

提示：采用手机 safari 微软翻译技术

1. 地平线: facebook 的开源强化学习平台

作者:[jason gauci](#), [edoard conti](#), [yitaoliang](#), [kittipat virochsiri](#),
[yuchen he](#) , [zachary kasen](#), [vivek narayanan](#), [xiaohui ye](#)

文摘: 本文介绍了 "地平线"、"脸谱" 的开源应用强化学习 (rl) 平台。horizon 是一个端到端平台, 旨在解决行业应用的 rl 问题, 在这些问题上, 数据集很大 (数百万到数十亿的观测), 反馈回路速度很慢 (与模拟器相比), 并且必须谨慎地进行实验, 因为它们不是在模拟器中运行的. 与其他 rl 平台不同的是, "地平线" 的设计考虑到了生产用例的最高考虑。该平台包含训练流行的深度 rl 算法的工作流, 包括数据预处理、特征转换、分布式培训、反事实策略评估和优化服务。我们还展示了使用 horizon 培训的模型在 facebook 上的表现明显优于和取代监督学习系统的真实例子。

少

2018 年 11 月 1 日提交;最初宣布 2018 年 11 月。

2. 用于 jpeg 图像的基于灰度的图像加密的加密然后压缩系统

作者:[tatsuya chuman](#), [warit sirichhotedumrong](#), [hitoshi kiya](#)

摘要: 提出了一种基于块竞争的加密方案, 以增强具有 jpeg 压缩功能的加密时压缩 (etc) 系统的安全性, 从而使我们能够通过不受信任的通道提供商 (如社交网络) 安全地传输图像服务提供商。与传统方案相比, 该方案允许使用更小的块大小和更大的块数。使用该方案加密的图像由于使用灰度图像, 即使原始图像有三个颜色通道, 也包含较少的颜色信息。这些功能增强了抵御各种攻击的安全性, 如拼图求解器和暴力攻击。在一个实验中, 对拼图求解器攻击的安全性进行了评估。加密图像被上传到脸书和推特, 然后从脸书和推特上下载, 结果表明, 拟议的方案对 etc 系统有效。少

2018 年 11 月 1 日提交;最初宣布 2018 年 11 月。

3. 新闻文章中的社交媒体来源的大规模研究

作者:[md main uddin rony](#), [mohammad yousuf](#), [naeemul hassan](#)

摘要: 在本研究中, 我们密切关注社交媒体内容在美国新闻媒体中的使用来源或参考。具体而言, 我们审查了 153 家美国媒体在 2013–2017 5 年期间发表的约 6 万篇新闻文章, 并分析了与其他来源相比社交媒体内容作为来源的使用情况。我们设计了一个社交媒体源提取算法, 并调查了不同新闻主题的社交媒体源使用的程度和性质。我们的研究结果显示, 社交媒体内容在新闻中的使用在五年内几乎翻了一番。不可靠的媒体比主流媒体更依赖社交媒

体。主流和不可靠的网站都更喜欢推特而不是脸书作为信息来源。

少

2018 年 10 月 30 日提交;最初宣布 2018 年 10 月。

4. 基于灰度的加密在加密压缩系统中的压缩性能

作者:[warit sirichotedumrong](#), [tatsuya chuman](#), [hitoshi kiya](#)

摘要: 本文考虑了一种新的基于灰度的基于灰度的基于图像加密的加密系统——当时的压缩 (etc) 系统, 具有 jpeg 压缩功能。首先, 从色彩空间的选择角度讨论了灰度图像的生成方法。此外, 还提出了一种新的基于灰度图像的 jpeg 量化表, 以提供更好的压缩性能。此外, 还讨论和评估了上传到社交网络服务 (sns) 并从 sns 下载的图像的质量。在实验中, 使用各种压缩参数和量化表对加密图像进行压缩, 并上传到推特和脸书。结果表明, 颜色空间的选择和提出的量化表不仅可以提高上传图像的压缩性能, 还可以提高下载图像的压缩性能。少

2018 年 10 月 30 日提交;最初宣布 2018 年 10 月。

5. 跟踪实体的历史与演变: 以实体为中心的大型社交媒体档案的时间分析

作者:[pavlos fafaliotis](#), [vasileios iosifidis](#), [kostas stefanidis](#), [eirini ntoutsis](#)

摘要: 2015 年, 希腊总理的受欢迎程度如何演变? 在那段时间里, 对他的主要情绪有什么不同? 是否有任何有争议的时期? 在这期间, 还有哪些实体与他有关系? 要回答这些问题, 需要分析有关查询实体的存档文档和数据, 例如旧新闻文章或社交媒体档案。特别是, 在推特和**脸书**等社交网络中发布的用户生成的内容可被视为我们社会的全面文件, 因此, 对这些存档数据的有意义的分析方法具有巨大的价值。社会学家、历史学家和其他有关方面, 他们想要研究实体和事件的历史和演变。为此, 本文提出了一种以实体为中心的方法来分析社交媒体档案, 并确定了一些措施, 以便研究不同时期和不同方面(如受欢迎程度、态度)如何在社交媒体中反映实体的情况, 有争议, 以及与其他实体的联系。一项使用四年大型推特档案的案例研究说明了这种以实体为中心的多方面分析所能获得的见解。少

2018 年 10 月 24 日提交;最初宣布 2018 年 10 月。

6. 2017 年俄罗斯赞助内容在美国的战略与传播分析

作者: [亚历山大·斯班格尔](#), [gireeja ranade](#), [besmira nushi](#), [adam fourney](#), [eric horvitz](#)

摘要: 总部设在俄罗斯的互联网研究机构 (ira) 在 2016 年总统选举前后在美国开展了广泛的宣传活动。该组织创建了一套广泛的互联网属性: 网络域名、**facebook** 页面和 twitter 机器人, 通过购买 **facebook** 广告、推特和搜索引擎为其域名编制索引来接收

流量。我们调查 2017 年 ira 活动的范围, 将 **facebook** 和 twitter 的数据与 internet explorer 11 和 edge 浏览器以及 bing. com 搜索引擎的日志结合起来。这些研究既证明了恶意行为者可以很容易地利用社交媒体和搜索引擎开展宣传活动, 也证明了通过将内容和活动资源从多个互联网融合起来来跟踪和理解此类活动的能力服务。我们展示了跨平台分析如何能够为操纵民主国家的观点和选举的企图提供前所未有的视角。少

2018 年 10 月 23 日提交;最初宣布 2018 年 10 月。

7. 莫诺雷波斯: 多元文学述评

作者: [gleison brito](#), [ricardo terra](#), [marco tulio valente](#)

摘要: 单片存储库 (monorespo) 被谷歌和脸谱等大公司以及巴贝尔和恩伯等热门开源项目所使用。本研究概述了单点的定义和特点及其好处和挑战。于是, 我们对主要是灰色的文学进行了多声文学综述。我们的发现有四个方面。首先, monorepos 是包含多个相关或不相关的项目的单个存储库, 它们共享相同的依赖关系。其次, 集中化和标准化是一些关键特征。第三, 主要优点包括简化的依赖关系、跨项目更改的协调以及轻松的重构。第四, 代码运行状况、代码库复杂性以及用于开发和执行的工具投资被认为是主要的挑战。少

2018 年 10 月 22 日提交;最初宣布 2018 年 10 月。

8. 在伪装面前捕捉松散的同步行为

作者:潘基文, 孙娇, 刘欣

摘要: 欺诈对社交网络和其他在线应用程序的业务产生了严重的不利影响。用户可以通过在微博上购买 "僵尸追随者" 成为假名人。商家可以通过亚马逊上的虚假评论来提升自己的声誉。这种现象在 **facebook**、yelp 和 tripadvisor 等方面也明显存在。在所有这些情况下, 欺诈者都试图通过伪造他们控制的假账户与目标客户之间的互动来操纵平台的排名机制。少

2018 年 10 月 20 日提交;最初宣布 2018 年 10 月。

9. 协作语义 web 信息体系结构的 web 应用程序

作者:Massimiliano dal mas

摘要: 本文分析了 web 应用程序上的信息可视化原型设计, 用于社区在一个协作环境中, 代表了 **facebook**、instagram、推特等实际社交网络的演变。克林丁, 维吉尼亚人这项工作的目的是确定基于语义 web 的信息可视化的 web 应用程序的最常见功能, 并讨论它们如何在 "协作" 环境中支持用户的要求。ui 上下文感知开发的解决方案基于 "联合含义", 该理解为社区内容的创建者和社区内容的用户的联合解释, 这得益于上下文和接口自适应, 而上下文和界面适应使用的是与语义网。通过对概念的验证

原型, 可以表明所提出的方法方法也可以很容易地应用于使用不同语言和/或组件技术构建的现有演示组件。少

2018 年 10 月 18 日提交;最初宣布 2018 年 10 月。

10. 用户和组织联合身份采用的障碍审查

作者: [john sherlock](#), [manoj muniswamaiah](#), [lauren clarke](#) , [shawn cicoria](#)

摘要: 查看身份即服务 (idaas) 和联合身份管理 (fim) 以及组织、用户和普通用户之间的接受。虽然 fim 在教育、商业和政府组织中表现出了接受, 但一般民众的行动并没有看到作为前者的信任程度。在极端的例子中, 接受的障碍或使能因素可能允许使用您的 facebook 凭据登录银行并处理业务? 少

2018 年 10 月 14 日提交;最初宣布 2018 年 10 月。

11. 预测性测试选择

作者: [Mateusz machalica](#), [alex samylkin](#), [meredith porth](#), [satish chandra](#)

摘要: 基于变化的测试是 facebook 持续集成的关键组成部分。然而, 大量的测试, 加上提交给我们的整体存储库的大量更改率, 使得对每个更改运行所有可能受到影响的测试是不可行的。我们提出了一种新的预测测试选择策略, 它选择一个测试子集来对提

交给连续集成系统的每个更改进行测试。该策略是使用基本的机器学习技术从大量历史测试结果数据集中学习的。该策略部署在生产环境中，将测试代码更改的总体基础架构成本降低了两倍，同时保证 95% 以上的单个测试失败和 99.9% 以上的错误更改仍向开发人员报告。我们在这里介绍的方法也解释了测试结果的非确定性，也被称为测试浮躁。少

2018 年 10 月 11 日提交;最初宣布 2018 年 10 月。

12. 重定向 2 自身: 通过异地间接访问保护用户上传内容的知识产权

作者:[georgeos kon](#) 中移动于 [angelos d.Keromytis](#) , [georgeos Portokalidis](#)

摘要: 社交网络服务吸引了数百万用户，包括个人、专业人士和公司，每天上传大量内容，如文字、图片和视频。内容创建者保留他们与这些网络共享的内容上的知识产权 (ip) 权利，但是，他们通常会含蓄地授予他们，有时使用该内容的许可过于宽泛，从而使服务能够在可能的情况下使用这些内容。不可取的方式。例如，**facebook** 声称所有用户提供的内容都有可转让的、可分许可的、免版税的全球许可。专业内容创作者和摄影师一样，尤其受到影响。在本文中，我们提出了一种在不影响用户功能损失的情况下将用户数据与社交网络服务脱钩的设计。我们的设计表明，在能够在服务条款和整体环境（例如，不同的位置）下托管用户生成的内容的第三方，用户数据被排除在社交网络服务之外，以

更好地满足用户的需求和愿望.同时, 间接架构无缝集成到社交网络服务中, 无需服务器端的任何合作, 因此用户可以像在现场托管时一样透明地访问异地数据。我们已经实现了我们的设计, 作为 chrome 浏览器的扩展, 称为 redirect2own, 并表明它在访问 "重定向" 内容时产生的开销可以忽略不计。我们提供的扩展作为自由软件和它的代码作为一个开源项目。少

2018 年 10 月 10 日提交;最初宣布 2018 年 10 月。

13. ai 辩论中的 30 年周期

作者:[jean-marie chauvet](#)

摘要: 在过去的几年里, 人工智能的兴起和这一领域学术突破的成功是不可避免的。向 ai 初创企业投入了大量资金。许多现有的科技公司——包括谷歌、亚马逊、脸谱和微软等巨头——已经开设了新的研究实验室。这些日常工作和娱乐工具的迅速变化促使人们对基础技术本身的兴趣不断上升;记者不知疲倦地写 ai, 公司——无论科技性质或非——品牌自己与人工智能, 机器学习或深度学习, 只要他们有机会。面对媒体的这一报道, 几位分析师开始对过度解读 ai 的火爆成功和公众对这一话题的报道有时不佳表示担忧。本文简要回顾了人工智能和机器学习的记录, 发现了早期戏剧性的成功模式, 其次是哲学批判和意想不到的困难, 如果不是彻头彻尾的停滞, 30 年后几乎回到了时钟自 1958 年以来的周期。少

2018 年 10 月 8 日提交;最初宣布 2018 年 10 月。

14. 文本中的个性方面识别

作者:w 斯 [leley ramos dos santos](#), [ivandre paraboni](#)

摘要: 基本的五大个性特征及其方面被认为是相关的广泛的语言特征, 因此, 从文本中识别个性特征是一个既定的 nlp 任务。然而, 获取方面信息可能需要大量的个性清单, 因此, 现有的计算模型通常仅限于承认五个主要个性类别。基于这些观察, 本文调查了从较短的个性清单中获得的(低成本)巴西 **facebook** 语料库对若干个性方面的认可。在此过程中, 我们比较了一些个性方面的识别模型, 并为该领域的进一步研究提供了初步参考结果。

少

2018 年 10 月 6 日提交;最初宣布 2018 年 10 月。

15. 加密压缩系统社交网络服务的图像操作规范

作 者 :[tatsuya chuman](#), [kenta iida](#), [warit siichhotedumrong](#), [hitoshi kiya](#)

摘要: 当时的加密压缩(etc)系统已被建议通过不受信任的通道提供程序安全地传输图像。在本研究中, etc 系统被应用于推特等进行图像操纵的社交媒体。对 etc 系统中使用的基于块竞争的加密方案进行了评价, 认为这些方案对社交媒体上的图像处理具有

鲁棒性。其目的是调查 5 个社交网络服务提供商 (**sns**)、**facebook**、推特、google +、tumblr 和 flickr 是如何操纵图像的, 并确定上传到 sns 提供商的加密图像是否可以避免被这些信息传输而扭曲操作。在一项实验中, 加密和非加密的 jpeg 图像被上传到各种 sns 提供商。结果表明, etc 系统适用于五个 sns 提供者。少

2018 年 10 月 4 日提交;最初宣布 2018 年 10 月。

16. "参议员, 我们销售广告": 2016 年俄罗斯脸谱广告活动分析

作者:[ritam dutt](#), [ashok deb](#), [emilio ferrara](#)

摘要: 美国民主的关键方面之一是自由和公正的选举, 使一任总统能够和平移交权力。2016 年美国总统选举之所以突出, 是因为在选举之前、期间和之后都有外国的影响力。这种怀疑影响的很大一部分是通过社交媒体进行的。在本文中, 我们特别研究了据称俄罗斯政府购买的 3, 500 个 **facebook** 广告。这些广告由美国国会众议院情报委员会于 2018 年 5 月 10 日发布。我们使用自然语言处理技术对广告进行了分析, 以确定与最有效的广告相关的文本和语义特征。随着时间的推移, 我们将广告集中到各种广告系列和与之相关的标记方中。我们还在个人、活动和政党的基础上研究了广告的有效性。最有效的广告往往有不太积极的情绪, 关注过去的事件, 在本质上更具体、更个性化。更有效的运动也

表现出如此相似的特点。竞选的持续时间和广告的推广表明，人们希望播下分裂的阵营，而不是左右选举。少

2018 年 9 月 26 日提交;最初宣布 2018 年 9 月。

17. 如何成为著名的 instagram: 双重关注的流行后预测

作者:张忠平,陈天朗, 郑州,李嘉新,罗洁波

摘要: 随着社交应用的数量不断增加，人们越来越愿意在脸谱、instagram 和微信等社交媒体平台上分享日常照片和活动。在社交媒体数据挖掘中，后人气预测受到了数据科学家和心理学家的高度关注。现有的研究更侧重于探索用户群体的后受欢迎程度，包括时间信息、用户连接、评论数量等综合因素。但是，这些框架不适合指导特定用户发布热门帖子，因为此用户的属性是固定的。因此，以前的框架只能回答“帖子是否受欢迎”的问题，而不是“如何通过流行帖子成名”的问题。在本文中，我们的目标是预测特定用户的帖子的受欢迎程度，并挖掘受欢迎程度背后的模式。为此，我们首先从 instagram 收集数据。然后，我们设计一个方法来计算用户环境，表示特定用户非常可能发布的内容。在相关数据的基础上，设计了一种新的双关注模型，将图像、标题和用户环境结合起来。双关注模型主要由图像标题对的显性注意和用户环境的隐性关注两部分组成。设计了一个分层结构，将显性注意部分和隐性注意部分结合起来。我们进行了一系列的实验，以验

证我们的模型的有效性, 并研究可能影响流行的因素。分类结果表明, 我们的模型比基线强, 统计分析确定了什么样的图片或标题可以帮助用户获得相对较高的 "喜欢" 数字。少

2018 年 9 月 25 日提交;最初宣布 2018 年 9 月。

18. 脸谱公共页面中的多视图社区检测

作者:zhige xin, george barnett, jon w. chapman, s. felix wu

摘要: 社会网络中的社区检测由于其在揭示人们如何联系和互动方面的重要性而得到了广泛的研究。然而, 脸谱公共网页对社区结构的关注很少。在本研究中, 我们调查了 facebook 新闻组页面中的社区检测问题。特别是为了处理用户活动的多样性, 我们采用多视图聚类来集成不同的视图, 例如, 喜欢帖子和喜欢评论。在本研究中, 我们不仅在一个给定的单页上, 而且在多个页面中探索社区结构。结果表明, 该方法能有效减少分离, 提高社区结构质量。少

