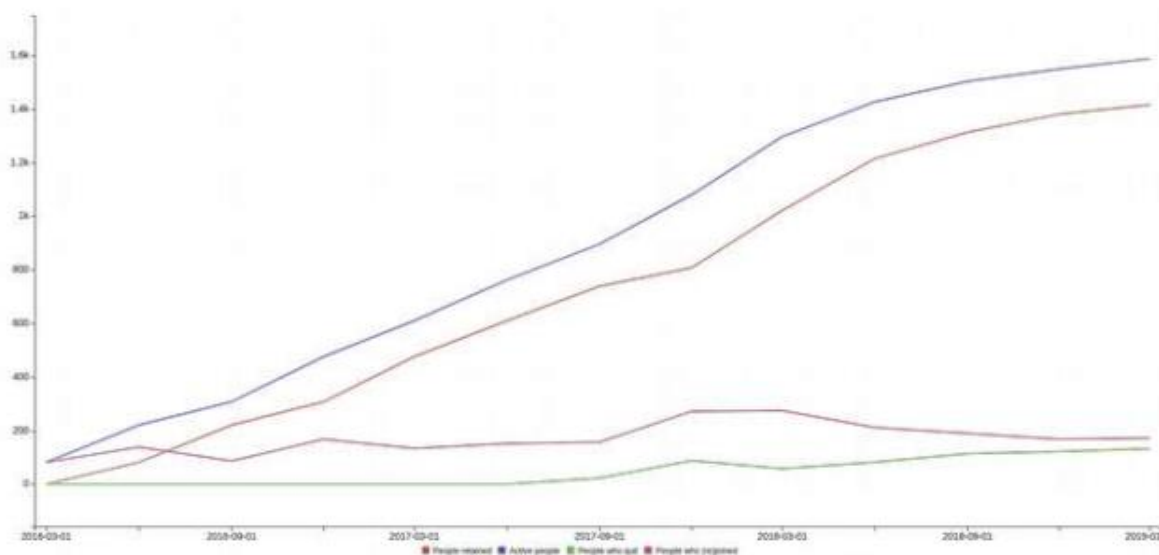
上图显示除了活跃人数和留存人数都在逐渐增长。留存率也在增长，这和 Apache HTTP 服务器的曲线不同。

在写此报告时，2019 年的数据如下：

- 631 个活跃贡献者
- 535 人留在社区里
- 66 人重新加入
- 36 人退出

#### 4.2.5 标志 5：非中国发起的孵化项目之留存贡献者的邮件列表统计

下图展示了社区里留存并对邮件列表讨论作出贡献的人员数量。



**图 4.14：非中国发起的孵化项目之留存贡献者的邮件列表统计 2012-2019**

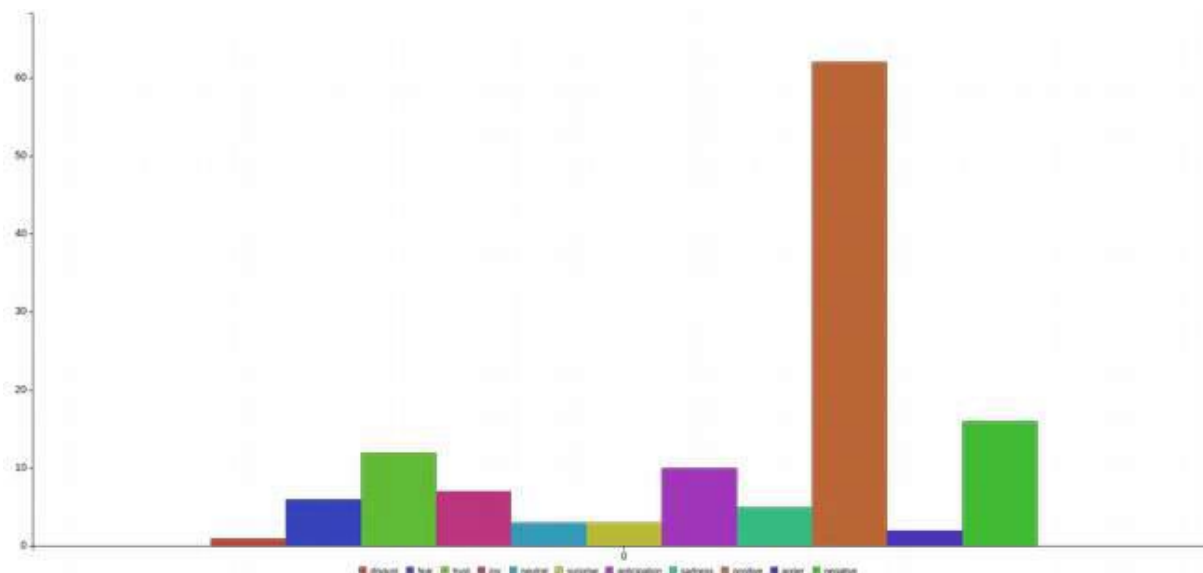
这张图看起来并不是 Apache HTTP 服务器的基准线。相反地，它随着时间的推移逐渐增加，而且还在上升。这意味着贡献者的数量正在积极增长，而且他们正在被留存。