2018 年 10 月 5 日提交;v1 于 2018 年 9 月 23 日提交;最初宣布 2018 年 9 月。

19. pp-dblp: 基于 dblp 的属性公钥专用网络建模与生成

作者:黄新,姜嘉新,崔拜龙, 徐建良,张志伟,宋云雅

摘要: 在许多在线社交网络（如 **facebook**、google +、twitter 和 instagam）中，用户更愿意隐藏他的部分或全部关系，这使得公共用户甚至朋友都看不到这种私人关系。这就产生了一个新的图形模型，称为公私网络，每个用户都有自己的网络视角，包括私人连接。近年来，公网分析引起了人们对文献的浓厚兴趣。研究了大量重要的图形计算问题（如最短路径、中心点、页面排名和可达性树）。但是，由于数据源和隐私问题有限，建议的方法不会在真实世界数据集上进行测试，而是通过随机选择顶点作为私有顶点在合成数据集上进行测试。因此，在评估效率和效力时，公私网络的实际数据集对于此类算法至关重要，也是迫切需要的。在本文中，我们从现实世界的 dblp 记录中生成了四个公私网络，称为 ppdblp。我们将发表的文章视为公共信息，并将正在进行的合作视为隐藏的信息，这只有作者才知道。我们发布的 ppdblp 数据集为以公平的方式验证各种高效的公私分析算法提供了前景。此外，在广泛存在的归因图的推动下，我们提出了一个先进的属性式公用图模型，其中顶点不仅具有私有边缘，而且具有私有属性。我们还讨论了属性公式图形上的未决问题。在我们生成的真实世界数据集中的初步实验结果验证了公私模型和最先进算法的有效性和效率。少

2018 年 9 月 20 日提交;最初宣布 2018 年 9 月。

20. 对主动攻击下社交网络隐私措施的分析与评价

作者:[bhaskar dasgupta](#), [nasim mobasheri](#) , [ismael g. yero](#)

摘要: 不幸的是, 在现代互联网时代, facebook、推特和 linkedin 等复杂的相互关联的社交网络的广泛使用也为恶意实体侵犯此类网络用户的隐私打开了大门. 在本文中, 我们从理论上和经验上调查了最近在 (信息科学, 328, 403–417, 2016) 中引入的主动攻击下的大型网络的侵犯隐私措施。我们的理论结果表明, 如果网络的拓扑结构不包含循环, 负责防止隐私侵犯的网络管理员在设计网络时必须非常小心。我们的实证结果揭示了八个真正的社交网络以及经典的 erdos–renyi 模型和巴拉巴西–阿尔伯特生成的无标度随机网络所产生的大量合成网络的侵犯隐私行为特性首选附件模型。少

2018 年 9 月 16 日提交;最初宣布 2018 年 9 月。

21. 社交媒体信息传播的趋势

作者:[hunt allcott](#), [matthew gentzkow](#), [chuan yu](#)

摘要: 我们衡量 2015 年 1 月至 2018 年 7 月期间在**脸书**和**推特**上传播错误信息的趋势。我们关注的是 570 网站的故事, 这些网站被确定为虚假故事的制作者。截至 2016 年底,**脸谱**和**推特**上与这些网站的互动稳步上升。随后, 脸谱网上的互动大幅下降, 而**微博**上的互动比例继续上升, 脸谱与推特股票的比例下降了约 60%。我们看不到其他新闻、商业或文化网站的类似模式, 随着时间的推

移, 这些网站的互动相对稳定, 在选举前后, 两个平台上都遵循了类似的趋势。少

2018 年 9 月 16 日提交;最初宣布 2018 年 9 月。

22. 检查互动: 社交媒体中的在线新闻媒体协同作用

作者 :praboda rajapaksha, reza farahbakhsh, noel crespi, bruno defude

摘要: 社交媒体的日益普及从根本上改变了新闻内容的传播方式, 包括具有新维度的互动尝试。迄今为止, 报纸、电视、广播等传统新闻媒体已经利用社交媒体、博客、网站等, 将其活动适应了网络新闻媒体。本文对全球大众新闻媒体的社交媒体存在提供了一些见解。尽管这些大型新闻媒体在很大程度上通过社交媒体环境传播内容, 但对社交媒体社区新闻项目制作人、提供者和消费者的了解甚少。为了更好地了解这些互动, 这项工作旨在分析推特和脸书这两个大型社交媒体的新闻项目。为此, 我们从 48 家新闻媒体收集了在推特和脸书上发表的所有帖子, 利用 152k 推特和 80k 脸书帖子的数据集进行描述性和预测性分析。我们探索了一套新闻媒体, 这些媒体自己在社交媒体上发源内容, 那些将新闻项目分发给其他新闻媒体的新闻媒体, 以及那些消费其他新闻媒体新闻内容和分享副本的新闻媒体。我们根据新闻媒体社区内的帖子数量、追随者数量和互动数量, 提出了一个预测模型, 以提高新闻媒体在读者中的受欢迎程度。结果表明, 新闻媒体应分散自

己的内容，首先在社交媒体上发布，成为大众新闻媒体，从新闻读者那里获得更多对新闻项目的吸引力。少

2018 年 9 月 16 日提交;最初宣布 2018 年 9 月。

23. 网络网络时间表中的公平: 度量、模型与机制设计

作者 :eduardo hargreaves, claudio agosti, daniel menasché, giovanni negia, 亚历山大 reiffers-maton, eitan altman

摘要: 脸书新闻源个性化算法每天都会对数百万网民的生活方式、情绪和观点产生重大影响。尽管如此，这种算法的行为缺乏透明度，激励测量，建模和分析，以了解和改善其性能。在本文中，我们提出了一个可重复的方法，包括测量，分析模型和基于公平的新闻饲料设计。该模型利用了实时 (ttl) 计数器的多功能性和分析可追踪性，以捕捉发布商在动态消息上的可见性和占用率。测量用于参数化和验证所提出模型的表达能力。然后，我们进行假设分析，以评估用户根据从模型中获得的基线产生的可见性和占用偏差。我们的研究表明，存在着明显的偏差，在新闻饲料的顶端位置上更加突出。此外，我们发现，即使是那些被故意设定为对其政治观点中立的用户，这种偏见也是不可忽视的，这促使人们提出了一种新颖和更透明的基于公平的新闻饲料设计。少

2018 年 10 月 30 日提交;v1 于 2018 年 9 月 11 日提交;最初宣布 2018 年 9 月。

24. 改进的基本块重新排序

作者:andy newell, sergey pujrev

摘要: 基本块重新排序是配置文件引导二进制优化的重要步骤。最先进的的基本块重新排序是最大限度地增加通过分支的数量。但是,我们证明,此类排序可能会对指令和 i-tlb 缓存施加次优性能。我们提出了一种新的算法,该算法依赖于一个结合了落差和缓存行为影响的模型。由于现代处理器缓存的细节相当复杂,而且往往是未知的,我们展示了如何使用机器学习来选择参数,以最好地消除不同的缓存效果,从而最大限度地提高二进制性能。对各种应用程序(包括 **facebook** 生产工作负载、开源编译器 clang 和 spec cpu 2006 基准)进行了广泛的评估,这表明新方法优于现有的块重新排序技术,从而改进了由此产生的大规模数据中心应用程序的性能。我们有开放来源的代码的新算法作为后链路二进制优化工具, bolt 的一部分。少

2018 年 9 月 12 日提交;最初宣布 2018 年 9 月。

25. 利用上下文完整性注释分析隐私策略

作者 :yan shvartzshnaider, noah apthorpe, nick feamster, helen nissenbaum

摘要: 在本文中, 我们展示了使用上下文完整性 (ci) 理论对隐私政策陈述进行注释和评价的有效性。我们使用 ci 注释进行案例研究, 以比较剑桥分析丑闻前后 **facebook** 的隐私政策。更新后的**脸书**隐私政策提供了更多细节, 说明正在传输哪些信息、由谁传输、由谁传递、向谁传输、在什么条件下传输。但是, 一些隐私声明通过在单个声明中包含许多 ci 参数, 规定了大量不可理解的信息流。由于使用了模糊的术语或完全省略了上下文参数, 其他语句导致信息流不完整。然后, 我们证明众包可以有效地生成规模的 ci 隐私政策注释。我们在 17 家公司的 48 项隐私政策摘录中测试了 ci 注释任务, 共有 141 名人群工人。由此产生的高精度注释表明, 众包可用于生成大量附加注释的隐私政策, 供今后研究使用。少

2018 年 9 月 6 日提交;最初宣布 2018 年 9 月。

26. 在 facebook 上捕获 Altmetrics 值参与度的挑战

作者:[asura enkhbayar](#), [juan pablo alperin](#)

摘要: 此前的研究表明, 尽管脸谱很受欢迎,但分享学术内容的频率较低。为了研究这种差异, 我们通过使用多个 url 查询 facebook api 来探索来自科学网络的 **103**, 539 篇文章的随机集合, 从而探索参与人数。我们确定了两个主要的挑战领域: 将文章映射到 url, 将 url 映射到 **facebook** 内部的对象。然后, 我们探

讨了数据集中的三个问题案例: (1) 标识任何给定 url 的着陆页, (2) 将等效 url 映射到不同 facebook 对象的实例, 以及 (3) 将不同文章映射到同一 url 上的实例脸谱对象。我们发现, 在脸谱网上至少分享过一次的所有文章中, 11.8 的参与数字由于这些问题而不可靠。此外, 我们无法确定数据中 11.6 的文章的 url。综合来看, 这三个问题案例占 103 539 篇测试文章的 12.3, 这些文章的参与编号是不能依赖的。鉴于我们只测试了少量问题案例和 url 变体, 我们的结果指出了那些希望通过可用 api 以编程方式收集 facebook 指标的人所面临的巨大挑战。少

2018 年 9 月 4 日提交;最初宣布 2018 年 9 月。

27. 论微目标社会分型广告--以俄罗斯链接的脸谱广告活动为例

作者 :[flife n. ribeiro](#) , [koustuv saha](#), [mahmoud reza babaei](#), [lucas henrique](#), [johnnatan méméas](#), [oana goga](#), [fabricio benevenuto](#), [krishna p. gummadi](#), [elissa m. 雷德迈尔斯](#)

摘要: 有针对性的广告旨在提高广告客户与客户匹配的效率。然而, 有针对性的广告也可能被恶意广告商滥用, 以有效地接触到容易受到虚假故事的人, 煽动不满, 并煽动社会冲突。由于非目标和非弱势人群不会看到有针对性的广告, 因此恶意广告可能不会被举报, 其效果也不会被发现。这项工作审查了一个具体的恶意广告案例, 探讨了 2016 年美国大选前俄罗斯情报研究机构 (ira) 的政

治广告在多大程度上利用了**脸谱**有针对性的广告基础设施有效地针对弱势亚群的分裂或两极分化的话题（如移民、基于种族的警务）的广告。特别是，我们做了以下工作：(a) 我们进行美国人口普查代表性调查，以描述具有不同政治意识形态的用户如何报告、批准和感知 ira 广告内容的真实性。我们的调查显示，许多广告是 "分裂"：他们引起非常不同的反应，从属于不同社会突出群体的人。(b) 我们描述了这些有分歧的广告是如何针对对现状特别感到不满的亚人口的。我们的调查结果支持了现有的提高内容透明度和政治广告针对性的呼吁。(c) 我们特别关注 **facebook** 广告 api 如何促进此类定位。我们展示了 **facebook** 如何汇总大量有关用户的个人数据，并将其提供给广告商，从而实现了此类恶意定位。少

2018 年 8 月 31 日提交;v1 于 2018 年 8 月 28 日提交;**最初宣布** 2018 年 8 月。

28. 微信：视觉分析，以更好地预测微信文章的信息级联预测

作者:李全,吴子明,李玲玲, 克里斯坦托肖恩 n,曲华敏, 马晓娟

摘要: **脸谱**和微信等社交媒体使数百万用户能够以前所未有的规模创建、消费和传播在线信息。由于关注的零和性质，社交媒体上丰富的信息加剧了微信公共文章（即帖子）的竞争，以吸引用户的关注。因此，只有一小部分信息往往变得极其受欢迎，而其余

的信息却没有被注意到或迅速消失。这种典型的 "长尾" 现象在社交媒体中非常普遍。因此, 近年来, 人们越来越有兴趣预测社交媒体帖子的受欢迎程度的未来趋势, 并了解影响帖子受欢迎程度的因素。然而, 现有的预测模型要么依赖于繁琐的功能工程, 要么依赖于复杂的参数调整, 而这些都很难理解和改进。本文通过结合视觉推理支持用户之间的通信, 并对基于点过程的模型进行了研究和增强, 以获得更好的预测结果。该系统支持用户发现模型背后的工作机制, 并根据获得的见解相应地提高预测精度。我们使用现实的微信文章来演示系统的有效性, 并在大规模的微信文章中验证改进后的模型。我们还征求和总结了微信领域专家的反馈意见。少

2018 年 8 月 27 日提交;最初宣布 2018 年 8 月。

29. "我应该担心吗?" 帐户安全事件响应的跨文化考察

作者:[elissa m. redmiles](#)

摘要: 数字安全技术能够识别和防止对用户帐户的许多威胁。然而, 一些威胁仍然是, 为了提供可靠的安全, 需要人为干预: 例如, 通过用户注意警告信息或完成二级认证程序。虽然之前的工作已经广泛地探索了人们对数字安全威胁的心理模型, 但我们对用户对无线威胁的精确、即时响应过程了解甚少。在这项工作中, 我们对最近在其真实 **facebook** 帐户上遇到可疑登录事件的用户进行

了一系列定性访谈 (n 增幅 67), 以探索这一帐户安全事件响应过程。我们发现, 来自五个国家的参与者——在线和离线文化不同——之间有一个共同的进程, 使我们能够确定未来技术发展的领域, 以最佳方式支持用户安全。我们提供了更多关于事件响应信息查找的独特性质、已知攻击者威胁模型以及从大型跨文化数字安全定性研究中吸取的经验教训。少

2018 年 9 月 4 日提交;v1 于 2018 年 8 月 24 日提交;最初宣布 2018 年 8 月。

30. 俄语脸谱的情绪指数

作者:[亚历山大·潘琴科](#)

摘要: 情绪指数衡量语料库中的平均情绪水平。我们引入了四个这样的指标, 并利用它们来衡量一个人在一定时期内根据社交网络中的帖子的平均 "积极程度"。本文首次提出了一个文本, 而不是基于数据的情绪指数。此外, 本研究还首次对俄语**脸谱**的情绪指数进行了大规模研究。我们的结果与之前的英语实验是一致的。少

2018 年 8 月 23 日提交;最初宣布 2018 年 8 月。

31. 光明之海: 在黑暗中实用的设备到设备的安全引导

作者:[flor alvarez](#), [max Kolhagen](#) , [matthias hollick](#)

摘要: 当今信息和通信系统中引导安全的实用解决方案在很大程度上取决于用于身份验证的集中式服务以及密钥和信任管理。对于移动用户来说尤其如此。截至 2018 年初, 谷歌或脸书等身份提供商的活跃用户基础各为 20 亿, 移动运营商的用户数量超过 50 亿独特用户。如果这些集中式服务由于自然或人为制造的灾难、大规模停电或全国范围的审查而完全 "黑暗", 用户将无法在其移动设备上实现引导安全。例如, 现有的分布式解决方案, 即所谓的信任网, 其轻量级程度不够高。此外, 它们既不支持移动设备上的交叉应用, 也不支持使用硬件安全模块对关键材料进行强大的保护。我们提出了一种实用的轻量级方案, 用于无线引导设备到设备的安全性, 从而实现安全的分布式自组织网络。它是为 "在黑暗中" 操作而量身定制的, 为关键材料提供了强大的保护, 同时也提供了直观的手段来构建轻量级的信任网络。sol 特别适合于地方或城市作业, 如协调应急反应, 这有助于限制错误信息的传播。作为概念的证明, 我们在 android 平台上实施了 sol, 从而在真正的移动设备上测试了其可行性。我们通过模拟进一步评估其关键性能方面。少

2018 年 8 月 14 日提交;最初宣布 2018 年 8 月。

32. 不应憎恨: 打击网络仇恨言论

作者: [binny mathew](#), [hardik tharad](#), [subham Prajwal](#), [prajwal singhania](#), [suan kalyan maity](#), [pawan goyal](#), [animesh mukherje](#)

摘要: 社交媒体中的仇恨内容越来越多。虽然脸谱、推特、谷歌试图采取几个步骤来解决这一仇恨内容,但他们最经常冒着侵犯言论自由的风险。另一方面,反言论为在不丧失言论自由的情况下解决网络仇恨问题提供了有效途径。因此,这些平台的另一种策略可以是促进反言论,作为对仇恨内容的防御。然而,要想成功地推广到这样的反言论,就必须对其在网络世界中的动态有一个深刻的了解。缺乏精心策划的数据在很大程度上阻碍了这种理解。在本文中,我们使用 youtube 的评论创建并发布了第一个用于反演讲的数据集。数据包含 9438 手动注释,其中标签指示注释是否为反演讲。这些数据使我们能够首次对反言论的语言结构进行严格的测量研究。这一分析得出了各种有趣的见解,如:反言论评论收到的不反言论评论得到的喜欢是双倍的,对于某些社区来说,大多数的反言论评论往往是仇恨言论,不同类型的反言论并不都是同样有效的,发布反言论的用户语言选择与详细的心理语言学分析所揭示的发布反言论的用户有很大的不同。最后,我们构建了一组机器学习模型,这些模型能够自动检测到 youtube 视频中的反言论, f1-分数为 0.73。少

2018 年 8 月 13 日提交;最初宣布 2018 年 8 月。

33. 讨论进展如何: 社交媒体对话中的话语行为分类

作者: [subhabrata dutta](#) , [tanmoy chakraborty](#) , [Dipankar das](#)

摘要: 提出了一种新的基于关注的分层 lstm 模型, 对社交媒体对话中的话语行为序列进行分类, 旨在利用句子层面以外的文本意义从在线讨论中挖掘数据。这项任务的独特性在于对非正式文本讨论中可能的语用作用进行完整分类, 而不是提取问题答案、姿态检测或讽刺识别, 而这些都是非常重要的具体任务。早期尝试了 reddit 讨论数据集。我们在相同的数据上训练我们的模型, 并在两个不同的数据集上显示测试结果, 一个来自 reddit, 一个来自 facebook。我们提出的模型在领域独立性方面优于前一个模型;在不使用平台相关结构特征的情况下, 我们具有单词相关性关注机制的分层 lstm 分别实现了 f1=28% 和 66% 的 f1 分数, 以预测雷德特和 **facebook** 讨论中评论的话语作用。提出并分析了经常性和卷积体系结构在同一任务中学习话语表示的有效性, 并采用了不同的词和注释嵌入方案。我们的关注机制使我们能够根据文本片段在话语中的角色来探讨它们的相关性排序。我们提出了一个人类注释器实验, 以揭示有关建模和数据注释的重要观察。利用我们基于文本的语篇识别模型, 探讨了异构非文本特征(如位置、时间、信息倾向等)在 **facebook** 上的在线讨论中的作用。少