#### 4.2.6 标志 6a: 2019/4/30 非中国发起的孵化项目之情感分析

下图是截至 2019 年 4 月 30 日非中国人发起的孵化项目的综合情绪分析。

最高的情绪是积极（62），消极（15）是第二，其次是信任（12）和期待（10）和快乐（7）。

底部的五种情绪展现的是悲伤（5），惊讶（3），中性（3），愤怒（2）和厌恶（1）。

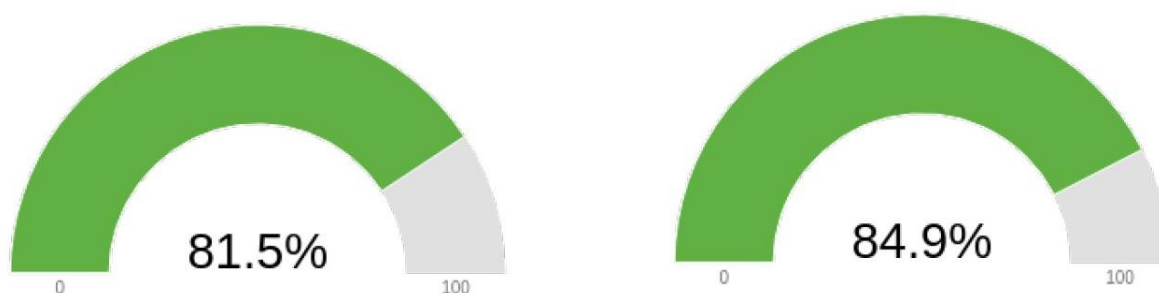


**图 4.15: 2019/4/30 非中国发起的孵化项目之情感分析**

注：3 个最强烈的情感与 Apache HTTP 服务器项目相同，并且它们的分布也类似。其他情绪有一些波动，但并不显著。

#### 4.2.7 标志 6b: 2019/4/30 非中国发起的孵化项目之情感分析比较

下图显示了截至 2019 年 4 月 30 日，非中国发起的孵化项目之情绪对比分析。



**图 4.16: 非中国发起的孵化项目与 ASF 的情绪对比**

上图显示了非中国发起的孵化项目的相对情绪。他们的沟通情绪相对来说是非常积极的（81.5）

当这个情绪与 Apache Kibble 中的其他 ASF 项目（目前为 63）进行比较时，情绪（84.9）

显示了 ASF 项目情绪强度的一般水平。因此，这意味着非中国发起的孵化项目表现出与其他 ASF 项目相似的情绪强度水平。

4.2.8 标志 7: 非中国发起的孵化项目之随时间推移的情感分析

下图是 2018 年 11 月 - 2019 年 4 月非中国发起的孵化项目之情绪分析。

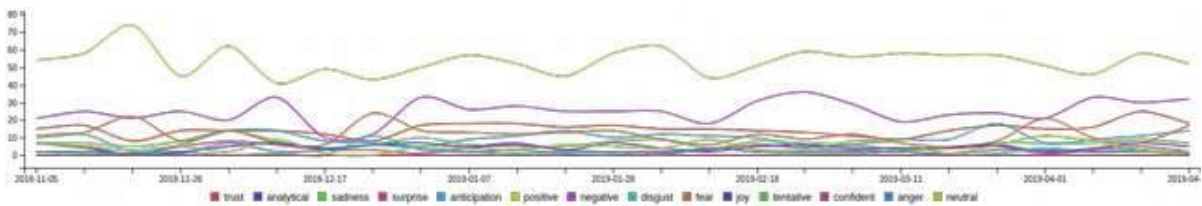


图 4.17: 2018 年 11 月 - 2019 年 4 月 非中国发起的孵化项目之随时间推移的情感分析

随着时间的推移，分析表明，积极的沟通方式已经确立，并且是表达的最强烈的情绪。这与 Apache HTTP 服务器项目的基准线模型非常相似。

4.2.9 标志 8: 非中国发起的孵化项目的关键词提取

下图显示了截至 2019 年 4 月 30 日，非中国人发起的孵化项目中最常用的关键词提取分析。这些词句可以细分为几个方面：

- 一般或标准词句（如留言、感谢、电子邮件、问候、姓名等）。
- 技术讨论（如：NULL，附加命令，发布，错误，文件，仓库，代码，测试）

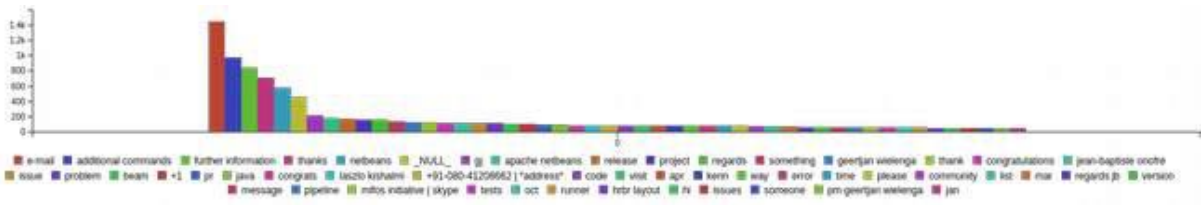


图 4.18：非中国发起的孵化项目之关键词句提取 2019/04/30

- 协作（例如，补充意见、进一步的信息、祝贺、项目、文件、请、附件、改进报告）。

- 文化（如社区、+1、列表）。

有几个文化指标与 ASF 的价值观相一致，具体如下：

- 开放性：在公开场合进行技术对话。
- 共识：+1 = 表明有共识。
- 协作：礼貌沟通、信息请求、问题分配、问题解决。
- 社区：邮件列表是交流的媒介。

这个关键词提取（KPE）概况与 Apache HTTP 服务器项目非常相似，许多常用的词句都在使用，包括‘社区’。而 +1 共识指标具有文化上的意义。

注：关键词提取（KPE）也与情绪分析相一致，突出了积极的沟通风格。有很多协作性的词句被提取出来，负面的词句在此分析中并未凸显出重要性。

## 4.3 中国发起的孵化项目

### 4.3.1 标志 1：中国发起的孵化项目之代码库的小马系数

下图显示了经过 Apache 孵化流程的中国发起的孵化项目之代码库的小马系数（Pony Factor）。

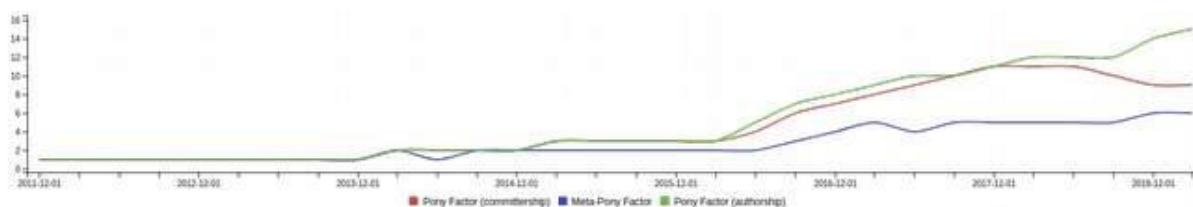


图 4.19：中国发起的孵化项目之代码库的小马系数 2011-2019

与 Apache HTTP 服务器项目一样，在项目的初始阶段，Pony Factor 很低（1），并随着时间的推移逐渐增加到目前的水平（15），是非中国发起的孵化项目的一半。请注意，该图的实际

形状与 Apache HTTP 服务器项目的曲线不一样，但它确实类似于非中国发起的孵化项目的一般曲线。

提交者的小马系数与作者的曲线走势相同，在提交者的小马系数下降到目前 9 的水平之前，与作者的曲线密切吻合。

请注意，Meta Pony Factor 不是恒定的 1，而是逐渐增加的，目前的数值是 6。

### 4.3.2 标志 2：中国发起的孵化项目之邮件列表的小马系数

下图显示了中国发起的孵化项目的小马系数。

注：邮件列表的统计涵盖了 2015 年至 2019 年这段时间。

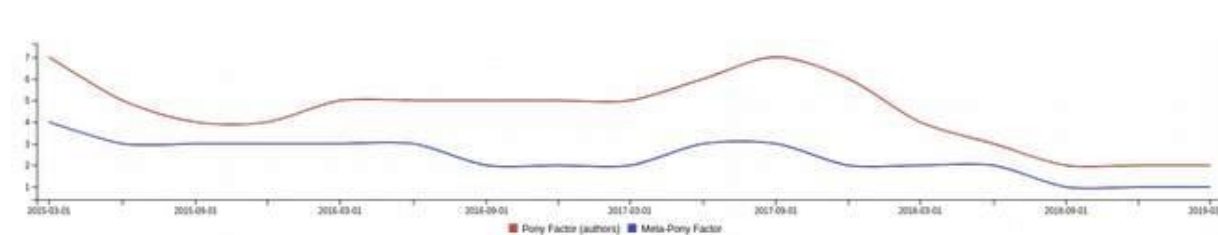


图 4.20：中国发起的孵化项目之邮件列表的小马系数 2015-2019

这里看起来与 Apache HTTP 服务器项目和非中国发起的孵化项目都有很大不同。小马系数开始是最高的，然后减少，现在已经降到了 2 的低点。

这突出表明，主要的电子邮件流量是由极少数人产生的。这很有趣，因为它可能表明对话也许发生在其他地方（即不是所有社区的人都能说流利的英语），或者也许社区需要很少的沟通来启动工作。

（译者注：也有可能是中国的贡献者更习惯使用即时通讯工具，如微信或 QQ）

### 4.3.3 标志 3：中国发起的孵化项目之贡献者经验

下图显示了贡献者对代码库贡献的时间长度的分类：

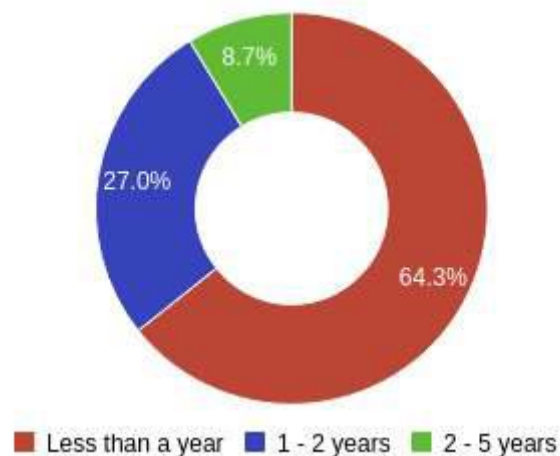


图 4.21：中国发起的孵化项目之贡献者经验

- 超过 64% 的贡献者对项目的贡献时间不到一年
- 27% 的贡献者已经贡献了 1-2 年的时间
- 超过 8% 的贡献者有 2-5 年的经验。

注：这与 Apache HTTP 服务器项目的面貌有很大区别。新的贡献者比率甚至高于本研究报告中非中国发起的孵化项目。

#### 4.3.4 标志 4：中国发起的孵化项目之代码库的贡献者留存

下图显示了有多少人被留存为社区的一部分并为代码库做出贡献。

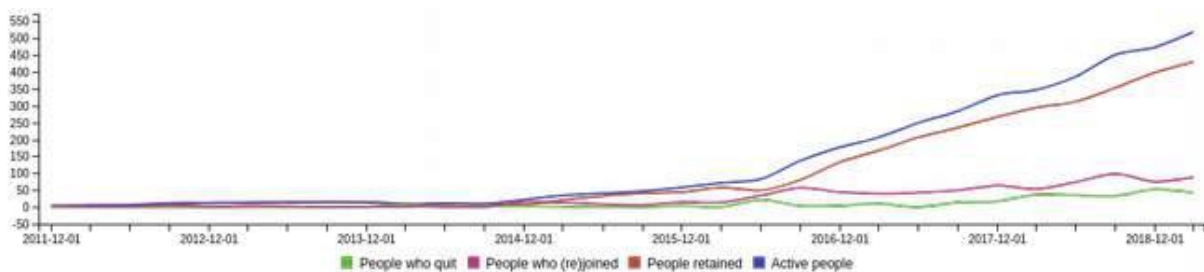


图 4.22：中国发起的孵化项目之代码库的贡献者留存 2011-2019

这表明，在 2014 年项目进入 Apache 孵化器之前，是一条平坦的曲线。此后，随着时间的推



移，活跃的人和留存的人逐渐增加。随着贡献者留存率的不断上升，它与 Apache HTTP Server 的情况非常不同。在撰写本文时，目前 2019 年的数据显示了：