2018 年 8 月 7 日提交;最初宣布 2018 年 8 月。

34. 你吃药了吗? –从微博上检测个人入药量

作者:debanjan mahata, jpper friedrichs, rajiv rahn shah, jing jiang

摘要: 挖掘推特、文章和脸谱网帖等社交媒体信息, 以获取健康和药物相关信息, 对药物警戒研究产生了浓厚的兴趣。社交媒体网站(如推特)被用于监测药物滥用、吸毒的不良反应和分析与毒品有关的情绪表达。这些研究大多是根据大量人口的汇总结果而不是特定的个人群体进行的。为了在个人层面或特定群体进行研究, 确定提及用户服用药物的职位是必要的。为了实现这一目标, 我们开发了一个分类器, 用于识别在推特上提到的个人药物摄入量。我们在一个带注释的数据集上训练了一个浅卷积神经网络(cnn)模型的堆叠组合。我们使用随机搜索来调整 cnn 模型的超参数, 并为预测任务提供了一系列最佳模型。我们的系统产生最先进的结果, 平均 f-分数为 0.693。我们认为, 所开发的分类器可直接用于心理学、健康信息学、药物警戒和情感计算等领域, 用于跟踪社交媒体中表示医学的患者的情绪、情绪和情绪。少

2018 年 8 月 2 日提交;最初宣布 2018 年 8 月。

35. 一种用于情绪、情感和强度预测的多任务组合框架

作者:md shad akhtar, deepanway ghosal, asif ekbal, pushpak bhattacharyya , sadao kurohashi

摘要: 本文通过多任务整合框架, 探讨了情绪和情绪分析的三个问题, 即 "情绪分类与强度"、"价值"、"情感的觉醒与支配" 和 "

情绪的价值与觉醒"。潜在的问题包括两个粒度（即粗粒度和细粒度）和各种领域（即推特、**脸书**帖子、新闻标题、博客、信件等）。该组合模型旨在利用三个深度学习模型（即 cnn、lstm 和 gru）的学习表示和手工制作的特征表示来进行预测。在基准数据集上的实验结果表明了我们提出的多任务集成框架的有效性。对于大多数问题和领域，我们平均比单任务系统获得 2–3 个点的性能改进。少

2018 年 10 月 15 日提交;v1 于 2018 年 8 月 3 日提交;最初宣布 2018 年 8 月。

36. poprank: 在 facebook 上对网页的影响和用户的参与程度进行排名

作者 :[andrea zacaria](#), [micra del vicario](#) , [walter quattrociochi](#), [antonio sca](#), [luciano pietronero](#)

摘要: 用户在网上往往获得符合其信仰体系的信息，而忽视不同的信息。这样的动态可能会影响页面的受欢迎程度。在本文中，我们介绍了一种算法，我们称之为 poprank，以评估 **facebook** 页面的影响，以及用户的参与的基础上，他们的相互作用。poprank 背后的想法是，i) 高影响力页面吸引了许多低参与度的用户，这意味着他们收到了很少发表评论的用户的评论，而 ii) 高参与度的用户与高影响力页面进行交互，即他们大多是评论页面具有很高的人气。由此产生的页面排名可以预测页面将收到的评论数及

其帖子的数量。页面影响被证明略微依赖于页面的信息内容（例如，科学与阴谋），但独立于用户的两极分化。少

2018 年 7 月 31 日提交;最初宣布 2018 年 8 月。

37.ritual-uh 在 trac 2018 共享任务: 侵略识别

作者: [niloofar safi samghabadi](#), [deepthi mave](#), [sudipta kar](#), [thamar solorio](#)

摘要: 本文介绍了我们的 "trac 2018 侵略识别共享任务" 系统。我们为英语数据集提供的最佳系统使用词汇和语义功能的组合。然而，对于只使用词汇功能的印地语数据，我们得到了最好的结果。我们获得了加权 f1——英语脸谱任务的 0.5921（排名第 12 位）、英语社交媒体任务的 0.5921（排名第 6 位）、印度语脸书任务的 0.5921（排名第一）和印度语社交媒体任务的 0.5921（排名第 12 位）的 0.5921（排名第 2 位）。少

2018 年 7 月 31 日提交;最初宣布 2018 年 7 月。

38. pythia v0.1: 2018 年 v0.1 挑战赛的获胜参赛作品

作者: [yu jiang](#) , [vivek natarajan](#), [xinleichen](#) , [marcus rohrbach](#), [dhruv batra](#), [devi parikh](#)

摘要: 本文介绍了 pythia v0.1, facebook ai research (fair) a-star 团队在 2018 年 v0.1 挑战赛中的获奖作品。我们的出发

点是对自下而上的自上而下（向上）模型进行模块化的重新实现。我们证明，通过对模型体系结构和学习速率计划进行细微但重要的更改、微调图像功能和添加数据增强，可以显著提高 v2.0 v2.0 数据集上向上模型的性能——从 65.75% 到 70.22%。此外，通过使用不同功能和不同数据集训练的各种模型组合，我们能够显著提高 "标准" 的组合方式（即具有不同随机种子的同一模型）1.31。总体而言，我们在 v2.0 v2.0 数据集的测试时实现了 7.27%。我们的完整代码（培训、评估、数据扩充、组合）和预先培训的模型可在以下网站上公开使用：
<https://github.com/facebookresearch/pythia>

2018 年 7 月 27 日提交;v1 于 2018 年 7 月 26 日提交;**最初宣布** 2018 年 7 月。

39. 脸谱新闻饲料中的偏见：意大利选举案例研究

作者 :[eduardo hargreaves](#), [claudio agosti](#), [daniel menasché](#), [giovanni negia](#), 亚历山大 [reiffers-maton](#), [eitan altman](#)

摘要: 脸书新闻源个性化算法每天都会对数百万网民的生活方式、情绪和观点产生重大影响。尽管如此，这类算法的行为通常缺乏透明度，激励测量，建模和分析，以了解和改善其属性。在本文中，我们提出了一种可重现的方法，包括测量和分析模型，以捕捉发布者在动态消息上的可见性。首先，利用测量对所提出的模型进

行参数化和验证, 验证了模型的表现力。然后, 我们进行假设分析, 以评估用户根据从模型派生的基线产生的可见性偏差。我们的研究表明, 存在着明显的偏差, 在新闻饲料的顶端位置上更加突出。此外, 我们发现, 即使是那些被故意设定为中立的用户, 他们的政治观点也是不可忽视的。少

2018年7月27日提交;v1 于 2018年7月22日提交;**最初宣布** 2018年7月。

40. 智能空间中基于对话的复杂事件管理

作者:[andrésousa lago](#), [hugo sereno ferreira](#)

摘要: 智能空间管理可以通过多种方式完成。一方面, 有对话助理, 如 google 助手或 amazon 亚历克莎, 使用户能够舒适地与智能空间交互, 只有他们的声音, 但这些功能有限, 通常仅限于简单的命令。另一方面, 有一些可视化的接口, 如 ibm 的 node-red, 可在不同设备之间启用复杂的功能和依赖关系。然而, 这些都是有限的, 因为它们要求用户对智能设备的工作方式有技术知识, 而且系统的接口更复杂, 更难以使用, 因为他们需要一台计算机。该项目提出了一个新的对话助理-贾维斯-结合了当前助理的易用性和视觉平台的操作复杂性。jarvis 的目标是通过提供直观的命令和有用的功能, 使管理智能空间变得更加容易。jarvis 与现有的用户界面 (如 google 助手、slack 或

facebookmessenger) 集成, 使其能够很容易地与现有系统集成。贾维斯还提供了一个创新的功能——因果关系查询——使用户能够问它为什么会发生一些事情。例如, 用户可以询问 "为什么开灯", 以了解系统是如何工作的。少

2018 年 7 月 18 日提交;最初宣布 2018 年 7 月。

41. bolt: 一种实用的数据中心及其后的二进制优化器

作者:maksim panchenko, rafael auler, bill nell, guilherme ottoni

摘要: 随着计算继续向数据中心发展, 大规模应用程序的性能优化变得更加重要。数据中心应用程序通常非常大和复杂, 这使得代码布局成为提高其性能的重要优化。这促使最近对改进编译时和链接时间代码布局的实用技术进行了研究。尽管后链接优化器在过去取得了一些成功, 但最近的工作没有在现代数据中心应用程序的背景下探讨它们的好处。在本文中, 我们提出了 bolt, 一个基于 llvm 框架的后链接优化器。利用基于样本的分析, bolt 提高了实际应用的性能, 即使是使用反馈驱动优化 (fdo) 和链路时间优化 (lto) 构建的高度优化的二进制文件也是如此。我们证明, 链接后性能的改进是对传统编译器优化的补充, 即使后者是在整个程序级别上完成的, 并且存在配置文件信息。我们在 **facebook** 数据中心工作负载和开源编译器上对 bolt 进行了评估。对于数据中心应用程序, bolt 在配置文件引导函数重新排序和 lto 的基

础上实现了高达 8.0 的性能加速。对于 gcc 和 clang 编译器, 我们的评估显示, bolt 在 fdo 和 lto 之上将其二进制文件的速度提高了 20.4%, 如果二进制文件是在没有 fdo 和 lto 的情况下构建的, 则可加快 52.1。少

2018 年 10 月 12 日提交;v1 于 2018 年 7 月 17 日提交;**最初宣布** 2018 年 7 月。

42. 巴西研究人员撰写的出版物的标准表现: cnpq 生产力奖学金持有者分析

作者:[ronaldo ferreira araujo](#), [marcelo alves](#)

摘要:属于健康与生物科学知识领域。至于消息来源, 门德利的表现很好, 其次是推特, 脸谱 的报道也很低

2018 年 7 月 17 日提交;最初宣布 2018 年 7 月。

43. 注意侵略性检测的 lstm

作者:[n 克里斯·尼希尔](#), [ramit Nikhil](#), [Nikhil kumar nirala](#), [rohan kirnani](#)

文摘: 在本文中, 我们描述了在脸谱网帖中提交的侵略识别共享任务的系统和 nusnik 团队的评论。以前的工作表明, lstm 在自然语言处理任务中取得了显著的性能。我们部署了一个 lstm 模型, 上面有一个关注单元。我们的系统在印地语子任务中分别排

名第六和第 4 位, 用于 **facebook** 评论和通用社交媒体数据的子任务。在相应的英语子任务中排名第 17 位和第十位。少

2018 年 7 月 16 日提交;最初宣布 2018 年 7 月。

44. 计算机专业学生利用 ngt 减少交通堵塞的模拟获取隐性知识

作者:leon andretti abdillah, ifit novita sari, dia eka indriani

摘要: 学科知识管理系统是信息系统学习计划的主干课程之一, 是计算机科学的师资。本课程在第 5 (5) 学期为学生提供 a 级课程。知识是由知识的显性和隐性组成的。为了获取隐性知识, 可以通过涉及多种方法来实现, 即: 1) 头脑风暴, 2) 名义群技术。这项研究还涉及一些社会信息技术, 例如: (1)**脸谱**、(2) wordpress、3) dropbox 和 4) youtube。知识主题的主题是如何减少交通拥堵。在通过 2 (2) 轮仿真回合后, 本研究获得了与车辆停车有关的 8 (8) 项思路, 以减少交通拥堵。利用 ngt 进行的隐性知识捕获模拟能够为所有相关小组成员提供可接受的建议。本研究中使用社会信息技术的情况也得到了小组成员的好评。少

2018 年 7 月 12 日提交;最初宣布 2018 年 7 月。

45. 深度学习中的社会影响预测

作者:邱洁忠,唐健, 马浩,董玉晓东,关山, 唐杰

摘要: 脸谱、推特、微信和微博等社交和信息网络活动已成为我们日常生活中不可或缺的一部分，我们可以很容易地接触到朋友的行为，进而受到朋友的影响。因此，对每个用户进行有效的社会影响预测对于在线推荐和广告等各种应用至关重要。传统的社会影响预测方法通常设计各种手工制作的规则，以提取特定于用户和网络的功能。然而，它们的有效性在很大程度上依赖于领域专家的知识。因此，通常很难将它们推广到不同的领域。在深度神经网络最近在广泛的计算应用中取得成功的启发下，我们设计了一个端到端框架 deepinf，以了解用户预测社会影响的潜在特征表示。一般来说，deepinf 将用户的本地网络作为图形神经网络的输入，以了解她潜在的社会代表性。我们设计的策略是将网络结构和用户特定的特征整合到卷积神经和注意力网络中。在开放学术图表、twitter、微博和 digg 上进行的大量实验表明，拟议的端到端模型 deepinf 明显优于传统功能基于工程的方法，表明代表学习在社会应用中的有效性。少

2018 年 7 月 15 日提交;最初宣布 2018 年 7 月。

46. 差距在哪里？基于代码研究的基础设施系统映射研究

作者: [akund rahman](#), [rezvan mahdavi-hezaveh](#), [laurie williams](#)

摘要: 上下文: 基础结构作为代码 (iac) 是一种自动配置系统依赖关系和设置本地和远程实例的做法。从业者认为 iac 是实施

devops 实践的基本支柱,可帮助他们快速向最终用户交付软件和服务。github、mozilla、**facebook**、**google** 和 netflix 等信息技术组织采用了 iac。对现有 iac 研究进行系统的绘图研究可以帮助研究人员确定与 iac 相关的潜在研究领域,例如, iac 脚本中可能出现的缺陷和安全缺陷领域。目的:通过对与基础设施相关的研究进行系统的摸底研究,帮助研究人员确定与基础设施有关的研究领域作为代码 (iac)。方法:我们通过搜索六个学者数据库来进行研究。我们使用七个搜索字符串收集一组 33,887 出版物。通过系统地应用包容和排除标准,我们确定了 31 份与 iac 相关的出版物。我们通过应用定性分析来确定这些出版物中涉及的主题。结果:我们确定了在与 ias 相关的出版物中研究的四个主题:(i) 框架工具作为代码的基础结构;(ii) 使用基础设施作为编码;(三) 与作为代码的基础设施有关的实证研究;(iv) 在基础结构中作为代码进行测试。根据我们的分析,在所研究的 31 篇出版物中,52% 的出版物提出了一个框架或工具来实现 iac 的实践或扩展现有 iac 工具的功能。结论:由于缺陷和安全缺陷会对 devops 的部署和开发环境以及其他主题产生严重影响,我们观察到需要进行研究,研究 iac 的缺陷和安全缺陷。少

2018 年 7 月 12 日提交;最初宣布 2018 年 7 月。

47. 产品指标是否以与案例研究相当的方式评估社会影响？基于英国卓越研究框架 (ref) 数据的 altmetrics 的收敛有效性的实证检验

作者: [lutz bornmann](#), [robin haunschild](#), [jonathan adams](#)

摘要: Altmetrics 被提出来作为评估研究的社会影响的一种方式。尽管在不同的上下文中, altmetrics 已经被用作影响或关注指标, 但它们是否真正反映或反映了社会影响仍不清楚。这项研究基于英国研究卓越框架 (ref) 的产品指标、引文计数、研究成果和案例研究数据, 以及同行对研究产出和社会影响的 ref 评估。我们使用两个 ref 数据集研究了 altmetrics 的收敛有效性: 作为研究输出 (pro) 提交给 ref 的出版物和案例研究 (pcs) 中引用的出版物。旨在展示社会影响的案例研究应引用最相关的研究论文。我们使用 mhq 的指标来评估影响——这是一个为多个零的计数数据引入的指标。分析的第一部分的结果显示, 新闻媒体以及 **facebook**、博客、维基百科和政策相关文档中的提及对 pcs 的 mhq 值高于 pro。因此, 标准指标似乎对这些数据具有收敛性。在分析的第二部分中, altmetrics 与 ref 评审人员在 pcs 上的平均得分相关。负或接近零相关性质疑在该上下文中的有效性。我们建议, 它们可能反映了社会影响的不同方面 (可以说是未知的关注), 而审查人员 (他们对社会中的研究与行动之间的因果联系感兴趣) 看到了这一点。少

2018 年 10 月 16 日提交;v1 于 2018 年 7 月 11 日提交;最初宣布 2018 年 7 月。

48. 在右倾斜的回声室内: gab 的特征, 一个不受慢的社会系统

作者 :[lucas lima](#), [julio c. s.reis](#), [philipemelo](#), [fabricio murai](#), [leandro araújo](#), [pantelis vikatos](#), [fabrício benevenuto](#)

摘要: 推特和脸书等许多社交媒体系统的内容适度, 促使出现了一个名为 "gab" 的新的促进言论自由的社交网络系统。之后不久, gab 就因违反公司的仇恨言论政策而被从 google play store 中除名, 并因类似原因被苹果拒绝。在本文中, 我们描述了 gab, 旨在了解谁是加入它的用户, 以及他们在这个系统中共享什么样的内容。我们的研究结果表明, gab 是一个非常注重政治的系统, 它承载着被禁止进入其他社交网络的用户, 其中一些网络是由于可能出现仇恨言论和与极端主义有关联的情况。我们首次衡量右倾回声室内的新闻传播情况, 调查一种社交媒体, 在这种媒体中, 读者很少接触到跨越意识形态界限的内容, 而是获得强化其当前潮流的内容政治或社会观点。少

2018 年 7 月 10 日提交;最初宣布 2018 年 7 月。

49. 社交媒体上的事件检测与检索

作者 :[manos schinas](#), [symeon Papadopoulos](#), [yiannis kumpatsiaris](#), [permits mitkas](#)

摘要: 近年来, 我们目睹了推特、**脸书**和 youtube 等社交媒体平台的迅速采用, 以及这些平台被用作全世界数十亿人日常生活的一部分。考虑到人们习惯利用这些平台分享想法、日常活动和经验, 用户生成的内容数量达到了前所未有的水平, 其中很大一部分与现实世界有关, 这并不奇怪事件, 即在特定时间和地点发生的行动或事件。鉴于事件在我们生活中的关键作用, 对周围社交媒体内容进行注释和组织的任务对于确保实时和未来访问有关感兴趣事件的多媒体内容至关重要。在本章中, 我们介绍了近年来的一些研究工作, 这些工作解决了两个主要问题: a) 事件检测和 (b) 基于事件的媒体检索和总结。给定社交媒体项目的存档集合或实时流, 事件检测方法的目的是以描述以前未知的事件集的形式标识这些事件。一般来说, 这些事件可以是任何类型的, 但也有一些针对特定类型事件的方法。给定目标事件, 事件摘要的目标首先是识别相关内容, 然后以简洁的方式表示, 选择最吸引人和最具代表性的内容。少