- 518 名活跃者
- 430 留存者
- 88 人重新加入
- 43 人退出

注：代码库创建于 2011 年，但项目从 2014 年起进入孵化器，这就是贡献者留存的开始。这似乎表明了由于进入孵化器并保持在孵化状态，导致社区发生了一些事情。

#### 4.3.5 标志 5：中国发起的孵化项目之留存贡献者的邮件列表统计

下图显示了有多少人被留存为社区的一部分并在邮件列表讨论中做出贡献。

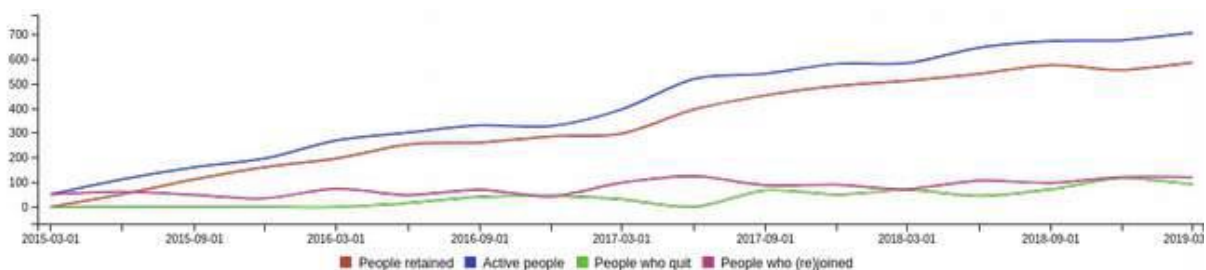


图 4.23: 中国发起的孵化项目之留存贡献者的邮件列表统计 2012 - 2019

这张图不看 Apache HTTP 服务器的基准线，但与非中国发起的孵化项目控制组相似。它随着时间的推移逐渐增加，并且仍在上升。这意味着贡献者的数量正在积极增长，而且他们正在被留存下来。

注：贡献者的数量是 705，大约是 Apache 孵化器对照组（1,588）的一半。

4.3.6 标志 6a: 2019/4/30 中国发起的孵化项目之情感分析

下图是截至 2019 年 4 月 30 日非中方发起的孵化项目的综合情绪分析。

最高的情绪是积极 ( 64 ), 第二是信任 ( 15 ), 其次是消极 ( 12 ), 以及期待 ( 11 ) 和恐惧 ( 9 )。

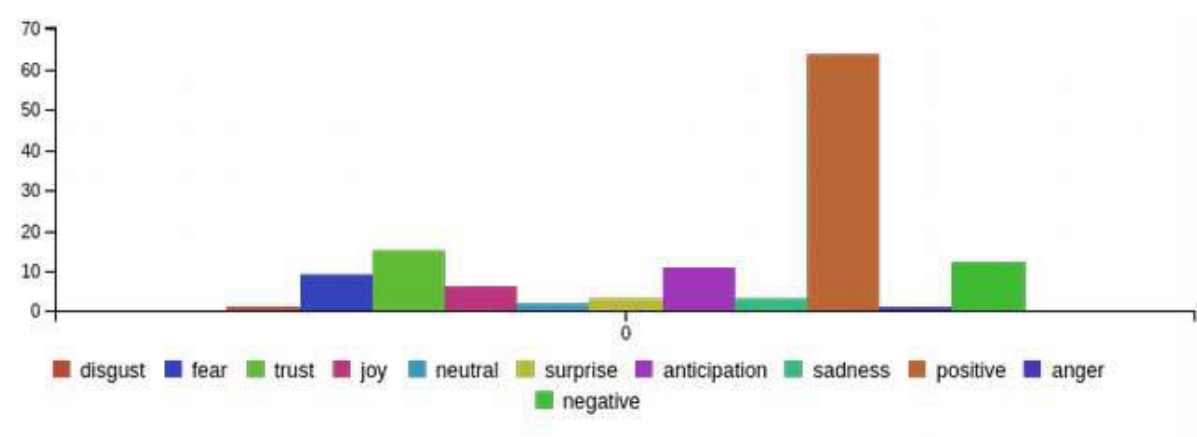


图 4.24: 2019/04/30 中国发起的孵化项目之情感分析

显示的底部五种情绪是悲伤 ( 3 ), 惊讶 ( 3 ), 中性 ( 2 ), 愤怒 ( 1 ) 和厌恶 ( 1 )。

注：这三种最强烈的情绪与 Apache HTTP 服务器和非中国人发起的孵化项目一样，而且分布也差不多。其他情绪有一些波动，但并不明显。

4.3.7 标志 6b: 2019 年 4 月 30 日的中国发起的孵化项目之情感分析比较

下图显示了截至 2019 年 4 月 30 日，中国发起的孵化项目的情绪对比分析。

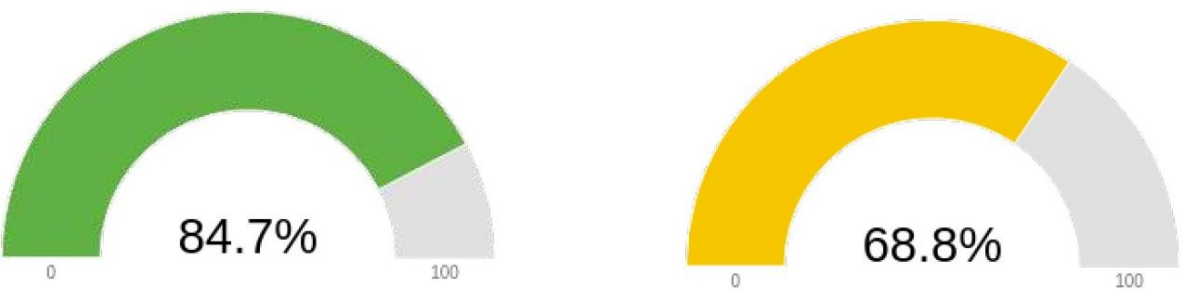


图 4.25: 中国发起的孵化项目与 ASF 的情绪比较

上图显示了中国发起的孵化项目的相对情绪。相对来说，他们沟通的情绪是非常积极的（84.7），并且高于非中国发起的孵化项目。

当这个情绪与 Apache Kibble 中的其他 ASF 项目的情绪（目前为 63）进行比较时，情绪（68.8）显示出所表达的情绪与其他 ASF 项目的强度不在同一水平。因此，这意味着中国发起的孵化项目在表达他们的情绪时，并不像其他 ASF 项目那样强烈。

### 4.3.8 标志 7: 中国发起的孵化项目之随时间推移的情感分析

下图是 2018 年 11 月至 2019 年 4 月中国人发起的孵化项目的情绪分析。

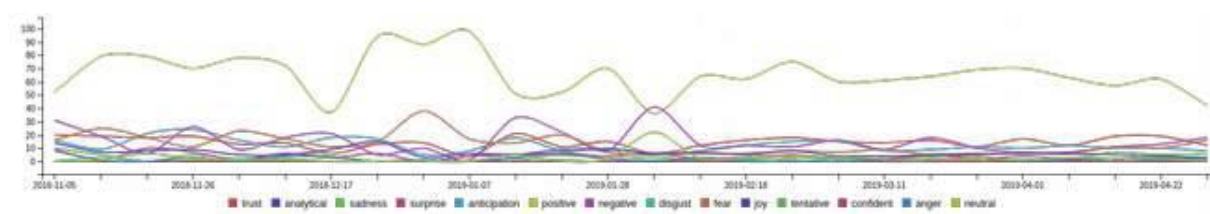
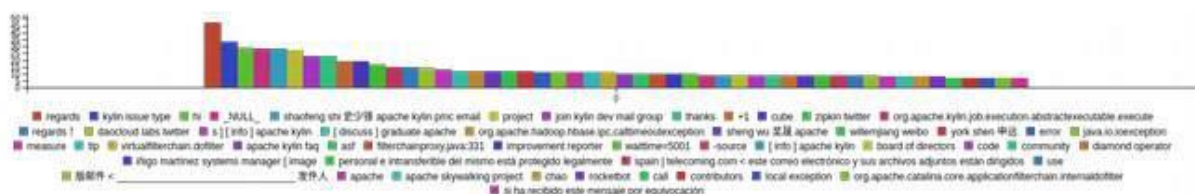