2018 年 7 月 10 日提交;最初宣布 2018 年 7 月。

50. 利用利他行为对游客 sns 社区的推荐系统

作者:[takumi ichimura](#), [takuya uemoto](#), [shin kamada](#)

文摘: 利用基于智能手机的用户参与传感系统, 已经开发了基于 sns 的观光信息推荐系统。该系统可以通过我们开发的智能手机

应用程序将吸引游客的信息发布到指定的脸谱页面。脸谱网的用户对观光感兴趣,可以从四面八方的信息空间蜂拥而至。但是,sns 上的社区中的活动仅由称为中心的指定人员支持。提出了旅游行为的激活方法,以促进人们的发展。我们开发了由陆军蚂蚁启发的具有利他行为的多智能体系统仿真系统。陆军蚂蚁采取利他行为喂养行为,以压制自私的行为,以多个用户共同使用的共同对象。本文介绍了一些仿真确定的利他主义行为,以振兴 sns 社区。通过一些实验仿真结果,研究了社区振兴过程的有效性。

少

2018 年 7 月 11 日提交;v1 于 2018 年 7 月 10 日提交;最初宣布 2018 年 7 月。

51. 网络社交网络中的配对互动探析

作者:luka humski, damir pintar , mihaela vranić

摘要: 在过去的几十年里,社会学家试图通过分析社交网络来解释人类的行为,这需要获得有关人际关系的数据。在出现在线社交网络之前,这在这一研究领域构成了一大障碍,极大地促进了收集此类数据的进程。如今,通过在 osn 上爬过公共配置文件,可以构建一个社交图,其中 osn 上的 "朋友" 表示为连接的节点。osn 连接不一定表示现实生活中的密切关系,但使用 osn 交互记录可能会揭示现实生活中的关系强度,这一话题激发了最近

的一些研究。尽管如此, 已发表的研究目前缺乏对 osn 交互记录的广泛的探索性分析, 即通过不同的 osn 交互方式全面概述用户的交互。在本文中, 我们利用所进行的广泛社会实验的结果提供了这样的概述, 该实验成功地收集了 3, 200 多名 **facebook** 用户与 1, 40 多万朋友互动的记录。我们的探索性分析侧重于提取 13 个交互参数的种群分布和相关参数, 为未来针对这一研究领域的研究提供有价值的在线社交网络交互见解。少

2018 年 6 月 30 日提交;最初宣布 2018 年 7 月。

52. 基于互动的网络网络用户友谊强度的确定

作者:[sanja krakan](#), [luka humski](#), [zoran skoč ir](#)

摘要: 在线社交网络 (osn) 是现代通信最流行的形式之一, 最著名的是**脸谱**.有关 osn 上用户之间的连接的信息通常非常少。只有在用户连接的情况下, 才知道, 而连接的强度是未知的。所描述的研究的目的是在分析 osn 用户之间的相互作用的基础上, 确定和量化他们之间的友谊强度。我们建立了一个数学模型, 它使用: 监督机器学习算法随机森林, 实验确定的重要性, 通信参数和系数的每个相互作用参数的基础上, 通过研究的答案进行一项调查。在设计友谊强度计算模型时考虑用户意见, 是一种与以往文献研究相反的新方法。以确定所提供的对中的更好的朋友为例, 验证了该模型的准确性。少

2018 年 6 月 29 日提交;最初宣布 2018 年 7 月。

53. wpse:通过浏览器端安全监控强化 web 协议

作者:[stafano calzavara](#), [riccardo focardi](#), [matteo maffei](#), [clara schanidewind](#), [marco Squarcina](#), [mauro tempesta](#)

摘要: 我们提供 wpse, 这是一个用于 web 协议的浏览器端安全监视器, 旨在确保符合预期的协议流, 以及消息的机密性和完整性属性。我们正式证明 wpse 具有足够的表现力, 可以保护 web 应用程序免受各种协议实现错误和 web 攻击的影响。我们讨论了 wpse 在 oauth 2.0 和 saml 2.0 上可以防止的攻击的具体示例, 包括我们通过在 wpse 中形式化协议规范发现的对 saml 2.0 的 google 实现的新攻击。此外, 我们还利用 wpse 在野外对 oauth 2.0 进行了广泛的实验评价。在 90 个经过测试的网站中, 我们发现了 55 个网站的安全缺陷 (61.1), 包括 **facebook** pixel 等跟踪库引入的新的关键漏洞, 所有这些漏洞都可由 wpse 修复。最后, 我们展示了 wpse 在 83 个网站 (92.2) 上的完美工作, 7 个兼容性问题是由自定义实现偏离 oauth 2.0 规范引起的, 其中一个问题引入了一个关键的漏洞。少

2018 年 6 月 24 日提交;最初宣布 2018 年 6 月。

54. 物联网 (iot) 和云计算支持的灾难管理

作者:raj gaire, chigulapalli sriharsha, deepak puthal, hendra wijaya, jongkil kim, prateeksha keshari, Wijaya ranjan, rajkumar buyya, ratan k. ghosh , r. k. shyamasundar , surya nepal

摘要: 灾害管理需要近实时的信息传播, 以便在适当的时间向合适的人提供应急服务。信息和通信技术的最新进展使从各种来源收集实时信息。例如, 部署在字段中的传感器收集有关环境的数据。同样, 推特和脸书等社交网络也可以帮助从灾区民众那里收集数据。一方面, 灾害中的情况认识不足已被确定为人为错误的主要因素之一, 造成生命损失和关键基础设施破坏等严重后果。另一方面, 社交媒体和移动设备的日益普及, 以及物联网的普遍性, 意味着有更多的出站流量来源, 这最终导致了数据泛滥, 从之后不久开始灾害事件的发生, 导致信息海啸问题。此外, 安全和隐私在克服滥用系统侵犯数据或克服滥用信息的行为方面发挥着至关重要的作用。....在本章中, 我们提供了这样一个情况感知应用程序, 以支持灾难管理数据生命周期, 即从数据接收和处理到警报传播。我们利用云计算、物联网和社交计算技术实现可扩展、高效和可用的情况感知应用程序, 称为 cloud4bigdata。少

2018 年 6 月 19 日提交;最初宣布 2018 年 6 月。

55. 基于灰度的社交网络服务块拼接图像加密

作者 :warit siichhotedumrong, tatsuya chuman, shoko Imaizumi , hitoshi kiya

摘要: 本文提出了一种新的块拼接加密方案, 该方案增强了 jpeg 图像加密压缩 (etc) 系统的安全性, 例如, 该系统用于通过不受信任的信道提供程序安全地传输图像。与传统方法相比, 该方法允许使用更小的块大小和更大的块数量。此外, 使用拟议方案加密的图像即使在原始图像有三个颜色通道的情况下, 由于使用灰度, 也会包含较少的颜色信息。这些功能增强了抵御各种攻击的安全性, 如拼图求解器和暴力攻击。一项实验的结果表明, 拟议的 etc 系统方案是在推特和脸书上上传到, 然后从推特和脸书下载的。少

2018 年 6 月 10 日提交;最初宣布 2018 年 6 月。

56. 一种验证社会网络隐私漏洞的逻辑框架

作者:内斯托尔·卡塔尼奥

摘要: 我们提出了一种新的方法来处理在社交网络中出现的传递性许可委派威胁, 当第三方用户授予内容权限从而违反了内容所有者的隐私政策时出现的威胁。在 facebook 这样的社交网络中, 这些类型的隐私泄露往往是无意的, 因此, 需要 (更多) 就地机制, 使社交网络用户认识到更改其隐私政策的后果。我们的方法在使用正式方法工具和技术方面是独一无二的。它基于社交网络的谓词逻辑定义, 该定义可满足现有社交网络的共同方面, 如用户、社交网络内容、友谊、内容权限和内容传输。我们的方法在

yice 中实施。对于谓词逻辑模型, 我们制定了一个用于验证社交网络用户所拥有内容的权限流的安全策略, 并演示了如何验证此安全策略。少

2018 年 6 月 9 日提交;最初宣布 2018 年 6 月。

57. 在 360°视频中进行监督薄弱的显著性预测的立方体填充

作者:郑显子,赵春鸿,董金东,胡凯文,刘廷鲁, 孙敏

摘要: 360°视频中的自动显著性预测对于视点引导应用程序 (例如 **facebook** 360 指南) 至关重要。我们提出了一个时空网络, 它是 (1) 弱监督训练和 (2) 为 360°观看球量身定做的。请注意, 大多数现有方法的可伸缩性较低, 因为它们依赖于注释显著性地图进行培训。最重要的是, 它们将 360°球体转换为 2d 图像 (例如, 单个等量矩形图像或多个独立的普通视场 (nfov) 图像), 从而引入失真和图像边界。相反, 我们提出了一个简单而有效的立方体填充物 (cp) 技术, 如下所示。首先, 我们使用透视投影在立方体的六个面上渲染 360°视图。因此, 它引入了很少的失真。然后, 我们将所有六个面连接在一起, 同时利用多维数据集上的面之间的连接进行图像填充 (即多维数据集填充), 以实现卷积、池、卷积 lstm 图层。这样, cp 在不引入图像边界的同时, 几乎适用于所有的卷积神经网络 (cnn) 结构。为了评估我们的方法, 我们提出了 wild-360, 一个新的 360°视频显著性数据集, 包含具有显著性的

视频与显著性热图注释。在实验中，我们的方法在速度和质量上都优于基线方法。少

2018 年 6 月 4 日提交;最初宣布 2018 年 6 月。

58. 私有和公共 web 存档的聚合框架

作者:matt kelly, michael l. nelson , micmile c. weigle

摘要: 个人和私人网络档案激增, 原因是创建这些档案的工具越来越多, 而且认识到因特网档案和其他公共网络档案无法捕捉到个性化(如脸书)和私人档案(如银行)网页。我们引入了一个框架, 以缓解私有、个人和公共 web 存档中的聚合问题, 同时又不影响私有捕获中包含的潜在敏感信息。我们修改 memento 语法和语义, 以允许历史地图丰富, 以考虑要表示的其他属性, 包括取消引用专用 web 存档捕获的要求。我们提供了一种方法, 让用户进一步参与超越时间的维度中的档案捕获谈判。我们引入了一个用于存档查询优先级和短路的模型, 当需要将私有和个人 web 存档捕获与通过 memento 的公共 web 存档中的存档捕获聚合时。这类谈判对 web 存档来说是新颖的, 可以更无缝地聚合各种类型的 web 存档, 从而更准确地传达过去 web 的情况。少

2018 年 6 月 3 日提交;最初宣布 2018 年 6 月。

59. 协作多模式深度学习, 实现 facebook 市场中的个性化产品检索

作者: [陆正](#), [赵坦](#), [kun han](#), [任茂](#)

摘要: 脸书作为受青睐的客户对客户 (c2c) 产品交易平台, 市场在消费者中的势头迅速增强。它背后的推荐系统有助于显著改善用户体验。建立脸谱市场的推荐系统具有挑战性, 原因有二: 1) 可扩展性: 脸谱市场的产品数量巨大。每天需要为数百万用户打分和推荐数万种产品; 2) 冷启动: c2c 产品的使用寿命非常短, 用户对产品的活动很少。因此, 很难积累足够的产品水平信号来推荐, 我们正面临着一个重大的冷启动问题。本文提出通过构建基于协作多模式深度学习的协同检索系统来解决可扩展性和冷启动问题, 在该系统中, 用户和产品的紧凑型嵌入是由多模态内容进行训练的信息。该系统显示出在线和离线实验的基准有了显著的改进: 在在线实验中, 它将买方向卖家发出的消息数量增加了 + 26.95; 在离线实验中, 预测精度提高了 + 9.58。少

2018 年 5 月 30 日提交; 最初宣布 2018 年 5 月。

60. 用于讽刺检测的英语-印地语代码混合推文语料库

作者: [Akhtar swami](#), [ankush khandelwal](#), [vinay singh](#), [syed sarfaraz akhtar](#), [manish shrivastava](#)

摘要: 像推特和脸谱这样的社交媒体平台已经成为人们用来表达不同话题的最大两种媒体。如此大的用户数据的生成使得情绪分析和意见挖掘等 nlp 任务变得更加重要。在社交媒体上使用讽刺手段最近已成为一种流行的趋势。使用讽刺可以逆转文本所隐含的含义和极性, 这对许多 nlp 任务构成了挑战。文本中的讽刺检测任务对商业和安全服务都越来越重要。我们提出了第一个英语-印地语代码混合数据集的推特标记为存在的讽刺和讽刺, 其中每个令牌也带有语言标签注释。我们提出了一个基线的自限定分类系统, 该系统是使用相同的数据集开发的, 在使用随机森林分类器并执行 10 倍交叉验证后, 平均 f 分数达到了 78.4。少

2018 年 5 月 30 日提交;最初宣布 2018 年 5 月。

61. 历时语感进化的无监督检测

作者: [jean-françois delpech](#)

摘要: 大多数词都有几种意义和内涵, 由于语义的转变而随着时间的推移而演变, 因此, 多年来, 密切相关的词可能会获得不同甚至相反的含义。这种演变与语言和文化变化的研究非常相关, 但目前可用于历时语义分析的工具具有显著的内在局限性, 不适合实时分析。在本文中, 我们演示了随机向量技术的线性如何能够构建一致的单词嵌入 (或语义空间) 的时间序列, 然后可以在不损失任何时间段精度的情况下进行线性比较和组合历时语义转换。

我们展示了这种方法如何产生多义词（如 amazon 或 apple）的时间轨迹，使元素间出现语义漂移和性别偏见，揭示了飓风或总统等稳定概念的不断变化的实例化。这种非常快速、线性的方法可以很容易地分布在许多处理器上，以便在推特或脸书等社交媒体的实时流中遵循；由此产生的、与时间相关的语义空间可以通过简单的加减法随意组合。少

2018 年 5 月 30 日提交;v1 于 2018 年 5 月 29 日提交;最初宣布 2018 年 5 月。

62. 网络上的网络辩论中的社会机器人效应

作者:andree thieltges, or 斯蒂斯·帕帕基里亚科普洛斯, juan carlos medina serrano, simon hegelich

摘要: 2018 年从伊朗的大规模抗议活动开始，让人们想起了所谓的 "阿拉伯之春"，这对马格里布国家、叙利亚和埃及产生了革命性的影响。许多报告和科学考试认为，推特等在线社交网络 (osn 's) 在发表这些抗议活动的幕后黑手方面发挥着关键作用。除此之外，还有证据表明，在社会机器人和假账的帮助下，有针对性地操纵舆论。因此，很明显，有人问，是否有人企图操纵与伊朗在 osn 抗议有关的舆论制造过程，使用社会机器人，以及这些操纵将如何影响整个话语。根据我们展示的与 "伊朗" 主题有关的约 900, 000 条推特的样本，有推特简介，必须被视为社会机器人账户。通过使用文本挖掘方法，我们表明这些社交机器人对辩论中

的负面情绪负有责任。因此, 我们想说明社交机器人对微博上政治讨论的明显影响。少

2018 年 5 月 25 日提交;最初宣布 2018 年 5 月。

63. 质量控制的大众标签时尚评论

作者 :[iurii chemushenko](#), [felix a.gers](#), [亚历山大 löser](#), [alessandro checco](#)

摘要: 我们提出了一种新的方法, 在时尚领域的高品质标签与人群工作者, 而不是专家。我们专注于基于表的情绪分析任务。我们的方法过滤掉了来自人群工人的不准确输入, 但我们保留了不同的工人标签, 以捕捉意见固有的高度变异性。我们以 **facebook** 的 **fasttext** 框架为基准, 展示了标记数据的质量。少

2018 年 4 月 5 日提交;最初宣布 2018 年 5 月。

64. 从脸书状态推断人类特征

作者:[andrew cutler](#), [brian kulis](#)

文摘: 本文探讨了使用语言模型从用户的**脸谱**状态更新中预测 20 个人类特征的方法。这些数据是由 y 大选项目收集的, 包括用户的身份以及他们的个性、性别、政治认同、宗教、种族、对生活的满意度、智商、自我披露、公平和占星术的信仰。一个单一的可解释模型满足了预测性别和个性等经过充分研究的任务的最新

结果;并设定了其他特质的标准,如智商、轰动效应、政治认同和对生活的满意度。此外,还为每个特征发布高度加权的单词。这些列表对于创建有关人类行为的假设以及了解模型正在提取的信息非常宝贵。我们使用性能和提取的功能来分析构建在社交媒体上的模型。我们探讨的现实世界中的问题包括性别分类偏差和剑桥分析的心理模型的使用。少

2018年7月25日提交;v1于2018年5月22日提交;**最初宣布** 2018年5月。

65. 两极分化等级: 基于 facebook 的欧洲新闻消费研究

作者 :ana lucia schmidt, fabiana zollo, antonio sca, walter quattrociochi

摘要: www 的出现改变了我们产生和获取信息的方式。最近的研究表明,用户倾向于选择与其信仰体系一致的信息,围绕忽视不同信息的共同叙述形成由志同道合的人组成的两极分化的群体。在这种环境下,用户合作制定和加强他们的共同叙述,试图揭穿效率低下的东西。这样的配置甚至发生在网上新闻的消费中,考虑到 63% 的用户直接从社交媒体访问新闻,一个假设是,两极分化的加剧允许进一步传播错误信息。沿着这条道路,我们关注不同欧洲国家(意大利、法国、西班牙和德国)脸谱网上新闻媒体周围用户的两极分化。首先,我们比较网页的发布行为 and 用户在不同国家/地区的互动模式,并观察不同的发布、喜欢和评论率。

其次, 我们探讨了用户与不同页面 (即选择性曝光) 互动的趋势, 以及特定页面周围产生的两极分化社区的出现。然后, 我们引入一个新的指标——即两极分化等级——来衡量每个国家的社区两极分化。我们发现, 意大利是最两极分化的国家, 其次是法国、德国和最后的西班牙。最后, 我们提出了一个变种的有界置信度模型, 通过考虑用户对新闻的参与和信任来模拟这些社区的出现。我们的研究表明, 对信息播音员的信任对防止用户在线两极分化起着关键作用。少

2018 年 5 月 21 日提交;最初宣布 2018 年 5 月。

66. 异构图的半继承学习及其在 facebook 新闻源中的应用

作者:郑菊,李嘉明,布拉姆·瓦斯蒂, 郭胜波

摘要: 基于图形的半监督学习是一个基本的机器学习问题, 已经得到了很好的研究。大多数研究集中在同质网络 (例如引文网络、朋友网络) 上。本文针对具有异构节点类型特征的图形, 提出了一种基于图形的半监督深度学习算法——异构嵌入标签传播 (help) 算法。通过经验验证, 我们用 facebook 用户域交互图证明了该方法在域分类任务中的有效性, 并将所提出的 help 算法的性能与最先进的算法进行了比较。我们证明, help 算法提高了多个任务的预测性能, 同时也提高了语义上有意义的嵌入, 这些嵌入对下游分类或回归任务具有鉴别力。少

2018 年 7 月 4 日提交;v1 于 2018 年 5 月 18 日提交;最初宣布 2018 年 5 月。

67. 扩散协议是否控制着级联的增长?

作者 :[justincheng](#), [jon kleinberg](#), [jure leskovec](#), [david liben-nowell](#), [bogdanstate](#), [karthik subbian](#), [lada adamic](#)

摘要: 随着人们相互共享信息, 大型级联可以在在线社交网络中发展。虽然简单的级联已经得到了广泛的研究, 但社交媒体上的各种级联行为却更加多样化。在这里, 我们研究扩散协议, 或社会交换, 使信息传输, 影响级联增长, 类似于通信协议如何定义信息如何从一个点传输到另一个点。研究了 **facebook** 上 98 种最大的信息级联, 我们发现了广泛的扩散协议——从级联重塑图像 (使用一个简单的点击单个按钮进行传播的协议) 到 als 冰桶挑战赛 (其挑战)。扩散协议涉及个人创建和发布视频, 然后提名特定的其他人这样做。我们发现了重复出现的扩散协议类别, 并确定了构建这些协议的两个关键平衡因素, 这对级联的增长有影响: 参与级联所需的努力, 以及保持在琪级的社会成本场边。需要更大的个人努力的协议减缓了级联的传播, 而那些强加更大的不参与的社会成本的协议增加了级联采用的可能性。传输的可预测性也随协议的不同而不同。但无论机构如何, 我们分析中的级联都有类似的复制数量 ("1.8), 这意味着较低的接触率可以用较高的每次接触率抵消。最后, 我们展示了级联的结构如何不仅可以区

分这些协议, 还可以通过分支过程进行建模。总之, 这些发现为了解各种信息级联如何在整个网络中实现实质性采用提供了一个框架。少

2018 年 5 月 18 日提交;最初宣布 2018 年 5 月。

68. 利用社会信号改进矩阵分解项目推荐

作者:王泽,李红

摘要: 尽管在过去的十年里, 推荐系统在产业界和学术界都得到了全面的研究, 但目前的推荐系统大多存在着各种问题: 1) 用户项目矩阵的数据稀疏性严重影响了推荐系统系统质量。因此, 大多数传统的推荐系统方法都无法处理对项目进行评级较少的用户, 这在推荐系统中被称为冷启动问题。2) 传统的推荐系统假定用户是独立的、分布相同的, 而忽略了用户之间的社会关系。然而, 在现实生活场景中, 由于 **facebook** 和 **twitter** 等社交网络服务呈指数级增长, 不同用户之间的社会关系在推荐系统任务中发挥着重要作用。在本工作中, 为了通过整合用户社交网络信息, 提供一个更好的推荐系统, 我们提出了一个具有用户社会连接约束的矩阵分解框架。对实际数据集的实验结果表明, 该方法在 **mae** 和 **rmse** 方面的性能明显优于最先进的方法, 特别是对冷启动用户而言。少

2018 年 5 月 16 日提交;最初宣布 2018 年 5 月。

69. 基于 facebook 反应的情绪分类器作为用于讽刺检测的 cue

作者: [po chen kuo](#), [fernando h. calderon alvarado](#), [yi-shin chen](#)

摘要: 在线社交媒体用户根据上下文对其中的内容做出反应。情绪或情绪在这些反应中发挥着重要作用, 这些反应让这些平台充满了固执己见的内容。为更好地利用这些数据, 正在不断开发不同的方法和应用程序。然而, 由于数据的性质、平台的多样性以及动态在线用户行为, 仍有许多问题需要解决。在发布评论之前, 从用户那里正确获得可靠的情感状态仍然是一个挑战。本工作引入了一种基于 **facebook** 反应和文本数据重叠的半监督多语言情感检测方法。通过由此产生的情绪检测系统, 我们评估了使用情绪和用户行为特征进行讽刺检测任务的可能性。收集和处理了 62 000 多个公共**脸书**网页帖子的 100 多万种英文和中文评论, 进行的实验显示了可接受的业绩指标。少

2018 年 5 月 4 日提交;最初宣布 2018 年 5 月。

70. 基于机器学习的脸谱后指标预测

作者: [emmanuel sam](#), [sergey yarushev](#), [sebastián Basterrech](#),
[亚历克西·阿韦金](#)

文摘: 在这篇简短的论文中, 我们评估了三种著名的机器学习技术在预测 **facebook** 帖子的影响方面的性能。社会媒体对社会行

为有着巨大的影响。因此, 建立一个预测社交媒体岗位影响的自动模型, 对社会是有益的。本文分析了支持向量回归 (svr)、回波状态网络 (esn) 和自适应网络模糊注入系统 (anfis) 三种常用技术预测后影响的效率。评估是在公共和知名的基准数据集上进行的。少

2018 年 5 月 15 日提交;最初宣布 2018 年 5 月。

71. 用户阻止被认为是有害的? 识别社交帐户的攻击性可控侧边通道

作者:watanabe takuya, eitaro shioji , mitsuaki 秋山, keito sasaoka, takeshi yagi, Mitsuaki mori

摘要: 本文介绍了一种实用的侧通道攻击, 用于识别攻击者网站访问者的社交 web 服务帐户。我们的攻击利用了广泛采用的用户阻止机制, 滥用其固有属性, 即某些页面根据用户是否被其他用户阻止而返回不同的 web 内容。我们的主要见解是, 攻击者准备的帐户可以持有攻击者控制的二进制状态, 即相对于同一服务上的任意用户进行阻塞;如果用户登录到服务, 则当用户访问攻击者的网站时, 可以通过常规的跨站点计时攻击将此状态作为位数据检索。我们概括并引用这样的属性作为可见性控制, 我们认为这是我们攻击的基本假设。在这种原始的基础上, 我们表明, 拥有一组受控帐户的攻击者可以对通过侧通道泄漏的数据进行完全和灵活的控制。利用这种机制, 我们表明, 在各种社交 web 服务上设计和实现强大的、大规模的用户识别攻击是可能的。为了验证

我们攻击的可行性, 我们使用 16 种流行的社交 web 服务进行了广泛的实证研究, 并证明其中至少有 12 项容易受到我们的攻击。弱势服务不仅包括推特和脸书等热门社交网站, 还包括提供社交功能的其他类型的网络服务, 如 ebay 和 xbox live。我们还证明, 攻击可以实现近 100% 的准确性, 并可以在足够短的时间内完成在实际设置。我们讨论了攻击的基本原则、实际方面、限制以及可能的防御。少

2018 年 5 月 14 日提交;最初宣布 2018 年 5 月。

72. 统计和语义模型在多文档汇总中的应用

作者:[divyanshu daiya](#), [anukarsh singh](#), [mukesh jadon](#)

摘要: 我们报告了一系列不同语义模型的实验, 在各种统计模型的基础上进行提取文本总结。尽管统计模型可以更好地捕捉文本周围的词的共现和分布, 但它们无法检测到整个句子的上下文和意义。语义模型帮助我们更好地洞察句子的语境。我们展示了如何调整不同模型之间的权重可以帮助我们在不同的基准上取得显著的成果。在给定的语料库上进一步学习语义模型中使用的预训练向量, 可以在性能上给出额外的尖峰。在不同的统计模型之间使用称重技术也进一步完善了我们的结果。对于统计模型, 我们使用了 tf/idf、textrank、jacards 若辛相似点。对于语义模型, 我们使用了基于文字网络的模型, 并提出了基于手套向量和

facebook 的无间发送的两个模型。我们在 duc 2004 数据集上测试了我们的方法, 生成了 100 个单词的摘要。我们讨论了系统、算法、分析, 并提出和测试了可能的改进。与其他总结者相比, 认为 roege 分数是比较的。少

2018 年 5 月 17 日提交;v1 于 2018 年 5 月 11 日提交;**最初宣布** 2018 年 5 月。

73. 数据中心存储系统中位损坏的实时恢复

作者: amy tai, andrew kryczka, shobhit kanaujia, chris petersen, mikhail 安东诺夫, muhammad Jamieson, kyle jamieson, michael j. freedman, asaf cidon

摘要: 由于闪存具有高性能和降低的比特成本, 它正在成为数据中心热数据的主要存储介质。然而, 闪存耐久性是一个永久的问题, 由于技术趋势, 后续几代闪存设备在遇到无法纠正的位错误之前, 其使用寿命逐渐缩短。在本文中, 我们建议通过允许设备暴露更高的误码率来延长闪存寿命。为此, 我们提出了一组新的策略, 它利用分布式存储系统中潜在的冗余从位损坏错误中恢复, 并将性能和恢复开销降至最低。在此过程中, 通过有效利用这些设备, 即使在这些设备开始暴露位错误后, 直放也能显著延长闪存设备的使用寿命。我们在两个实际存储系统上实现了毫不约束: zippydb (由 rocksdb 支持的分布式键值存储) 和 hdfs (分布式文件系统)。在 **facebook** 上的生产跟踪中进行测试时, 走低直流

将 zippydb 中的应用程序可见错误率降低了 10^2 以上, 恢复时间减少了 10^4 以上。直送还允许 hdfs 容忍 10^4 – 10^5 更高的误码率, 而不会遇到应用程序可见的错误。少

2018 年 5 月 8 日提交;v1 于 2018 年 5 月 7 日提交;**最初宣布** 2018 年 5 月。

74. #ILookLikeAnEngineer: 利用基于社交媒体的 hashtag 行动主义运动作为更好地了解工程多样性问题的镜头

作者:[aakdas malik](#), [aditya johri](#), [rajathanda](#), [habib karbasian](#), [hemant purohit](#)

摘要: 每年都投入大量时间和资源, 以改善一系列联邦和州机构、私营/非营利组织和基金会工程的多样性。尽管进行了几十年的投资, 但努力没有产生预期的回报——在 stem 劳动力需求不断增加的情况下, 少数群体的参与仍然滞后。近年来, 出现了新的数据流——包括推特、脸书和 instagam 在内的在线社交网络——作为公众社会行为和态度的关键传感器。近 87% 的美国人现在参与某种形式的社交媒体活动。因此, 社交网站已成为社会行动的有力指标, 社交媒体数据显示出研究公共卫生宣传、政治运动、人道主义危机和行动主义等许多问题的重大希望。我们认为, 同样可以利用社交媒体数据来更好地理解 and 改善工程多样性。作为说明该方法可行性的案例研究, 我们提出了一项运动的结果, #ILookLikeAnEngineer (使用推特数据——19 354 条原始推特和

29 529 条转发), 目的是增加工程工作场所的性别多样性。该运动为增加多样性和与相关受众联系的新颖方式的总体努力提供了持续的动力。我们的分析表明, 与科学、技术和数学相关的多样性举措吸引了包括个人、大公司、媒体和社区利益集团在内的各种实体的声音。少

2018 年 5 月 4 日提交;最初宣布 2018 年 5 月。

75. 可持续云计算: 基础和未来方向

作者: [Rajkumar buyya, sukhpal singh gill](#)

摘要: 微软、谷歌、脸书和亚马逊等主要云提供商严重依赖数据中心来支持不断增长的对其计算和应用服务的需求。然而, 运行如此庞大的基础设施所需的财务和碳足迹成本对云服务的可持续性产生了负面影响。现有的大部分工作主要集中在最大限度地减少服务器的能耗上。在本文中, 我们为所有资源(包括服务器、网络、存储、冷却系统)的整体管理设计了一个概念模型和实用设计指南, 以提高云数据中心(cdc)的能效并减少碳足迹)。此外, 我们还讨论了可持续云计算的能源和可靠性之间的相互关联的关系, 并强调了相关的研究问题。最后, 我们提出了该领域未来的研究方向, 并为进一步的实际发展奠定了基础。少

2018 年 5 月 1 日提交;最初宣布 2018 年 5 月。

76. 计算机生成的描述对视觉障碍者照片分享体验的影响

作者: [赵玉航](#), [吴少梅](#), [林赛·雷诺兹](#), [石丽·阿曾科特](#)

摘要: 和有视力的人一样, 视障人士也希望在社交网络服务上分享照片, 但很难识别和选择相册中的照片。我们的目标是通过将最先进的计算机生成的描述纳入 **facebook** 的照片共享功能来解决这一问题。我们采访了 12 名视障参与者, 了解他们的照片分享体验, 并为 **facebook** 移动应用程序设计了照片描述功能。我们在一项为期七天的日记研究中, 对 6 名参与者进行了评估。我们发现参与者使用这些描述来回忆和整理他们的照片, 但他们在没有有视力的人的情况下犹豫上传照片。除了有关照片内容的基本信息外, 参与者还想了解更多关于突出物体和人物的细节, 以及照片是否反映了他们的个人美学。我们从自我披露和自我呈现理论的角度讨论这些发现, 并提出新的计算机视觉研究方向, 以更好地支持视障人士的视觉内容共享。少

2018 年 5 月 3 日提交;最初宣布 2018 年 5 月。

77. 伊利诺伊州医院网站上的网络跟踪器流行率

作者: [罗伯特·罗宾逊](#)

摘要: 网络跟踪技术是普遍存在的, 由少数大型技术公司运营。这项技术, 以及对收集到的数据的使用, 都与影响选举、假新闻、歧

视甚至卫生决策有关。人们对这一技术是如何在医院或其他与健康有关的网站上部署的了解甚少。美国伊利诺伊州 210 家公立医院的网站使用网络跟踪器识别工具进行了评估。94% 的医院网站上发现了网络跟踪器，综合医院网站上平均有 3.5 个跟踪器。较小的关键接入医院的网站平均使用 2 个网络跟踪器。识别的最常见的网络跟踪器是谷歌分析，在 74% 的伊利诺伊州医院网站上发现。在发现的网络跟踪器中，88% 由谷歌运营，26% 由 **facebook** 运营。鉴于有消息披露了网页浏览配置文件是如何被使用和滥用的，搜索泡沫，以及算法歧视的可能性，医院领导和决策者必须仔细考虑使用第三方跟踪是否合适医院网站上的技术。少

2018 年 5 月 3 日提交;最初宣布 2018 年 5 月。

78. 源于脸书的暴力: 孟加拉国的一个案例研究

作者:[matur rahman minar](#), [jibon naher](#)

摘要: 脸谱和社交网络一样，是现代的一大创新。在所有社交网站中,脸谱是全世界最受欢迎的社交网络。孟加拉国也不例外。人们使用脸书有各种原因，如社交网络和通信、网上购物和商业、知识和经验分享等。然而，孟加拉国最近发生的一些事件源于或基于脸书的活动，导致纵火和暴力。在这些事件中，脸书这类社交网络主要被用作引发仇恨和暴力的工具。本案例研究讨论了这些与

技术有关的事件, 并建议今后可能采取的措施, 以防止此类暴力。

少

2018 年 3 月 31 日提交;最初宣布 2018 年 4 月。

79. 滥用嗅探: 自动检测和防御滥用 facebook 朋友

作者:[sajedul talukder](#), [bogdan carbunar](#)

摘要: 转化符利用社交网络好友关系从用户那里收集敏感数据, 并以包括假新闻、网络欺凌、恶意软件和宣传在内的滥用行为作为攻击他们的目标。例如, 在 80 名用户研究参与者中, 有 71 人至少有 1 名 **facebook** 好友, 无论是在**脸谱**还是现实生活中, 他们都从未与之互动, 或者他们认为这些朋友有可能滥用发布的照片或状态更新, 或发布攻击性信息, 虚假或恶意内容。我们引入 abusiniiff, 这是一个识别**被视为陌生人或虐待者**的 facebook 朋友的系统, 并通过不友好、不关注或限制此类朋友访问信息来保护用户。我们开发了一个调查问卷来检测被感知到的陌生人和朋友虐待。我们介绍了 **facebook** 的相互活动功能, 并表明他们可以训练监督学习算法来预测调查问卷的反应。我们通过几项用户研究对 AbuSniff 进行了评估, 共有来自 25 个国家的 263 名参与者参加了研究。在回答问卷后, 参与者同意分别在 91.6 和 91.6 的个案中取消对滥用者的关注和限制, 在 92.45% 的个案中, 有 92.45% 的个案同意不跟踪和限制滥用者, 并在 92.45 宗的个案

中,有沙箱或非朋友的非虐待者。在未回答调查问卷的情况下,参与者同意对预计是陌生人或虐待者的朋友采取 abusinff 建议的行动,在 78.2% 的案件中。abusinff 增加了参与者自我报告的拒绝陌生人和虐待者的邀请的意愿,他们对朋友虐待的影响的认识,以及他们认为可以保护免受朋友虐待的感觉。少

2018 年 4 月 26 日提交;最初宣布 2018 年 4 月。

80. 基于蜜罐的安全方法: 保护在线社交网络免受恶意配置文件的攻击

作者:fatna elmendili, nistrine maqran,younes el bouzekri el idrissi, habiba chaoui

摘要: 近年来, 社会网络的快速发展和指数化利用促使社会计算的扩张。在社交网络中, 用户通过边缘或链接相互联系, **facebook**、twitter、linkedin 是最受欢迎的社交网络网站。由于这些网站越来越受欢迎, 它们成为网络犯罪和攻击的目标。这主要是基于用户如何使用推特等这些网站。攻击者可以轻松访问和收集个人和敏感用户的信息。用户对安全设置的了解程度较低, 也最不关心安全设置。而且他们很容易成为身份侵犯的受害者。为了检测恶意用户或假配置文件, 不同的技术已被提出像我们的方法, 这是基于使用社交蜜罐发现恶意配置文件。在安全研究人员的启发下, 该方法使用蜜罐观察和分析网络中的恶意活动, 使用社交蜜罐捕获恶意用户。该方法的两个关键要素是: (1)