图 4.26：中国发起的孵化项目随时间变化的情感分析 2018 年 11 月 - 2019 年 4 月

随着时间的推移，分析表明，积极的沟通方式已经确立，并且是最强烈的情绪表达。这种情绪的变化确实比 Apache HTTP Server 和非中国发起的孵化项目要大得多。2019 年 2 月 4 日，负面情绪比正面情绪高。

### 4.3.9 标志 8: 中国发起的孵化项目之关键词提取

下图是截至 2019 年 4 月 30 日，中国发起的孵化项目中最常用的关键词句的提取分析。



**图 4.27：中国发起的孵化项目之关键词语的提取：2019/04/30**

这些词句可分为几个方面：

- 一般或标准（例如，信息、感谢、电子邮件、问候、姓名等）。
- 技术讨论（如 NULL、cube、错误、措施、局部异常、rocketbot 机器人、代码、改进报告）。
- 协作（如使用、项目、毕业讨论、问题、议题、呼吁、改进报告人、贡献者）。
- 文化（例如：社区、+1、列表、Apache）。

有几个文化指标与 ASF 的价值观相一致，例如：

- 开放：在公开场合做技术讨论
- 共识：+1 = 达成共识
- 协作：礼貌沟通、信息请求、问题分配、问题解决
- 社区：邮件列表是交流的媒介

## 4.4 绕过孵化的非中国项目

### 4.4.1 标志 1：绕过孵化的非中国项目之代码库的小马系数

下图显示了绕过 Apache 孵化过程的四个非中国发起的项目的代码库的小马系数（Pony Factor）。

这张图与 Apache HTTP 服务器项目基准线、非中国发起的孵化项目和中国发起的孵化项目的任何结果都不同。

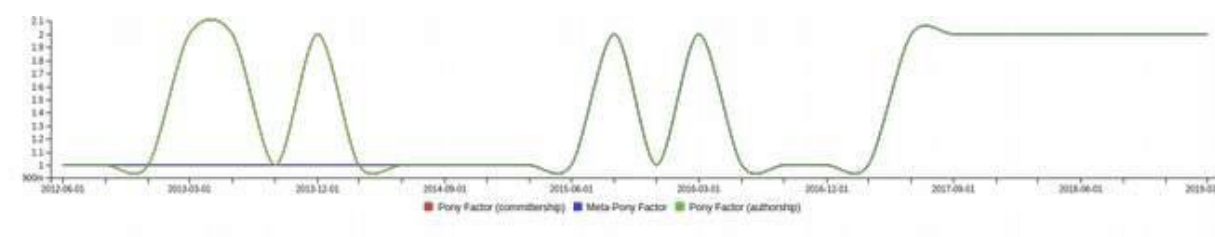


图 4.28：绕过孵化的非中国项目之代码库的小马系数 2012 - 2019

此处的小马系数的增加和减少似乎并不稳定，也许是由于关键人物间歇性地活动造成的，另一个原因可能是贡献者没有得到留存。

#### 4.4.2 标志 2：绕过孵化的非中国项目之邮件列表的小马系数

下图显示了非中国发起的绕过孵化期项目的小马系数（Pony Factor）。

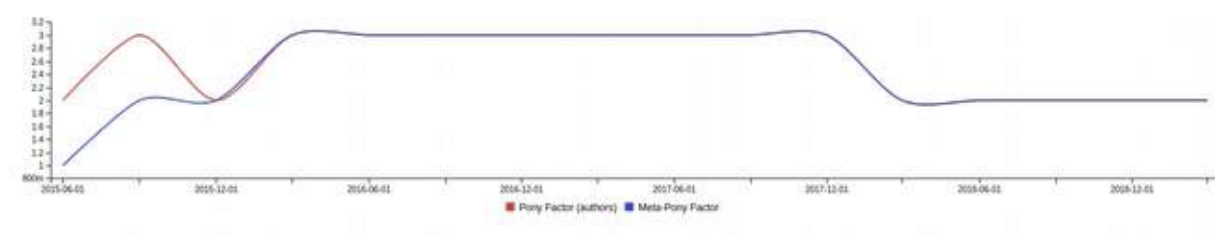


图 4.29：绕过孵化的非中国项目之邮件列表的小马系数 2015 - 2019

这个指标的曲线看起来与非中国发起的孵化项目和中国发起的孵化项目非常不同。然而，它确实有一些与 Apache HTTP 服务器曲线类似的特征，即小马系数上升，变平，然后降低，稳定在 2。

#### 4.4.3 标志 3：绕过孵化的非中国项目之贡献者经验

下图显示了贡献者对代码库贡献的时间长度的分类：

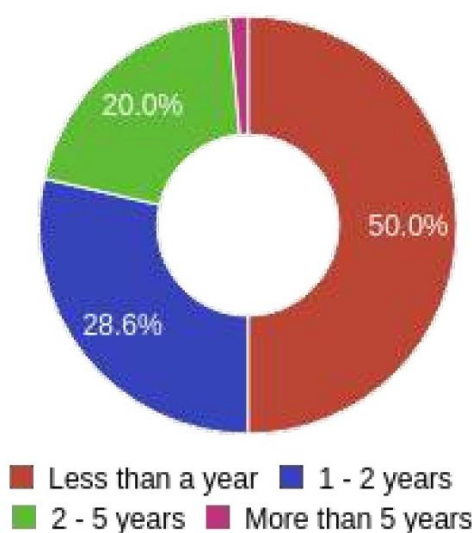


图 4.30：绕过孵化的非中国项目之贡献者经验

- 50% 的贡献者对项目的贡献不超过一年
- 超过 28% 的贡献者已经贡献了 1-2 年的时间
- 超过 20% 的贡献者参与了 2-5 年
- 已经参与了 5 年以上的人不到 2%

这里似乎有一个很好的混合和流动的贡献者，他们中的大多数人都不到一年的时间。

#### 4.4.4 标志 4：绕过孵化的非中国项目之代码库的贡献者留存

下图显示了有多少人被留存为社区的一部分并为代码库做出贡献。

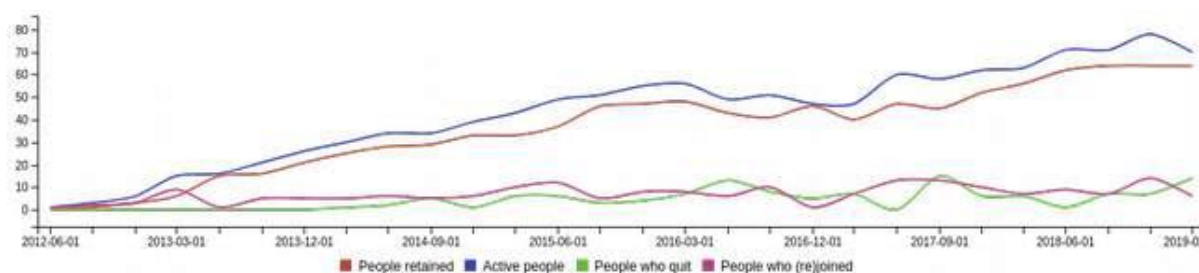


图 4.31：绕过孵化的非中国项目之代码库的贡献者留存 2011-2019

这显示了一个逐渐增加并略有下滑的趋势。这与非中国发起的孵化项目和中国发起的孵化项目的情况非常相似。

截至编写本报告时，目前 2019 年的数据显示：

- 70 名活跃者
- 64 人留存
- 6 人重新加入
- 14 人退出

注：与非中国发起的孵化项目（约 1,500 人）和中国发起的孵化项目（约 700 人）相比，这些项目的社区要小得多。

4.4.5 标志 5：绕过孵化的非中国项目之留存贡献者的邮件列表统计

下图显示了有多少人被留存为社区的一部分并为邮件列表的讨论做出贡献。

这张图不看 Apache HTTP 服务器的基准线，但与非中国发起的孵化项目和中国发起的孵化项目相似。邮件列表里的贡献者数随着时间的推移逐渐增加，并且仍在上升。这意味着贡献者的数量正在积极增长，而且他们正在被留存下来。