部署社交蜜罐, 以获取恶意配置文件的信息。(2) 分析这些恶意配置文件的特征以及用于创建分类器的已部署蜜罐的特征, 以便筛选现有配置文件并监视新配置文件。少

2018 年 4 月 26 日提交;最初宣布 2018 年 4 月。

81. 利用友好网络进行社会交互网络中的动态链接预测

作者:[ruthwik r. junuthula](#), [kevin s. xu](#), [vijay k. devabhaktuni](#)

摘要: 联机社交网络 (osn) 通常包含许多不同类型的用户之间的关系。在研究 facebook 等 osn 的结构时, 最常研究的两个网络是友谊和互动网络。友好网络中的链接预测问题已经得到了大量的研究。在独立于友谊网络的互动网络中, 也有过链接预测的研究。本文研究了友谊与互动网络相结合的预测能力。我们假设, 通过利用友谊网络, 可以提高交互网络中链接预测的准确性。我们增强了几种交互链接预测算法, 以整合友谊和预测的友谊。从 facebook 数据上的实验中, 我们发现将友谊融入交互链接预测算法可以提高准确性, 但与将当前的友谊结合起来相比, 将预测的友谊结合起来并不重要友谊。少

2018 年 4 月 23 日提交;最初宣布 2018 年 4 月。

82. 脸谱上的新词

作者:[nikita muravyev](#), [亚历山大·潘琴科](#), [sergei Obiedkov](#)

文摘: 本文对脸谱用户帖子中经常出现的新词和借词进行了研究。我们分析了 2006–2013 期间讲俄语的 facebook 用户撰写的数百万篇公开文章的数据集。从这些, 我们已经建立了最常见的狐狸词的词汇从 open 有你的公司字典缺失的假设是, 许多这样的词只是最近才开始普遍使用。对于以这种方式提取的所有单词, 这一假设肯定不是正确的; 出于这个原因, 我们手动筛选自动获得的列表, 以排除非俄语或不正确的引理单词, 以及其他词典记录的单词或发生在文本中的俄罗斯国家语料库。其结果是一个 168 个单词的列表, 这些单词可能被认为是新词。我们试图对这些新词进行词源分类 (不出意外, 它们中的大多数是最近从英语中借来的, 但也有不少由以前借来的词干组成的新词), 并确定了各种衍生词模式。我们还将单词分为几个大的主题领域, 即 "互联网"、"营销" 和 "多媒体" 等是词汇数量最多的领域之一。我们相信, 与在这一过程中收集的词基一起, 它们可以作为进一步研究新词和词汇过程的起点, 从而使它们被接受为主流语言。少

2018 年 4 月 13 日提交;最初宣布 2018 年 4 月。

83. mater, pater numquam: facebook 广告数据能告诉我们关于男性生育率的什么?

作者 :francesco rampazzo, emilio zagheni, ingmar weber, maria rita testa, francesco billari

摘要: 在许多发展中国家, 仍然缺乏关于出生率和其他人口指标的及时和准确的信息, 特别是在男性生育率方面。利用 **facebook** 广告平台的匿名和汇总数据, 我们对生育平均年龄 (mac) 进行了全球估算, 这是推迟生育的一个关键指标。我们的分析表明, 基于**脸书**数据的生育率衡量与我们有统计数据的国家基于传统数据的传统指标高度相关。例如, 使用**脸书**和联合国数据计算的 mac 数据的相关性分别为女性和男性 0.47 ($p = 4.02 \times 10^{-8}$) 和 0.47 ($p = 2.2 \times 10^{-15}$)。简单回归模型的样本验证外, 表明平均绝对百分比误差为 2.3%。我们使用线性模型和**脸书**数据来计算我们没有数据的国家的男性 mac 的估计数。少

2018 年 4 月 12 日提交;最初宣布 2018 年 4 月。

84. 从社交媒体数据中自动推断人类特征和行为

作者:潘世美,陶丁

摘要: 鉴于人类思维的复杂性及其行为灵活性, 它需要复杂的数据分析来筛选大量的人类行为证据来模拟人类的思维和预测人类的行为。目前, 人们在推特和**脸书**等社交媒体上花费了大量时间。因此, 他们生活和行为的许多方面都被数字化捕获, 并在这些平台上不断存档。这使得社交媒体成为大型、丰富和多样的人类行为证据的重要来源。本文综述了近年来应用机器学习从社交媒体

数据中推断人类特征和行为的工作。我们还将指出未来的几个研究方向。少

2018 年 4 月 11 日提交;最初宣布 2018 年 4 月。

85. 谁陷害了罗杰·驯鹿？通过代码段分类减少对脸书帖子的审查

作者 :[fabio del vigna](#), [marinella petrocchi](#), [alessandro tommasi](#), [cesare zavattari](#), [maurizio tesconi](#)

摘要: 本文认为网络新闻审查, 并集中在身份审查。混淆身份可能是出于不同的原因, 从军事到司法原因。在大多数情况下, 这恰好是为了保护个人不被敌对人民识别和迫害。然而, 作为以信息冗余为特点的协作网络, 同一事实由多个来源报告的情况并不少见, 这些来源在审查方面可能不适用同样的限制政策。另外, 社交网络用户披露个人信息的能力已被证明有能力, 这也导致了这样的现象, 即对新闻的评论可以揭示新闻本身隐瞒的数据。这给了我们一个方法, 可以找出被审查的新闻的主题是谁。我们建议采用文本分析方法进行调整, 以揭示被审查的身份。该方法在合成方案中进行了测试, 但该方案类似于实际使用案例。利用基于基于通过网页页面的帖子和评论的片段进行的上下文分类器的文本分析, 我们取得了很有希望的结果。尽管我们在其中操作的设置相当有限——例如只考虑非常短的片段——但我们的系统成功地检测到了被审查的名字, 在超过 50% 的调查案例中选择了 10 个不同的候选人名字。这优于两个参考基线的结果。本文报道的研究结

果,除了得到彻底实验方法的支持和对自己的兴趣之外,也为进一步调查网络审查的阴险问题铺平了道路。少

2018 年 4 月 10 日提交;最初宣布 2018 年 4 月。

86. 系统监测社交媒体: 2017 年德国联邦选举的案例

作者 :sebastian stier, arnim bleier, malte bonart, fabianmörsheim, mahdi bohlouli, margarita nizhegorodov, lisa posch, jürgenmaier, tobias rothmund , steffen staab

摘要: 大规模收集数字跟踪数据,同时遵守既定的学术标准,是一项相当艰巨的任务。在政治传播方面,重要的挑战是: (1) 界定与运动有关的社交媒体账户和职位 (内容有效性); (2) 实施开展相关社交媒体活动的场所 (构建(3) 捕获所有相关的社交媒体活动 (可靠性), (4) 共享尽可能多的数据以供重用和复制 (客观性)。 gesis-leibniz 社会科学研究所和科布伦茨-兰道大学电子民主方案的这一项目开展了这一努力。我们集中讨论了两个最具政治意义的社交媒体网络,即脸书和推特。少

2018 年 4 月 9 日提交;最初宣布 2018 年 4 月。

87. 代码混合社交媒体文本中单词变体的自动归一化

作者:rajat singh, nurendra choudhary, manish shrivastava

摘要: 推特和脸书等社交媒体平台在多语言社会中越来越受欢迎。这一趋势使南亚语言的波特曼托与英语结合在一起。多种语言作为代码混合数据的混合最近在研究社区中开始流行,用于各种 nlp 任务。代码混合数据由异常(如语法错误和拼写变化)组成。在本文中,我们利用了单词的上下文属性,其中单词的不同拼写变化在一个巨大的嘈杂的社交媒体文本中具有相似的上下文。我们使用单词的分布式表示以无监督的方式捕获属于同一上下文的不同变体。我们的实验表明,基于我们的方法对代码混合数据集进行预处理,可以提高最先进的语音部分标记(pos 标记)和情绪分析任务的性能。少

2018 年 4 月 2 日提交;最初宣布 2018 年 4 月。

88. 在阳光的阴影下: 描述垃圾邮件运动在网络上滥用电话号码

作者: [srishti gupta](#), [druv kuchhal](#) , [payas gupta](#), [mustaque ahamad](#), [manish gupta](#), [ponnurangam kuraguru](#)

摘要: 网络犯罪分子滥用在线社交网络(osn)引诱受害者进入各种垃圾邮件。在不同的垃圾邮件类型中,一个探索较少的领域是 osn 滥用,它利用电话通道来欺骗用户。电话号码通过 osn 进行广告,用户被骗拨打这些号码。为了扩大此类诈骗/垃圾邮件活动的覆盖面,电话号码在脸书、推特、谷歌加、flickr 和 youtube 等多个平台上做广告。在本文中,我们介绍了第一个跨平台活动的驱动特性,这些活动使用多个 osn 平台来覆盖受害者并使

用电话号码进行货币化。在 6 个月的时间里, 我们从推特、脸谱、谷歌加、优酷和 flickr 收集了 23m 帖子, 其中包含 180 万的不同电话号码。对这些帖子进行分组, 帮助我们确定 202 个在全球范围内开展的活动, 其中印度尼西亚、美国、印度和阿拉伯联合酋长国是最著名的发起者。我们发现, 尽管印尼的广告系列产生的数量最高 (320 万帖子), 但到目前为止, 在宣传印尼广告系列的账户中, 只有 1.6% 被暂停。通过检查跨多个 osn 的广告系列, 我们发现 twitter 检测和暂停的帐户比 **facebook** 多 93%。因此, 跨 osn 共享有关滥用相关用户帐户的情报有助于垃圾邮件检测。根据我们的数据集, 如果在 osn 之间共享情报, 大约有 35k 受害者和 880 万美元可以保存。通过分析基于电话号码的垃圾邮件活动在 osn 上运行, 我们重点介绍了 osn 上出现的各种未被探索的基于电话的攻击。少

2018 年 4 月 2 日提交;最初宣布 2018 年 4 月。

89. facebook 的广告平台: 新的攻击载体和干预需求

作者:[irfan faizullabhoy](#) , [Aleksandra korolova](#)

摘要: 随着新工具 (如自定义受众、行为定位和受众洞察) 的引入, 广告定位的功能变得越来越强大。虽然这对企业有利, 因为它能让人们获得更多相关的广告, 但这些工具的力量有缺点。在本文中, 我们关注三个缺点: 侵犯隐私、微目标 (即在特定个人不知道

他们是广告所涉及的唯一内容的情况下接触他们的能力) 以及接触边缘化群体的便利性。以 **facebook** 的广告系统为例, 我们展示了这种缺点的可行性。然后, 我们讨论 **facebook** 对我们负责任地披露调查结果的回应, 并呼吁在快速发展的广告定位生态系统中开展更多的政策、科学和工程工作, 以保护消费者。少

2018 年 3 月 27 日提交;最初宣布 2018 年 3 月。

90. 印度-英语代码混合数据的攻击注释语料库

作者: [ritesh kumar](#), [aishwarya n. reganti](#) , [akshit bhatia](#), [tushar maheshwari](#)

摘要: 随着网络互动的增加, 全球范围内的侵略事件和相关事件, 如手推车、网络欺凌、燃烧、仇恨言论等, 也在不断增加。虽然欺凌或仇恨言论等行为大多早在互联网之前, 但互联网的触角和范围使这些行为具有前所未有的力量 and 影响力, 足以影响数十亿人的生活。因此, 采取一些预防措施, 为使用网络的人们提供保障, 使网络总体上仍然是一个可行的沟通和连接媒介, 这一点至关重要。在本文中, 我们讨论了从印度两个最流行的社交网络和社交媒体平台推特和脸书中开发一个侵略标签和一个附加说明的 hindi-nic 代码混合数据语料库。语料库使用由 3 个顶级标记和 10 个级别 2 标记组成的分层标记集进行注释。最终的数据集包含大约 18k 推特和 21k 脸书评论, 正在发布, 供该领域进一步研究。少

2018 年 3 月 25 日提交;最初宣布 2018 年 3 月。

91. 支持向量机对罗兴亚人运动评语的情绪分析

作者:[hemayet ahmed chowdhury](#), [tanvir alam nibir](#), [md. saiful isam](#)

摘要: 罗兴亚人运动和危机在孟加拉国的政治和经济状况中引起了巨大的轩然。难民运动是一个反复发生的事件,脸谱等社交媒体上仍有大量意见形式的数据,对这些数据的分析很少。为了分析所有罗兴亚人相关帖子的评论,我们必须创建和修改一个基于支持向量机算法的分类器。该代码在巨蟒中实现,并使用科学学习库。目前还没有关于罗兴亚人分析的数据集,因此我们必须使用自己的数据集,即 2500 条正面评论和 2500 条负面评论。我们专门使用了具有线性内核的支持向量机。我们以前使用天真的贝叶斯算法在同一数据集中进行了一次实验,但没有产生令人印象深刻的结果。少

2018 年 3 月 22 日提交;最初宣布 2018 年 3 月。

92. 语境就是一切: 在语义空间中寻找意义统计

作者:[埃里克·泽利克曼](#)

文摘: 本文介绍了语境性的显著性 (co 萨尔), 它是一个简单而明确的度量词在语境中的重要性, 它在理论上更自然、更简单、更

准确地替代了 tf-idf。co 萨尔支持非常小的上下文 (20 个或更多的句子), 超出上下文词, 并且很容易计算。用双迁移短语和非迁移项标记生成的单词向量空间揭示了上下文意义重大的单词不成比例地定义短语。这种关系被应用于产生简单的加权小词包句子嵌入。这种模式的表现优于 SkipThought 和最好的模型训练无序句子在 facebook 的 senteval 的大多数测试, 击败 tf-idf 在所有可用的测试, 一般是相当的最先进的状态。本文还将 cosine 应用于句子和文档的总结, 并提高了视点感知余弦距离。运用意外词语是重要的前提, 认为 co 萨尔是 tf-idf 的替代品, 也是上下文单词重要性的直观度量。少

2018 年 4 月 12 日提交;v1 于 2018 年 3 月 22 日提交;**最初宣布** 2018 年 3 月。

93. 社交媒体不会撒谎: 通过网络异构数据预测 2016 年台湾大选

作者:郑谢,刘冠南, 吴俊杰,谭勇

摘要: 网络媒体的普及吸引了来自不同领域的研究人员, 探索人类行为, 做出有趣的预测。在这项研究中, 我们利用从各种网络平台收集的异质社交媒体数据来预测台湾 2016 年的总统选举。与现有的大多数研究不同的是, 我们对异质信息采取 "信号" 视图, 并采用卡尔曼滤波器将多个信号融合到候选人的日常投票预测中。我们还根据源于金融研究领域的所谓事件研究模型, 定量

地考虑影响选举的事件。我们得到了以下有趣的发现。首先，网络媒体的舆论在预测能力和及时性方面主导了台湾选举预测的传统民调。但离线民调仍然可以缓解网络观点的样本偏差。其次，尽管随着选举日的临近，网上信号趋于一致，但简单的脸谱"喜欢"一直是选举结果的最有力指标。第三，最具影响力的事件与两岸关系有着密切的联系，周子宇国旗事件发生后，在选举前一天播放了道歉视频，使蔡英文的得票率增加了 3.66。本研究证明了网络媒体在政治中的预测能力和信息融合的优势。卡尔曼滤波和事件研究方法的结合有助于数据驱动的政治分析范式，用于预测和归因目的。少

2018 年 4 月 3 日提交;v1 于 2018 年 3 月 21 日提交;最初宣布 2018 年 3 月。

94. 一种实证博弈理论分析的广义方法

作者 :karl tuyls, Perolat perolat, marc lanctot, joel z leibo, thore graepel

摘要: 本文为复杂多因素相互作用的经验博弈理论分析提供了理论边界。我们在经验元博弈中提供了见解，表明元博弈的纳什均衡是真正底层博弈的近似纳什均衡。我们调查并显示需要多少数据样本来获得基础游戏的足够近的近似值。此外，我们将元博弈分析方法扩展到非对称游戏。最先进的只考虑了经验游戏，其中代理可以访问相同的策略集，并且收益结构是对称的，这意味着

代理是可互换的。最后，我们对几个领域的泛化方法进行了实证说明，说明了几个版本的 alphago 算法（对称）的理论和进化动力学，以及人类玩的 blotto 上校游戏的动力学玩家在 facebook 上（对称），以及一个元游戏的例子在 leduc 扑克（不对称），生成的 psro 多智能体学习算法。少

2018 年 3 月 16 日提交;最初宣布 2018 年 3 月。

95. 照亮党派网站的生态系统

作者:[shweta bhatt](#), [sogar joglekar](#), [shehar bano](#), [nhanth sastry](#)

摘要: 本文旨在揭示其他新闻媒体生态系统，这些生态系统被认为在 2016 年美国总统选举期间通过虚假和有偏见的新闻报道影响了观点和信仰。我们检查了 668 超党派网站及其相应的 facebook 页面的大型专业策划列表，并确定了在这一生态系统中调节交通流的关键特征。我们发现了一种模式，即在选举前正在建立新的网站，之后就放弃了。这些网站形成了一个生态系统，通过 "喜欢" 对方的脸谱页面，创建了一个网站到另一个网站的链接。这些做法非常有效，可以以高度党派化的方式在生态系统内引导用户流量，右倾网站链接到其他右倾网站，并喜欢其他左倾网站，也可以通过类似的左倾网站链接到生态系统上的其他网站。左，从而形成一个过滤泡沫的新闻生产者之间类似于过滤泡沫，已被广泛观察到的消费者的党派新闻。虽然在左倾和右倾

网站上都有活动, 但右倾网站的演变更为多样, 占了被废弃网站和党派内部链接的不成比例的比例。我们还考察了超党派新闻消费者的人口特征, 发现美国一些人口较多的人口群体往往是倾向于更靠的网站消费者。少

2018 年 3 月 9 日提交;最初宣布 2018 年 3 月。

96. 大图的简明模糊表示: 一种维度约简方法

作者:faisal n. ab 布-khzam , rana h. mouawi

摘要: 使用大图形表示的海量数据在某些情况下超过了传统计算机的资源。特别是, 与节点数量相比, 边缘可能会占用相当多的内存。但是, 严格的边缘存储对于能够得出所需的结论可能并不总是必不可少的。类似的问题需要记录中的许多变量, 并尝试提取最明显的特征。据说, 这一数据的 "维度" 有所减少。按照考虑到相同目标的方法, 我们可以将图形表示映射到 k 维空间, 并通过测量欧几里得距离来回答相邻节点的查询。我们的答案的准确性会降低, 但会被模糊逻辑来补偿, 模糊逻辑给出了错误可能性的想法。此方法允许在内存中进行合理的表示, 同时保持相当数量的有用信息。通过在一些脸书图表上测试拟议的方法, 获得并报告了有希望的初步结果。少

2018 年 3 月 8 日提交;最初宣布 2018 年 3 月。

97.脸谱 (a) 直播? 社会广播真的是广播吗?

作者:Aravindh raman, gareth tyson, nasanth sastry

摘要: 直播时代又回来了, 但有两个重大变化。首先, 与传统电视广播不同的是, 内容现在通过互联网流播, 使其能够接触到更广泛的受众。其次, 由于各种用户生成的内容平台, 任何人都有可能参与进来, 将自己的内容流向世界。这种新出现的直播趋势通常发生在社交平台上, 用户主要从移动设备上进行现场社交广播, 让他们的朋友 (和公众) 能够实时参与流。随着这类平台的日益普及, 目前互联网基础设施的负担预计将成倍增加。考虑到这一点, 我们探索一个这样一个著名的平台——脸谱直播。我们收集 3tb 的数据, 代表一个月的全球活动, 并探索社会直播的特点。由此, 我们得出了简单而有效的原则, 可以减轻网络负担。然后, 我们通过捕获广播公司或制作视频的用户以及观众或与之交互的用户的地理位置, 剖析视频的全局和超本地属性。最后, 我们研究视频直播时的社会参与, 并区分同一视频点播时的关键方面。整篇论文的一个共同主题是, 尽管脸谱直播的名字很大, 但它的许多属性都偏离了直播和广播的概念。少

2018 年 3 月 7 日提交;最初宣布 2018 年 3 月。

98. 美国社交媒体的两极分化、党派关系与垃圾新闻消费

作者: [vidya narayanan](#), [vlad barash](#), [johnkelly](#) , [bence kollanyi](#), [isa-marianeudert](#), [phillip n. howard](#)

摘要: 什么样的社交媒体用户阅读垃圾新闻? 我们审查了在总统唐纳德·特朗普第一次国情咨文发表之前的三个月内, 最重要的垃圾新闻来源的传播情况。根据不断发布政治新闻和信息的来源, 我们发现这类内容的传播分布不均跨越意识形态的范围。我们在微博上证明 (1) 特朗普支持者网络分享了最广泛的已知垃圾新闻来源, 传播的垃圾新闻比所有其他团体加起来还要多;(2) 在 **facebook** 上, 与共和党网页不同的极硬的右页面共享最广泛的已知垃圾新闻来源, 传播的垃圾新闻比所有其他受众加起来还要多;(3) 平均而言, 微博上垃圾新闻的受众比 **facebook** 公共页面上的受众分享的已知垃圾新闻来源范围更广。少

2018 年 3 月 4 日提交;最初宣布 2018 年 3 月。

99. 多模式检索的学习装饰相关哈希代码

作者:[田大勇](#),[余军](#) , [唐元燕](#),[陶大成](#)

摘要: 在社交网络中, 异构多媒体数据相互关联, 如 youtube 中的视频及其相应的标签和 **facebook** 中的图像文本对。在大型数据集上跨多个模式检索最近的邻居成为一个热门而又具有挑战性的问题。哈希预计将是一个有效的解决方案, 因为它将数据表示为二进制代码。由于按位 xor 操作可以快速处理, 检索时间大大

缩短。现有的多模散列方法很少考虑哈希位之间的相关性。这种相关性对哈希代码有负面影响。当哈希代码长度变长时, 检索性能的改进会变慢。本文提出了多模哈希的最小相关正则化 (mcr)。首先, 使用 sigmoid 函数嵌入数据矩阵。然后, 将 mcr 应用于 sigmoid 函数的输出。由于 sigmoid 函数的输出近似于二进制代码矩阵, 因此所提出的 mcr 可以有效地修饰哈希码。实验表明, 随着码长的增加, 该方法的优越性越来越大。少

2018 年 3 月 1 日提交;最初宣布 2018 年 3 月。

100. 通过社交媒体保护土著人民的土著文化: 伊戈罗特人

作者:khavee agus botangen, shahper vodanovich, jian yu

摘要: 土著知识对人类社会可持续性的价值和相关性推动了对其的保护。这项工作探讨了利用脸书团体在散居国外的伊戈罗特人中促进土著知识的问题。尽管与不同社会同化面临挑战, 但虚拟社区有助于加强伊戈罗特移民与其传统文化的联系。对 20 个脸书团体的帖子进行调查, 确定并对通过社交媒体传播的土著文化要素进行分类。随后对 56 名 Igorot 移民进行的调查显示, 大众社交媒体在交流、振兴、实践和学习土著文化方面发挥着重要作用;激发有效的媒介来利用保存策略。少

2018 年 2 月 26 日提交;最初宣布 2018 年 2 月。

101. 基于 iMacros-based 的 facebook 用户行为分析数据爬虫的设计与实现

作者:mudasir ahmad wani, nancy Agarwal, suraiya jabin, syed zeeshan hussai

摘要: 在分析在线社交网络 (osn) 网站时, 获取所需的数据集仍然是研究人员面临的主要挑战。osn 服务提供商为检索数据而提供的应用程序编程接口 (api) 施加了一些不可避免的限制, 因此很难获得理想的数据集。在本文中, 我们提出了一个基于 imrocler 的基于 imrocler 的 iMacros 技术数据爬虫, 它能够在重新授权的法律框架内从 **facebook** 网站通过浏览器收集可访问的每一条信息。建议的爬虫解决了与 web 数据提取方法和 osn 服务提供商提供的大多数 api 相关的大多数挑战。从**脸书**用户简介中提取了两个广泛的章节, 即个人信息和隔离墙活动。收集到的数据被预先处理成两个数据集, 并对每个数据集进行统计分析, 以绘制语义知识, 并了解 **facebook** 用户的几个行为方面, 如主要由用户披露的信息种类、性别在揭示信息的模式、网络上的高发布内容、网络上的高度执行活动、个人和帖子属性之间的关系等方面的差异。据我们所知, 目前的工作是首次尝试详细描述爬虫设计和基于性别的信息揭示**脸谱**用户的行为。少

2018 年 2 月 18 日提交;最初宣布 2018 年 2 月。

102. 情感线: 多方对话的情感语料库

作者:陈胜国,许朝春, 郭传春,丁浩,黄, 顾伦伟

摘要: 感受情感是区分人和机器的一个重要特征。在所有用于情感检测的多模态资源中, 文本数据集是那些除了语义之外包含的附加信息最少的数据集, 因此被广泛用于测试所开发的系统。然而, 大多数文本情感数据集只包含单个单词、句子或文档的情感标签, 这使得讨论情感的语境流动具有挑战性。在本文中, 我们介绍了情感线, 第一个数据集的情感标签的所有话语中的所有发言只基于其文本内容。情感线对话是从朋友电视脚本和私人 facebook 信使对话收集。然后七种情绪中的一种, 六个 ekman 的基本情绪加上中性的情绪, 被 5 亚马逊 mturkers 标记在每个话语上。来自 2000 个对话的 29 245 个话语在 "情绪线" 中被标记为 "情感线"。本文还为情绪线上的情绪检测模型提供了几个强有力的基线。少

2018 年 5 月 30 日提交;v1 于 2018 年 2 月 22 日提交;**最初宣布** 2018 年 2 月。

103. 人机界面与智能机器的联学

作者:李昌成,王美辉,高丽伟,久保田那玉, 林鲁安, 北冈新雅, 王玉德, 苏顺峰

摘要: 机器学习已成为控制论系统非常流行的方法, 一直被认为是计算智能领域的重要研究。然而, 当涉及到智能机器, 它不只是

关于方法。我们需要考虑系统和控制论, 以及将人纳入循环。本文的目的如下: (1) 将 **facebook** 的开源 **facebook ai** 研究 (fair) 暗林程序与项目响应理论 (irt) 集成到新的开放学习系统, 即 ddf 学习系统中;(2) 将 ddf go 与机器人集成, 即机器人 ddf go 系统;(3) 邀请专业的围棋玩家参加活动, 用智能机器在现场玩围棋游戏。研究小组将把这项技术应用于教育, 比如玩游戏, 提高孩子们对学习数学、语言等话题的专注力。有了检测到的脑电波, 机器人就能说出一些对学生非常重要的词, 并在未来的课堂上帮助老师。少

2018 年 2 月 19 日提交;最初宣布 2018 年 2 月。

104.skylite: 低海拔无人机网络的端到端设计, 用于提供 lte 连接

作者:[karthikeyan sundaresan](#), [eugenechai](#), [ayon chakraborty](#) , [sampath Rangarajan](#)

摘要: 无人驾驶飞行器 (无人机) 有可能改变广域无线连接的格局, 将其带到连接不受或根本不存在的地区 (如农村地区) 或因灾害而受到损害的地区。虽然谷歌的 loon 项目和 **facebook** 的 aquila 项目是在这一方向上的高空、长耐久性的 uav 连接工作的例子, 但电信运营商 (如 at & amp; t 和 verizon) 一直在探索低空基于 uav 的用于按需部署的 lte 解决方案。可以理解的是, 这些项目尚处于初期阶段, 在实现和部署方面面临严峻挑战。本文档的目标是让读者了解这些非常规连接解决方案所带来的挑战

和潜力。我们的目标是探索这种基于无人机的连接网络的端到端设计，特别是在提供 lte 连接的低空无人机网络的背景下。具体来说，我们的目标是以跨周期的方式突出跨多个层（访问、核心网络和回程）的挑战，以及设计空间本身的丰富性和复杂性。为了帮助感兴趣的读者在这个复杂的设计空间中找到解决方案，我们还阐明了一个这样的端到端设计的概述，即 skylite——一个自组织的低海拔无人机网络，可在所需的地方提供优化的 lte 连接地区。

少

2018 年 2 月 20 日提交;v1 于 2018 年 2 月 16 日提交;最初宣布 2018 年 2 月。

105.facebook 在欧洲广告中的使用敏感数据

作者:[joségonzález cabañas](#), [angel cuevas](#), [rubén cuevas](#)

摘要: 即将出台的《欧洲一般数据保护条例》禁止处理和利用某些类别的个人数据（健康、政治取向、性偏好、宗教信仰、族裔血统等），因为这一点很明显。恶意使用此类信息可能会产生的隐私风险。这些类别称为敏感的个人数据。**facebook** 最近在西班牙因收集、存储和处理用于广告目的的敏感个人数据而被罚款 120 万欧元。本文量化了欧洲联盟（eu）中被贴上与敏感个人数据相关联的兴趣标签的 **facebook** 用户部分。我们的研究结果显示，**facebook** 为 73% 的欧盟用户贴上了敏感兴趣的标签。这相

当于欧盟总人口的 40%。我们还估计, 恶意第三方可能会公布已被分配敏感兴趣的 **facebook** 用户的身份, 其成本最低为每个用户 0.015 欧元。最后, 我们建议并实施一个网络浏览器扩展, 以告知 **facebook** 用户 **facebook** 分配给他们的敏感利益。少

2018 年 2 月 14 日提交;最初宣布 2018 年 2 月。

106. 涉及 facebook 讨论组的攻击策略和时间分析

作者:赖春明,王晓云,洪云峰, 林玉成,吴丽丽, 帕特里克·麦克丹尼尔,哈桑·卡姆

摘要: 在线社交网络 (osn) 讨论小组正在对政治对话产生重大影响。在没有访问控制机制的情况下, 任何用户都可以对任何 osn 线程做出贡献。个人可以利用此特性执行有针对性的攻击, 这增加了后续恶意行为 (如网络钓鱼和恶意软件分发) 的可能性。这类行动还将破坏媒体、政客及其支持者之间的桥梁。出于安全管理的考虑, 将恶意网络攻击与在线社会交往融合在一起, 带来了全新的挑战。在本文中, 我们描述了我们的建议, 一种新的方法来研究和理解攻击者使用恶意 url 在 **facebook** 讨论组中传播恶意 url 的策略。我们定义和分析与预测新闻媒体组织创建的线程上的攻击可能性相关的问题。我们使用宏观静态特征和职位和线程的微观动态演化的组合来确定精度超过 90% 的可能目标。我们的次要目标之一是在很短的 (10 分钟) 时间内做出这样的预测。

我们希望本文中提供的数据和分析能够更好地了解攻击者策略和足迹, 从而在处理社交网络上的网络攻击时开发新的系统管理方法。少

2018 年 2 月 13 日提交;最初宣布 2018 年 2 月。

107. 《关于军事和国家安全的垃圾新闻》：针对美国军人和退伍军人的社交媒体虚假信息运动

作者:[john d. gallacher](#), [vlad barash](#), [phillip n. howard](#), [john kelly](#)

摘要: 社交媒体为现役军人和退伍军人提供政治新闻和信息。我们分析了 twitter 和 **facebook** 用户的子组, 他们花时间从网站上消费垃圾新闻, 这些网站的目标是美国军人和老兵, 他们有阴谋论、错误信息和其他形式的关于军事的垃圾新闻。国家安全问题。

(1) 在微博上, 我们发现, 现任和前任军事人员与广泛的极端主义、以俄罗斯为重点和国际阴谋分组网络之间存在着重大和持续的互动。(2) 在**脸书**上, 我们发现军队和退伍军人公共网页与致力于政治阴谋的分组以及政治领域的双方之间进行了重大和持续的互动。(3) 在 **facebook** 上, 对阴谋论和政治权利最感兴趣的用戶似乎传播的垃圾新闻最多, 而军人或退伍军人的用戶是最复杂的新闻消费者,并通过网络分享很少的垃圾新闻。少

2018 年 2 月 10 日提交;最初宣布 2018 年 2 月。

108. 无消息内容的应用使用模式及预测模型研究

作者: [avi rosenfeld](#), [sigal sina](#), [david sarne](#), 或 [avidov](#), [sarit krous](#)

摘要: 互联网社交网络已成为一个无处不在的应用程序, 使人们能够轻松地共享文本、图片和音频和视频文件。受欢迎的网络包括 what ' **sapp**、**facebook**、reddit 和 linkedin。我们对 what ' sapp 社交网络的使用情况进行了广泛的研究, 该网络是一个正在迅速取代短信的互联网短信应用程序。为了更好地了解人们对网络的使用情况, 我们对来自 100 多个用户的 600 多万条消息进行了分析, 目的是利用活动数据建立人口预测模型。我们对数据进行了广泛的统计和数值分析, 发现不同性别和年龄的人在 whatsapp 使用方面存在显著差异。我们还将数据输入到 weka 数据挖掘包中, 研究了从决策树和贝叶斯网络算法创建的模型。我们发现, 在几乎所有的信息和群体属性中, 不同的性别和年龄人口结构的使用习惯有显著的不同。我们还注意到用户组行为的差异, 并创建了预测模型, 包括给定组具有相对较多的文件附件的可能性, 如果组包含更多的参与者、更高的活动频率、更快响应时间和更短的消息。我们成功地量化和预测了用户的性别和年龄人口。同样, 我们能够预测不同类型的组使用情况。所有模型都是在不分析消息内容的情况下构建的。我们对所有预测模型中包含的特定属性进行了详细讨论, 并根据这些结果提出了可能的应用程序。少

2018 年 2 月 9 日提交;最初宣布 2018 年 2 月。

109. 两极分化与虚假新闻: 潜在误报目标预警

作者 :[micha del vicario](#), [walter quattrociocchi](#), [antonio sca](#), [fabiana zollo](#)

摘要: 用户两极分化和确认偏差在网络社交媒体上的错误信息传播中起着至关重要的作用。我们的目的是利用这些信息, 事先确定骗局和假新闻的潜在目标。本文介绍了一个通用的框架, 以快速识别社交媒体上的两极分化的内容, 从而 "预测" 未来的假新闻话题。我们在一个庞大的意大利 **facebook** 数据集上验证了拟议方法的性能, 表明我们能够准确率 7% 地识别容易受到错误信息影响的主题。此外, 这类信息可能会作为一个新功能嵌入到一个额外的分类器, 能够识别假新闻, 准确率为 91%。我们的方法的新颖之处在于考虑到与用户在网络社交媒体上的行为相关的一系列特征, 为平滑两极分化和减少错误信息现象迈出了重要的第一步。少

2018 年 2 月 5 日提交;最初宣布 2018 年 2 月。

110. 基于文字层次融合的多模态情绪分析与强化学习

作者:[陈明海](#),[王森](#),[保罗·浦良](#),[塔达斯·巴尔特鲁什蒂斯](#), [amir Tadas](#), [louis-philip·莫朗西](#)

摘要: 随着 youtube 和 facebook 等视频共享网站的日益普及, 多式联运情绪分析越来越受到科学界的关注。与以往以语音段整体信息为中心的多模情分析作品——如单词袋表示和平均面部表情强度——相相反, 我们开发了一种新的多模态情绪深层体系结构在单词级别执行模式融合的分析。本文提出了由 2 个模块组成的具有时间注意 (gme-lstm (a)) 的门式多模态嵌入式 lstm 模型。门控多式联运嵌入在有嘈杂模式的情况下, 缓解了融合的困难。具有时间注意的 lstm 在输入模式之间以更精细的融合分辨率执行字级融合, 并注意到最重要的时间步骤。因此, gme-lstm (a) 能够通过时间更好地模拟语音的多模态结构, 并具有更好的情感理解。通过获得最先进的情绪分类和回归结果, 证明了该方法在情感强度和主观性分析 (cmu-mosi) 数据集上公开可用的多式联运语料库的有效性。对模型的定性分析强调了时间注意层在情绪预测中的重要性, 因为额外的声学 and 视觉模式是嘈杂的。我们还演示了门控多式联运嵌入在有选择地过滤这些嘈杂模式的有效性。我们的研究结果和分析为人类交际中的情绪分析研究开辟了新的领域, 为多模态融合提供了新的模型。少