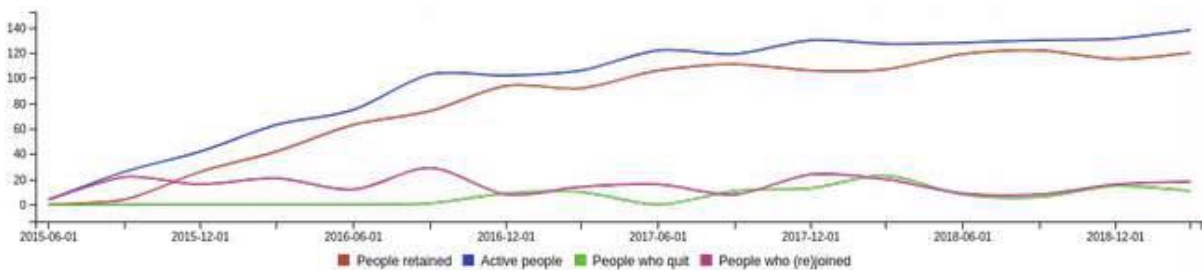


图 4.32：绕过孵化的非中国项目之邮件列表的贡献者留存 2012 - 2019

4.4.6 标志 6a: 2019/4/30 绕过孵化的非中国项目之情感分析

下图是截至 2019 年 4 月 30 日，绕过孵化的非中国发起的项目之综合情绪分析。

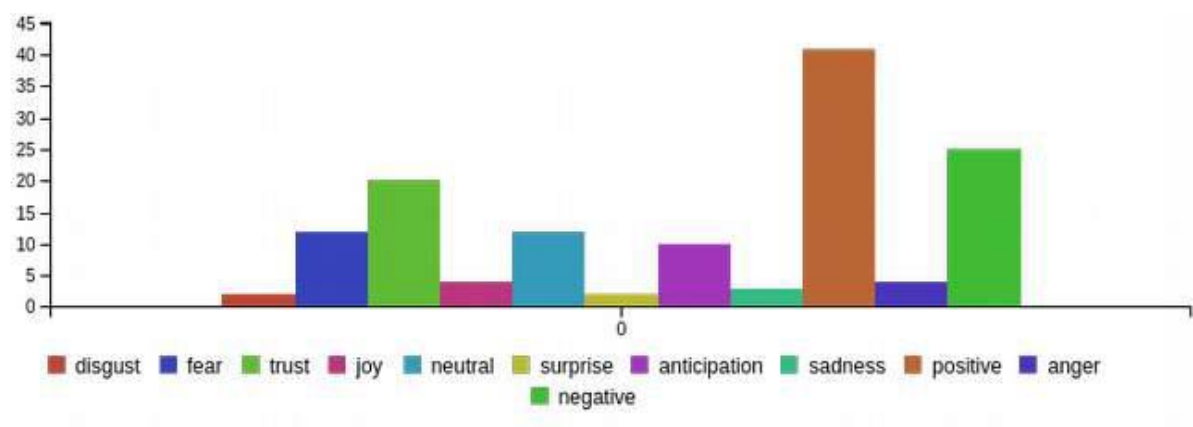


图 4.33：非中国发起的非孵化项目之情感分析 2019/4/30

最高的情绪情绪是积极的（45），第二是消极的（25），其次是信任（20），以及中性（21）和恐惧（12）。

底部的五种情绪是快乐（4），愤怒（4），悲伤（3），惊讶（2）和厌恶（2）。

4.4.7 标志 6b: 2019/4/30 绕过孵化的非中国项目之情感分析比较

下图是截至 2019 年 4 月 30 日，绕过孵化的非中国发起的项目的心情对比分析。

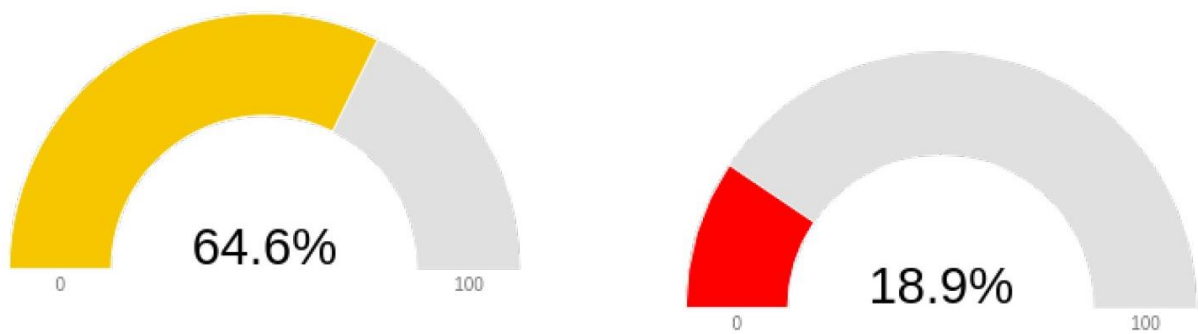


图 4.34：绕过孵化的非中国项目与 ASF 的情绪比较

上图显示了绕过孵化的非中国发起的项目的相对情绪。基于他们的沟通，相对情绪非常混杂



( 64.6 ) , 低于中国发起和非中国发起的两组孵化项目。

当这种情绪与 Apache Kibble 中的其他 ASF 项目 ( 目前为 63 ) 相比 , 其情绪 ( 18.9 ) 表明所表达的情绪与其他 ASF 项目的强度不在同一水平。这里的情绪要弱得多。

#### 4.4.8 标志 7: 绕过孵化的非中国项目之随时间推移的情感分析

下图是 2018 年 11 月至 2019 年 4 月绕过孵化的非中国发起项目的情绪分析。

随着时间的推移, 情绪看起来很不稳定。显示的最主要的情绪是积极的。有的时候, 积极性很低, 然而只有一个明显的高峰, 在一个短时期内, 消极性是最主要的。

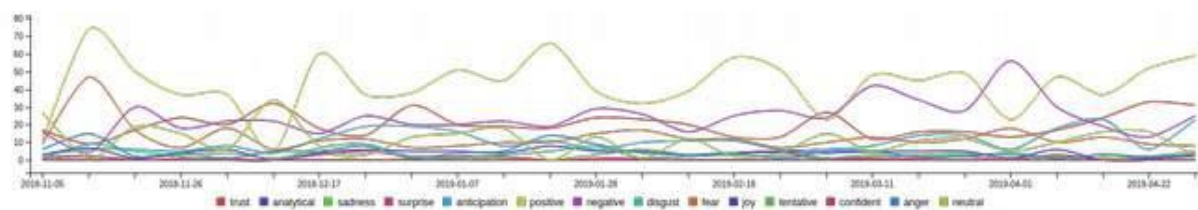


图 4.35 : 非中国人发起的非孵化项目随时间变化的情感分析 2018 年 11 月 - 2019 年 4 月

#### 4.4.9 标志 8: 绕过孵化的非中国项目之关键词句提取

下图是截至 2019 年 4 月 30 日, 绕过孵化的非中国人发起的项目中最常用的关键词句提取分析。

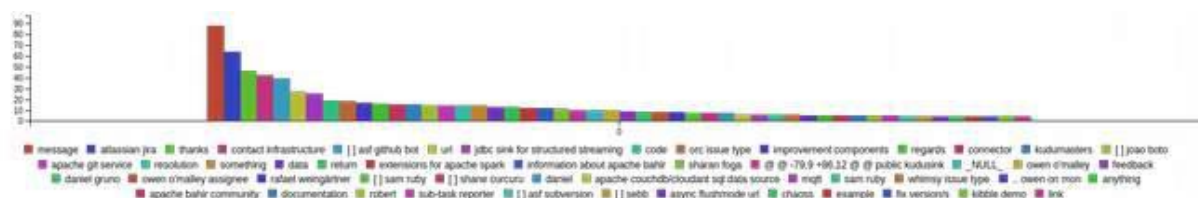


图 4.36 : 2019/04/30 绕过孵化的非中国项目之关键词句提取

这些词句可以细分为以下几个方面：

- 一般或标准 ( 如信息、感谢、名字等 )

- 技术讨论（如联系基础设施、代码、连接器、结构化流的 jdbc、apache git 服务数据源）
- 协作（如网址、问题类型、改进内容、解决方案、数据、反馈、文件、反馈）
- 文化 (Apache Bahir 社区)

与 ASF 价值观相一致的几个文化指标如下：

- 开放性：在公开场合进行技术对话
- 协作：礼貌沟通、信息请求、问题分配、问题解决
- 社区：Apache bahir 社区

很少使用文化语言，最重要的是在这些项目的关键词句提取（KPE）分析中没有 +1 共识指标。

## 第 5 章 分析结果

本研究的目的是为了更好地了解 Apache 软件基金会的开源文化，并分析一个具有不同文化特征的项目，以了解在该项目演变和毕业成为顶级项目的过程中是否会发生文化变化。

本研究聚焦于四个中国贡献的项目，这些项目都已经或正在接受 Apache 软件基金会的孵化。

我们选择了 12 个 ASF 项目，并将其分为三组，与 Apache HTTP 服务器项目之文化基准线进行比较。同时他们相互之间也进行了比较。这些小组的情况如下：

- 第 1 组：由中国贡献者发起的四个经过孵化的 Apache 项目。
- 第 2 组：不是由中国贡献者发起的四个经过孵化的 Apache 项目。
- 第 3 组：不是由中国人发起的四个没有经过孵化的 Apache 项目。

### 5.1 Apache HTTP 服务器项目之文化基准线

下面的结果显示了各组与 Apache HTTP 服务器项目之基准线的比较。

#### 5.1.1 小马系数

这三个小组中没有一个在代码库或电子邮件方面遵循 Apache HTTP 服务器的基准线。除了非孵化项目外，代码库作者和提交者之间存在着普遍的差异，这表明了孵化项目组也许没有那么频繁地根据贡献者的活动来承认其功绩。

在有差异的地方，中国人发起的项目显示出代码库作者和提交者之间的差距最小，这也许表明他们在承认功绩方面更加慷慨。这有可能是集体主义文化的一种表现，在这种文化中，加入集体的人更容易得到认可。

这三组的电子邮件的小马系数都是不同的，同时没有一个与 Apache HTTP Server 项目的基

准线相匹配。

### 5.1.2 贡献者留存

在这个领域，似乎没有一个小组遵循 Apache HTTP 服务器的基准线。所有三个小组的大多数贡献者都是由新的贡献者组成的。在中国发起的项目中，超过 90% 的贡献者的贡献时间少于 2 年，而在非中国发起的孵化项目中，这一比例为 85%，在非孵化的项目中为 70%，Apache HTTP Server 项目则为 19%。

这意味着这三组都能成功地吸引人们做出贡献。并有一个良好的人流进入他们的社区。

贡献者的留存率也很高，这与 Apache HTTP 服务器项目的基准线有很大的不同。

### 5.1.3 情感分析

这三组项目的情况都与 Apache HTTP 服务器的基准线相似。沟通方式总体上是积极的，尽管消极性被列入所有组别的前 5 名，但从长期来看，它并没有在项目中占主导地位。

随着时间的推移，中国发起的项目和非孵化项目的情绪并不像 Apache HTTP 服务器那样地保持前后一致。

从情绪的比较和情绪的强度来看，只有非孵化项目的小组与 Apache HTTP 服务器项目有类似的比较情绪。这表明，Apache HTTP 服务器项目和非孵化项目的情绪比其他所有项目都要低很多。

### 5.1.4 关键语句分析

所有这三组项目似乎都有类似于 Apache HTTP 服务器基准线的特征。他们都显示了标准的日常交流和技术互动的元素。有趣的是，它们也显示了 ASF 价值的文化表达，如开放性、协作和社区。

在基准线以及三个小组中的两个小组中出现的最重要的文化元素是 "+1" 指标。这是 ASF 所特有的，作为一种共识的指标。

唯一没有出现 "+1" 指标的组别是未孵化的项目组。

## 5.2 中国发起的孵化项目 vs. 非中国发起的孵化项目

下面的结果显示了中国发起的孵化项目与非中国发起的孵化项目的比较。

### 5.2.1 小马系数

中国发起的孵化项目和非中国发起的孵化项目之代码库的小马系数看起来都差不多。它们都在增加，而非中国发起的项目的曲线更陡峭，显示出更快的增长。

在中国发起的项目中，随着时间的推移，代码库作者和提交者之间的差距比非中国发起的项目要小得多。对于中国项目来说，这有可能是集体主义团体的表现。

目前，在作者和提交者之间存在着差距。在中国发起的项目中，差距为 6，而在非中国项目中，差距为 13。

然而，从电子邮件的小马系数来看，非中国发起的项目增长很快，已经稳定在 35。中国发起的项目的小马系数一直在下降，目前稳定在 2。对于中国发起的项目，这意味着他们的综合邮件列表上的电子邮件信息通常集中在 2 人左右。

一个小马系数显示了活动和社区的增长，另一个显示了有限的沟通，对于中国发起的项目来说，这可能意味着主要的电子邮件流量是由极少数人产生的。这很有趣，因为这可能表明，对话可能发生在其他地方，因为不是所有的社区都能说流利的英语，或者这些社区需要很少的沟通来启动工作。

(译者注：也有可能是中国的贡献者更习惯使用即时通讯工具，如微信或 QQ)

## 5.2.2 贡献者留存

中国发起的和非中国发起的项目都能很好地吸引和留住贡献者。

在非中国发起的项目中，超过 85% 的贡献者拥有不到 2 年的经验。在中国发起的项目中，这个数字甚至更高，91% 的贡献者拥有不到两年的经验。这意味着中国发起的项目在吸引新贡献者方面是最成功的。

在这两个群体中，也有一些人一直留在社区内，目前处于 2-5 年的范围内。随着项目的成熟，我们希望看到这个数字会有一些增长，因为新的贡献者与项目一起成长。

从贡献者留存时间来看，我们可以清楚地看到，对于中国发起的项目，当他们加入 Apache 孵化器时，他们的贡献者留存时间增加了。中国发起的项目和非中国发起的项目之间的贡献者群体规模非常相似：

- 非中国发起的孵化项目有 631 位活跃贡献者
- 中国发起的孵化项目有 518 位活跃贡献者

就电子邮件而言，中国发起的项目的贡献者人数为 705 人，不到非中国发起的项目的贡献者人数 ( 1,588 ) 的一半。

## 5.2.3 情感分析

在中国人发起的项目团体和非中国人发起的项目团体中，最重要的情绪是积极的。信任和消极的情绪也都很强烈。

对两组的时间分析表明，积极的沟通方式已经确立，是表达的最强烈的情绪。对于中国人发起的项目，积极性的强度确实比非中国人的项目变化大得多，在 2019 年 2 月期间，消极情绪一

度高于积极情绪。

两组的总体情绪都超过 80%，但与其他 Apache 项目相比，情绪的强度有很大不同。非中国发起的项目的比较情绪超过 84%，这意味着他们的情绪与其他 ASF 项目的情绪非常接近。对于中国发起的项目，比较情绪要低得多，只有 51%。这意味着中国发起的项目在表达他们的情绪时并不像其他 ASF 项目那样强烈。

在中国的项目中，这可能仅仅是在使用外语时通常会遇到的文化缄默症。所使用的措辞和句子类型可能缺乏情感，因为他们专注于交流具体的技术任务，而不是社交上的寒暄。

#### 5.2.4 关键语句分析

中国人发起的项目群体和非中国人发起的项目群体都有类似的特征。他们展示了标准的日常交流和技术互动的元素。有趣的是，他们也显示了 ASF 价值观的文化表达，如开放、合作和社区。

在中国发起和非中国发起的项目中出现的最重要的文化元素是 "+1" 指标。这是 ASF 所特有的，作为一种共识指标。

### 5.3 非中国发起的非孵化项目 vs 中国发起的孵化项目

#### 5.3.1 小马系数

非孵化项目之代码库的小马系数显示出不稳定的模式，增加和减少都不一致，这可能与关键人物的间歇性活动有关。中国发起的项目的小马系数正在增加，并且仍在增加。

非孵化项目显示，作者和提交者之间的小马系数没有区别。这可能是因为参与项目的人已经是 ASF 内部有经验的提交者或长期的指导者。绕过孵化的一个关键原因是，社区或大多数贡献者已经熟悉了 ASF 文化。

在中国发起的项目中，代码库作者和提交者之间几乎没有差距。这种比较很有意思，因为在 Apache HTTP 服务器的基准线中，作者和提交者之间没有任何区别，而在非孵化项目中，我们也看到了同样的模式。这似乎表明，中国发起的项目与 Apache HTTP 服务器项目和非孵化项目比非中国发起的项目更为相似。

非孵化项目的电子邮件的小马系数上升，稳定，但现在已经下降到与中国发起的项目相同（2）。这表明沟通是以少数人为中心的。

### 5.3.2 贡献者留存

在非孵化组和中国发起的组中，似乎都有一个很好的混合和流动的贡献者，他们中的大多数参与不到一年的时间。

中国发起的项目有 90% 以上的新贡献者，因此在吸引新的贡献者方面比非孵化项目更成功。非孵化项目有更多的贡献者组合，这表明贡献者过渡到长期参与其中。

非孵化项目和中国发起的项目之间的贡献者群体规模非常不同：

- 中国发起的项目有 518 位活跃贡献者
- 非孵化的项目有 70 位活跃贡献者

就邮件列表而言，中国发起的项目之邮件讨论的贡献者人数很高，为 705 人，而非孵化项目的贡献者人数为 138 人。

### 5.3.3 情感分析

中国发起的项目和非孵化项目组的首要情绪是积极的。信任和消极也是一种强烈的情绪。

对两组的时间分析表明，积极的沟通方式已经确立，是表达的最强烈的情绪。对于中国人发起的项目，积极性的强度变化比非中国人的项目大得多，在 2019 年 2 月期间，消极情绪一度



高于积极情绪。

非孵化项目也有负面情绪高于正面情绪的时候。这包括一个峰值，即负面情绪在很短的时间内是最主要的。

非孵化组的总体情绪评分为 63%，而中国发起的项目为 84%。对比 ASF 内其他项目的情绪强度，非孵化项目为 18%，而中国发起的项目则为 68%。

在这种情况下，非孵化项目的情绪水平并不像其他 ASF 项目那样强烈。这表明，中国发起的项目实际上比非孵化项目更能代表 ASF 的情绪。

这可能是由于在非孵化项目中使用的交流类型所造成的，这些项目缺乏情感，或者更专注于具体的技术任务。这也可能是因为现有的有限数量的贡献者已经了解如何一起工作，不需要大量的沟通。

### 5.3.4 关键词句分析

非孵化组和中国发起组都有类似的特征。他们展示了标准的日常交流和技术互动的元素。有趣的是，他们也显示了 ASF 价值的文化表达迹象，如开放性、协作和社区。

非孵化项目使用的文化语言非常少。在这个比较中，最重要的是 "+1" 指标出现在中国发起的项目中，而不是在非孵化项目中。这个指标是 ASF 所特有的，是衡量共识的指标。

这表明，非孵化项目要么不使用共识指标，要么使用得很少，可以忽略不计。

## 5.4 非中国发起的孵化项目 vs 非中国发起的非孵化项目

### 5.4.1 小马系数

非孵化项目之代码库的小马系数显示出一种不稳定的模式，不稳定的增加和减少可能与关键人

物的间歇性活动有关。非中国发起的孵化项目正在快速增长，并且仍在上升。

非孵化项目显示，作者和提交者的小马系数之间没有区别。这可能是因为参与项目的人已经是 ASF 内部有经验的提交者或长期的指导者。

目前，在非中国的孵化项目和非孵化的项目中，作者和提交者的小马系数有很大差距。

非中国发起的孵化项目的电子邮件小马系数迅速增长并稳定在 35，而非孵化项目的小马系数有所下降，目前为 2。

### 5.4.2 贡献者留存

在非孵化项目组和从中国的孵化项目组中，似乎都有很好的贡献者组合和流动，他们中的大多数人参与了不到一年的时间。

非中国发起的孵化项目组里有 85% 的贡献者拥有不到两年的经验，表明这些项目在吸引新的贡献者方面是成功的。而非孵化项目组有更多不同贡献时长的贡献者组合，这表明他们的贡献者逐渐地过渡到长期参与项目。

非孵化项目和非中国人发起的孵化项目之间的贡献者群体规模有显著的不同：

- 非中国人发起的孵化项目有 631 位活跃贡献者
- 非孵化项目则有 70 位活跃贡献者

就邮件列表而言，非中国发起的孵化项目的邮件讨论贡献者人数很高，达到 1,588 人，而未被孵化的项目则为 138 人。

### 5.4.3 情感分析

非中国人发起的孵化项目组和从孵化项目组的主要情绪是积极的。信任和负面情绪也作为强烈

的情绪出现。对两组的分析表明，积极的沟通方式已经确立，并且通常是最强烈的情绪表达。非孵化项目确实存在消极情绪高于积极情绪的情况，并且包括一个消极情绪在短时间内占主导地位的高峰。

非孵化组的总体情绪为 63%，而非中国发起的孵化项目为 81%。对比 ASF 内部其他项目的情绪强度，非孵化项目为 14%，而非中国发起的孵化项目为 84%。

在这种情况下，非中国发起的孵化项目实际上比非孵化的项目更能代表 ASF 的情绪。

#### 5.4.4 关键语句分析

非孵化项目组和由非中国人发起的孵化项目组都有类似的特征。他们显示了标准的日常交流和技术互动的元素。他们也在不同程度上表现出 ASF 价值观的文化表达，如开放、合作和社区。

非孵化的项目很少使用文化语言。在这个比较中最重要的是，“+1”指标出现在非中国发起的孵化项目中，而非孵化的项目没有。这个指标是 ASF 所特有的，是衡量共识的指标。

这表明，非孵化项目要么不使用共识指标，要么使用得太少，以至于没有意义。

## 第 6 章 结论

本研究的目的是为了更好地了解 Apache 软件基金会的开源文化，并分析一些具有不同文化特征的项目，以了解在其项目演化和毕业于顶级项目的过程中是否会发生文化变化。

毕业和成为顶级项目的一个关键部分涉及 Apache 行为和文化价值观的展示。

本研究聚焦于四个中国贡献的项目，这些项目都已经或正在接受 Apache 软件基金会的孵化。

本研究报告希望回答以下特定的研究问题：

1. 我们能找到哪些文化嵌入的证据？
2. 中国贡献的项目和非中国贡献的项目之间有多大的文化差异？
3. Apache 孵化器在将 ASF 文化嵌入中国贡献的项目中的成功程度如何？

本研究使用了一套工具和指标，根据第一个 ASF 项目 -- Apache HTTP 服务器 -- 所表现出来的价值观和行为，创建了一个文化基准线。

通过挖掘 ASF 项目归档资料中公开的数据，使用了一套 8 个指标来创建基准线。

对于 Apache HTTP 服务器项目的基准线，使用了以下指标来捕捉、强调和衡量：

- 小马系数 (Pony Factors) :
  - 社区的多样性，确认功绩得到了奖励，社区发展的迹象，贡献者留存
- 感性分析 (Sentient Analysis) :
  - 在社区互动中表现出的主要情绪、沟通方式、沟通的整体情绪 (消极、积极或中性)，随着时间的推移，潜在的文化元素的范围

- 关键词语分析 ( Key Phrase Analysis )
  - 识别正在使用的最常见的重要词句和单词，表明协作，识别独特文化语言的使用

我们选择了 12 个 ASF 项目，并将其分为三组，与 Apache HTTP 服务器文化基准线进行比较。各组之间也进行了相互比较。这些小组的情况如下：

- 四个经过孵化的 Apache 项目是由中国贡献者发起的
- 四个经过孵化的 Apache 项目，不是由中国人发起的
- 四个没有经过孵化的 Apache 项目，不是由中国人发起的

## 6.1 对研究问题的回应

### 6.1.1 我们能找到哪些文化嵌入的证据？

分析结果显示如下：

- 小马系数的增长表明功绩得到了认可
- 贡献者分类显示，新人正在被欢迎进入项目社区
- 贡献者留存率的细分显示，新贡献者正在转变为长期贡献者
- 正在使用文化语言，这包括 "+1" 共识指标

关键的发现是，孵化项目比非孵化项目更善于使用 "+1" 指标。

### 6.1.2 中国贡献的项目和非中国贡献的项目之间有多大的文化差异？

分析结果显示如下：

- 中国项目似乎比非中国项目能更快地认可功绩
- 在吸引新的贡献者方面，中国项目比非中国贡献的项目更成功
- 非中国项目比中国贡献的项目在更高、更频繁的水平上进行交流

( 译者按：主要指的是透过邮件列表的沟通 )

- 非中国项目的情绪和情绪的强度更能反映出其他 ASF 项目的情况，相比之下，中国贡献的项目似乎有点低调
- 中国的孵化项目和非中国的孵化项目都大力采用 ASF 文化语言，并经常使用 "+1" 共识指标

关键的发现是，中国的孵化项目和非中国贡献的孵化项目都强烈地采用了 ASF 的文化行为和语言，并且是 "+1" 共识指标的频繁使用者。

### 6.1.3 Apache 孵化器在将 Apache 文化植入中国贡献的项目中的成功程度如何？

分析表明，Apache 孵化器在将 Apache 文化植入中国贡献项目方面非常成功。中国贡献的项目：

- 是最善于识别其贡献者的功绩的
- 在吸引新的贡献者到他们的项目上是最成功的
- 已适应使用 Apache 文化语言
- 已经采用了 "+1" 共识指标，并正在大量使用它

关键的发现是只有被孵化的项目是 "+1" 共识指标的频繁使用者。

#### 【专家点评】

江波：社区的本质是共识，而共识的背后是社区文化的传承，包括共同的信仰、社区规则的建立和独特的话语体系，比如此处提到的 "+1" 共识指标正是验证 Apache 孵化器在将 Apache 文化植入中国贡献的项目中的成功程度的一个很有趣的指标。

姜宁：Apache 文化的植入与 ASF 孵化有密切的关系，孵化器为大家提供了一个学习 Apache Way 的场所，在进行孵化的过程中还是有很多文化冲突的，例如喜欢私下解决问题，不愿意把自己的问题以邮件的方式公开暴露出来。文章中的研究方法集中在邮件的分析

会有一定的局限性，因为很多新志愿者的招募都是通过线下或者 IM 平台来进行实现的。另外由于非英语母语的关系，来自于中国的开发者邮件中大多只是陈述问题，很少对自己的情感进行表露。“+1”让大家可以用一个比较简单的方式表达自己的看法，但在提案阶段的讨论，据我观察来自中国的开发者还是很少愿意直接地表达自己的看法。

## 6.2 摘要

开源实现并促进了分布式开发。它与一般软件开发的不同之处在于，它是一个开发者分享他们所创造的东西的环境，这样其他人就可以看到已经完成的东西并可选择是否自愿地参与贡献。这种代码的分享是“礼物文化”的基础，在这里，软件是被给予的礼物。正是这种软件的礼物，有助于凝聚和建立一个社区。(Pan & Bonk, 2007)。

这项研究考察了 ASF 的中国贡献项目和非中国贡献项目，发现了文化嵌入的证据。在情绪、情感、贡献者和语言方面，孵化的项目比非孵化的项目表现出更多的文化特征。

非孵化项目似乎没有那么注重社区的发展，但仍然能够保留贡献者。非孵化项目不是充满了有经验的贡献者，而是仍然吸引着新人。一个脆弱的领域是，这些社区似乎以少数人为中心。

在不同的项目组以不同的强度表达自己和自己的情感时，存在着沟通的差异。中国发起的项目似乎符合注重行动而非语言的模式。因此，虽然在沟通方式上有明显的差异，但注重作为一个社区携手工作来实现某些目标则是共同的中心思想。

那么，权力距离以及接受和期待不平等的能力是什么？事实上，在这种情况下，中国的项目是最善于认可功绩的，这似乎意味着他们懂得如何利用他们的权力来鼓励贡献者社区内的平等。

### 【专家点评】姜宁

由于层级管理的观念，和对于权威的崇拜，大部分的中国开源项目开发并不擅长将项目的权力分享给社区开发者。很少见到大厂程序员间自发组织起来的共同开发的项目，更多的是大厂间的同质竞争。

在每个项目进入孵化期之前，都有一个围绕它的现有社区。尽管它可能很小，但社会互动肯定已经发生，以使项目得以进展和生存。

在集体主义方面，ASF 文化和中国发起的项目可能没有那么大的差别，因为 ASF 的 "社区高于代码" 的口号似乎是这些项目中已经存在的东西。中国的项目在某些方面与 Apache HTTP Server 更为相似，这说明这种行为是自然发生的，而且已经内化了。

非中国人发起的项目的小马系数 (Pony Factors) 的差异可能与从个人主义思维过渡到集体主义或社区思维所需的时间有关。

其中一些项目社区可能起源于企业社区，因为一些企业已经创建了内部项目，只是后来将其开源。其结果是权力的过渡，因为项目的方向从企业主导转向社区主导。因此，这也可能是我们在本研究中看到的一些结果。

随着持续的增长和吸引新项目和新贡献者的能力，ASF 表明它有一些独特的东西，使个人愿意参与和参加。这种 "独特的东西" 是基于社会互动、文化和社区。

如果能够在本文成果的基础上，在以下领域进行进一步研究，将是非常有益的：

- 深入分析以 "+1" 指标为重点的关键词句，以追踪其与社区、邮件列表和相关系列议题的关系
- 调查一个项目社区或项目管理委员会 (PMC) 的初始规模是否会影响其通过 Apache 孵化器的速度和发展
- 对企业可能参与的项目进行调查，分析文化或社区行为的任何差异
- 对有经验的导师的项目进行分析，看他们是否比有经验不足的导师的项目在孵化过程中进展更快
- 分析与 ASF 非相关的开源项目，并与 ASF 项目进行比较，以确认 ASF 的价值观和行为是否具有文化的特殊性/独特性



- 对已进入衰退期的 ASF 项目进行调查，看是否可以用本文使用的任何指标来预测其衰退

## 附录 A（参考文献）

Carillo, K., & Okoli, C. (2009). THE OPEN SOURCE MOVEMENT: A REVOLUTION IN SOFTWARE DEVELOPMENT. *The Journal of Computer Information Systems*, 49(2), 1-9.

Sharifian, F. (2017). *Cultural Linguistics: Cultural Conceptualisations and Language*. Amsterdam: John Benjamins Publishing Company.

Hofstede, G., Hofstede, G. J. & Minkov M. (2015). *Cultures and organizations: Software of the mind* (Rev. and expand 3rd ed.). New York: McGraw-Hill.

Zhou, Y. (2011). AGAINST INTELLECTUAL MONOPOLY: FREE SOFTWARE IN CHINA. *World Review of Political Economy*, 2(2), 290-306.

Pan, G., & Bonk, C. J. (2007). The Emergence of Open-Source Software in China. *The International Review of Research in Open and Distributed Learning*, 8(1).

Ceraso, A., & Pruchnic, J. (2011). INTRODUCTION: OPEN SOURCE CULTURE AND AESTHETICS. *Criticism*, 53(3), 337-375.

Caldwell-Harris, C., Kronrod, A. & Yang, J. (2013). Do more, say less: Saying “I love you” in Chinese and American cultures. *Intercultural Pragmatics*, 10(1), pp. 41-69. Kelty, C. M. (2004). Culture’ s open sources: Software, copyright, and cultural critique. *Anthropological Quarterly*, 77(3), 499-506.

Incubation Policy, Apache Incubator

<https://incubator.apache.org/policy/incubation.html>

Apache Project Maturity Model <http://community.apache.org/apache-way/apacheproject-maturity-model.html>

Incubator Proposals

<https://wiki.apache.org/incubator/KylinProposal>

<https://cwiki.apache.org/confluence/display/incubator/DubboProposal>

<https://wiki.apache.org/incubator/SkyWalkingProposal>

<https://cwiki.apache.org/confluence/display/incubator/WeexProposal>

Allen, D. (2003). Cultural conundrum. *Nursing Standard* (through 2013), 17(43), 16-8.

Seidler, Katie. *Crime, Culture and Violence : Understanding How Masculinity and Identity Shapes Offending*, Australian Academic Press, 2010. ProQuest Ebook Central,

Rose, M. (2007). THE TRANSMISSION OF CULTURE IN A GLOBALIZED WORLD: THE CHALLENGE OF NEW TECHNOLOGIES. *Canadian Issues*, , 55- 57.

Murray, P. (2006). *Open sources 2.0: The continuing evolution*. Weybridge: Birchley Hall Press.