2018 年 2 月 3 日提交;最初宣布 2018 年 2 月。

11. 推特滥用行为的大规模众包与表征

作者: [antigoni-maria founta](#), [constantinos djouvas](#), [despoina chatzakou](#), [liasleontiadis](#), [jeremy blackburn](#), [gianluca stringini](#), [athena vakali](#), [michael sirivianos](#), [nicolas kourtellis](#)

摘要: 近年来, 攻击性、辱骂性和可恨性的语言、性别歧视、种族主义和其他类型的攻击性和网络欺凌行为越来越频繁地表现出来, 并在许多在线社交媒体平台上表现出来。事实上, 过去的科学工作侧重于在脸书和推特等大众媒体上研究这些形式。在这些工作的基础上, 我们以整体的方式对微博上各种形式的虐待行为进行了为期 8 个月的研究。从过去的工作出发, 我们检查了各种各样的标签方案, 这些方案同时涵盖了不同形式的虐待行为。我们提出了一种增量和迭代方法, 利用众包的强大功能, 使用一组与滥用相关的标签来注释大量的推特。事实上, 通过应用我们的方法, 包括统计分析的标签合并或消除, 我们确定了一组减少但稳健的标签。最后, 我们提供了我们收集和注释的 10 万条推特数据集的第一个概述和发现, 我们公开提供这些信息, 供进一步的科学探索。少

2018 年 4 月 15 日提交;v1 于 2018 年 2 月 1 日提交;**最初宣布** 2018 年 2 月。

112. 一种用于滥用检测的统一深度学习体系结构

作者 : [antigoni-maria founta](#), [despoina chatzakou](#), [nicolas kourtellis](#), [jeremy blackburn](#), [athena vakali](#), [ilas leontiadis](#)

摘要: 仇恨言论、攻击性语言、性别歧视、种族主义和其他类型的虐待行为已成为许多网络社交媒体平台的普遍现象。近年来, 这种不同的虐待行为以越来越频繁和程度的强度表现出来。这是因

为推特和脸书等大众媒体平台开放和愿意主持敏感或有争议议题的内容。然而，这些平台没有充分解决网上虐待行为的问题，它们对有效发现和阻止这种不当行为的反应仍然有限。在本文中，我们研究这个复杂的问题，遵循一个更全面的方法，其中考虑虐待行为的各个方面。为了使这种方法变得切实，我们专注于 twitter 数据，并从滥用发布行为的不同角度分析用户和文本属性。我们提出了一个深入的学习架构，它利用了各种各样的可用元数据，并将其与推文中自动提取的隐藏模式结合起来，以检测高度相互关联的多种滥用行为规范。我们以无缝、透明的方式应用这种统一的体系结构来检测不同类型的滥用行为（仇恨言论、性别歧视与种族主义、欺凌、讽刺等），而无需为每个任务对模型体系结构进行任何调整。我们使用多个数据集测试该方法，这些数据集可解决 twitter 上不同和多重滥用行为。我们的研究表明，它在很大程度上优于最先进的方法（根据数据集，auc 的改进在 21% 到 45% 之间）。少

2018 年 2 月 21 日提交;v1 于 2018 年 2 月 1 日提交;**最初宣布** 2018 年 2 月。

113. 基于图形的半监督学习在 facebook 上进行多账户检测

作者:王晓云,赖春明,洪云峰,谢卓瑞, 吴费利克斯

摘要: 在社交网络中, 一个用户可以通过在不同的新闻页面上积极评论, 创建多个账户来传播自己的观点并影响他人。将这种不正常活动降级对社交网络及其社区都有好处, 第一步是发现这些账户。然而, 检测具有挑战性, 因为这些帐户可能具有非常现实的名称和合理的活动模式。本文研究了三种不同的方法, 并提出了将图形嵌入与半监督学习结合起来, 以预测一对账户是否由同一用户创建。我们进行了广泛的实验分析, 以了解输入数据和算法参数/优化的变化如何影响预测性能。我们还发现, 局部信息对这种预测的重要性高于全局信息, 并指出了导致最佳结果的门槛。我们用来自中东的 6700 页脸书页面测试了拟议的方法, 对同名用户来说, 平均准确率为 0.996 页, 哥伦比亚联合自卫队 (曲线下面积) 为 0.996; 根据美国 2016 年选举数据集, 我们获得了最好的 auc 为 0.877 为不同名称的用户。少

2018 年 1 月 29 日提交;最初宣布 2018 年 1 月。

114. 从脸谱利益看移民同化

作者: [antoine dubois](#), [emilio zagheni](#), [kiran garimella](#), [ingmar weber](#)

摘要: 移徙者的同化是欧洲社会面临的一个重大挑战, 部分原因是近年来难民突然激增, 部分原因是长期的人口趋势。在本文中, 我们使用 **facebook** 的数据为广告商研究德国讲阿拉伯语的移民的同化程度, 从他们在网上表达的兴趣可以看出这一点。我们的

研究表明, 同化的梯度沿着人口关系, 使用的语言和原籍国。鉴于难以及时收集移民数据, 特别是与文化同化有关的特征, 我们开发的方法和我们提供开放的新研究领域的结果, 计算社会科学家完全有能力地址。少

2018 年 8 月 31 日提交;v1 于 2018 年 1 月 29 日提交;最初宣布 2018 年 1 月。

115. 社交媒体上的竞选活动: 政治家、听众以及在脸书和推特上调解政治传播

作者 :sebastian stier, arnim bleier , haiko lietz , markus strohmaier

摘要: 尽管相当多的研究集中在网络竞选上, 但目前仍不清楚政治家们如何在政治沟通中使用不同的社交媒体平台。本文以 2013 年德国联邦竞选为重点, 调查选举候选人是否涉及对大众观众最重要的话题, 以及他们的传播在多大程度上受到脸谱功能的影响和推特。根据竞选期间进行的有代表性的调查的不限成员名额的答复, 我们训练了一种可解释的贝叶斯语言模型, 以确定政治话题。将该模型应用于候选人及其直接受众的社交媒体信息, 我们发现两者都优先考虑的主题与大众受众不同。分析还显示, 政客们将脸书和推特用于不同目的。我们将各种发现与受众和社会技术环境的特殊性所引起的社交媒体政治传播的调解联系起来。少

2018 年 1 月 26 日提交;最初宣布 2018 年 1 月。

116. 减轻对 oauth 2.0 和 openid 连接的 csrf 攻击

作者:[李万鹏](#),[克里斯·米切尔](#),[陈晓明](#)

摘要: 数百万用户经常使用他们的 google、facebook 和 microsoft 帐户登录支持 oauth 2.0 和基于 openid 连接的单一登录的网站。因此, oauth 2.0 和 openid 连接的安全性至关重要, 在理论和实践中都得到了广泛的研究。不幸的是, 正如这些研究所表明的那样, 这两个方案的现实实施往往容易受到攻击, 特别是跨站点请求伪造 (csrf) 攻击。在本文中, 我们提出了一种新的技术, 可用于减轻 csrf 对 oauth 2.0 和 openid 连接的攻击。

少

2018 年 1 月 24 日提交;最初宣布 2018 年 1 月。

117. 基于随机游走的节点相似性学习网络

作者:[jeremy g. hoskins](#), [cameron musco](#), [christopher musco](#) ,
[charalampos e. tsourakakis](#)

摘要: 在线社交媒体领域的数字存在带来了巨大的隐私风险。在本工作中, 我们考虑对社交网络的隐私威胁, 在社交网络中, 攻击者可以访问基于随机步行的节点相似性的子集, 例如有效的电阻 (即, 上下班时间) 或个性化的评分分数。使用这些相似性, 攻击

者的目标是尽可能多地推断有关基础网络的信息，包括任何剩余的未知的对等节点相似性和边缘。对于有效的电阻度量，我们表明，只需一小部分测量值，攻击者就可以在社交网络中了解到很大一部分边缘，即使测量是嘈杂的。我们还表明，可以学习一个图形，准确匹配基础网络上的所有其他有效电阻。从数据挖掘的角度来看，第二个观察结果很有趣，因为准确计算所有有效电阻可能成本很高。作为一种替代方法，我们的图形仅从近似有效电阻的子集中获得，可作为代理，应用于使用有效电阻探测图形结构的应用，包括图形聚类、节点中心评估和异常检测。我们通过使用两个优化问题将图形学习目标形式化，从而获得结果。一种公式是凸的，可以在多项式时间内得到很好的求解。另一个不是，但我们用投影梯度和坐标下降有效地解决了它。我们在从 **facebook** 获得的一些社交网络上展示了这些方法的有效性。我们还讨论了如何将我们的方法推广到其他基于随机步行的相似性，如个性化 pagerank 。我们的代码可在 <https://github.com/cnmusco/graph-similarity-learning>。少

2018 年 1 月 22 日提交;最初宣布 2018 年 1 月。

118. 网络新闻源的多来源社会反馈

作者:nuno moniz , luís torgo

摘要: 社交媒体平台的兴起导致用户生成的内容大量, 使得与信息检索、推荐系统、数据挖掘和机器学习等领域有关的研究激增。然而, 缺乏全面的基线数据集, 无法进行彻底的评价比较, 已成为一个重要问题。在本文中, 我们展示了谷歌新闻和雅虎新闻等知名聚合器的大量新闻项目数据集, 以及他们各自在多个平台上的社会反馈:脸谱、谷歌 + 和 linkedin。所收集的数据涉及 2015 年 11 月至 2016 年 7 月期间的 8 个月, 约有 100, 000 个新闻项目涉及四个不同的主题: 经济、微软、奥巴马和巴勒斯坦。此数据集是为预测分析任务中的评估比较量身定制的, 但允许在其他研究领域执行任务, 如主题检测和跟踪、短篇情感分析、第一故事检测或新闻推荐。少

2018 年 1 月 22 日提交;最初宣布 2018 年 1 月。

119. 大规模灾害后二手车的公共情绪与需求: 基于脸谱页面的社交媒体情绪分析

作者:[yuya shibuya](#), [Hideyuki tanaka](#)

摘要: 在一场大规模灾难发生后, 已经有各种研究分析公众情绪。然而, 很少有研究关注社交媒体上的公众情绪与其对现实世界中人们活动的结果之间的关系。本文利用脸谱页面对 2011 年东日本大地震和海啸后的长期情绪进行了分析, 目的是调查公众情绪与受影响地区民众实际需求之间的相关性。水灾难。此外, 我们还

试图分析不同类型的与灾害有关的通信是否在物理世界中对人们的活动造成了不同类型的关系。我们的分析显示,与地理信息有关的传播情绪可能会受到受损区域内情绪的影响,与受损地区二手车价格呈正相关。另一方面,灾害兴趣通信的情绪与二手车价格呈负相关,这种情绪可能受到对灾害感兴趣的人的更多影响,但在受损地区之外。这一结果可以解释为,当人们开始复苏时,二手车价格会上涨,因为他们的情绪变得更加积极。这项研究表明,对于长期的灾后恢复分析,我们需要考虑直接受灾害影响的当地人和没有直接受灾害影响的非本地人发布的在线通信的不同特点。

少

2018 年 1 月 22 日提交;最初宣布 2018 年 1 月。

120. 解开社交媒体机器人: 指导研究和政策的一种类型学

作者:[robert gorwa](#), [douglas guilbeault](#)

摘要: 在大选举期间数字影响力操作的广泛报道中,决策者、学者和记者对社交媒体“机器人”的政治影响越来越感兴趣。最近,脸谱和推特等平台公司被传唤为机器人作证,作为对2016年美国总统选举期间数字化外国操纵行为调查的一部分。面对公众和立法者越来越大的压力,这些公司被指示打击显然是恶意的机器人账户。但正如本文所表明的,自上世纪90年代最早的关于机器人的著作以来,对于机器人到底是什么,机器人到底做了什么,一

直存在着相当大的困惑。我们认为，多种形式的歧义是当代与药物相关政策背后的许多复杂性的原因，在制定成功的政策干预之前，对机器人有更全面的理解——特别是如何它们是定义和测量的——将是必需的。在本文中，我们提供了不同类型机器人的历史和类型，提供了明确的指南，以更好地对政治自动化进行分类，并解开它可能对当代技术政策产生的影响，并概述了主要的挑战和未来与机器人有关的研究人员和立法者都将面临的模糊之处。少

2018 年 7 月 28 日提交;v1 于 2018 年 1 月 21 日提交;**最初宣布** 2018 年 1 月。

121. 危机信息学的无监督哈希塔格检索和可视化

作者:[姚谷](#),[梅扬克·凯里瓦尔](#)

摘要: 在推特这样的社交媒体中，哈希标签承载着大量的语义信息，可以很容易地与正文区分开来。以一种有意义的方式探索和可视化哈希标签的空间可以为数据集提供重要的洞察，尤其是在危机情况下。在本演示论文中，我们提出了一个功能强大的原型 hashviz，它接收了在危机局势（如拉斯维加斯枪击案）后收集的推特，并使用了快速文本包技巧语义嵌入算法（从[脸书研究](#)）将单词和哈希标签嵌入到向量空间中。用这种方法获得的哈希标记向量可以在二维中使用 t-SNE 维数约简算法进行可视化。尽管存在多个 twitter 可视化平台，但 hashviz 的特点是简单、可扩

展、交互式和便携, 足以部署在服务器上, 用于在任意灾难发生后收集的百万推特公司, 而无需特殊安装、技术专长、人工监督或昂贵的软件或基础设施投资。虽然简单, 但我们表明, hashviz 提供了一种直观的方式来总结和洞察正在发展的危机局势。hashviz 也是完全无人监督的, 不需要手动输入就可以从原始语料库转到可视化和搜索界面。我们以最近的拉斯维加斯大屠杀为例, 仅使用客户端的 web 浏览器来说明 hashviz 的潜力。少

2018 年 1 月 17 日提交;最初宣布 2018 年 1 月。

122. 深度搜索: 基于内容的图像搜索和检索

作者:tanya piplani, david bamman

摘要: 今天的互联网大多由包括视频和图像在内的数字媒体组成。随着像素成为大多数交易在互联网上发生的货币, 有一种相对轻松地浏览这片信息海洋的方式变得越来越重要。youtube 每分钟上传 400 个小时的视频, 在 instagam、**facebook** 等网站上浏览了数百万张图片。在深度学习和成功领域的最新进步的启发下, 我们在图像字幕、机器翻译、单词 2vec、跳过思想等各种问题上取得了成功, 我们提出了一种基于深度学习的自然语言处理模式允许用户输入要搜索的图像类型的描述, 并作为响应, 系统检索与查询在语义上和上下文上相关的所有图像。以下各节介绍了两种方法。少

2018 年 1 月 11 日提交;v1 于 2018 年 1 月 9 日提交;最初宣布 2018 年 1 月。

123. 脸书疫苗接种辩论的两极分化

作者: [ana lucia schmidt](#), [fabiana zollo](#), [antonio sca](#), [comelia betsch](#), [walter quattrociochi](#)

摘要: 疫苗犹豫已被认为是一个重大的全球健康威胁。在社交媒体上获得任何类型的信息都被认为是犹豫的潜在强大影响因素。最近在疫苗接种以外的其他领域进行的研究表明, 在没有中介的情况下通过互联网获取大量内容, 就会导致两极分化群体中用户的严重隔离。用户选择符合自己的信仰体系的信息, 往往忽视不同的信息。本文对社交媒体在疫苗接种领域的使用是否存在两极分化进行了评估。我们在 **facebook** 上进行彻底的定量分析, 分析 260 万用户在 7 年 5 个月的时间内与 298.018 帖子互动。我们使用社区检测算法自动检测用户活动中出现的社区, 并量化社区随着时间的推移的凝聚力。我们的研究表明, 疫苗的含量消费以回声室效应为主, 两极分化多年来有所增加。社区产生于用户的消费习惯, 即大多数用户只使用有利于疫苗或不支持疫苗的信息, 而不是两者兼而有之。回声室的存在或许可以解释为什么提供准确信息的社交媒体运动可能覆盖面有限, 可能只在分组中有效, 甚至可能煽动意见的进一步两极分化。将不同的信息引入分

组是无视的,可能产生事与愿违的效果,进一步加强了分组内现有的意见。少

2018 年 1 月 9 日提交;最初宣布 2018 年 1 月。

124. 社交媒体关注增加文章访问量: 对 peerj 艺术家级推荐数据的调查

作者:[王贤文](#),[崔云雪](#),[李庆春](#),[郭新辉](#)

摘要: 为了更好地了解社交媒体在学术文章传播中的作用,利用 345 天收集的 110 篇 peerj 文章的每日更新推荐数据,分析了社交媒体关注度与学术论文关注的关系。由社交媒体执导的文章访客。我们的研究表明,peerj 文章的社交媒体存在率很高。约 68.88% 的论文至少从该杂志的官方账户 @PeerJ 以外的推特账户收到一条推特。社交媒体的关注增加了学术文章的传播。Altmetrics 不仅可以作为传统引文措施的补充,而且在增加文章下载量和促进学术文章的影响方面发挥着重要作用。不同社交媒体平台的网络关注之间也存在显著的相关性。**脸谱**份额较多的文章往往会得到更多的推特。时间趋势表明,社会关注在出版后立即出现,但不会持续很久,社交媒体指导的文章观点也是如此。少

2018 年 1 月 8 日提交;最初宣布 2018 年 1 月。

125. 利用卷积和经常性神经网络和支持向量机确定警察部门脸书网帖的应急阶段

作者: [nicolai pogrebnyakov](#), [edgar maldonado](#)

摘要: 在应急响应中对社交媒体员额进行分类是一个重要的实际问题: 准确的分类可以帮助自动处理此类信息, 并帮助其他应急人员和公众及时对紧急情况作出反应。这项研究的重点是对美国警察部门的脸谱信息进行分类。随机选择的 5,000 条消息用于培训分类器, 这些分类器区分了四类消息: 应急准备、响应和恢复以及一般参与消息。用单词袋和单词 2vec 表示特征, 并使用支持向量机 (svm) 和卷积 (cnn) 和递归神经网络 (rnn) 构建模型。性能最好的分类器是具有自定义训练的 word2vec 模型来表示特征的 mn, 它实现了 0.839 的 f1 度量。少

2018 年 1 月 24 日提交;v1 于 2018 年 1 月 2 日提交;最初宣布 2018 年 1 月。

126. 敌对用户社交博客和聊天安全检测模型的建立

作者: [shubhankar gupta](#), [nitin](#)

摘要: 在世界范围内, 大量的人通过在线聊天进行互动。社交和专业平台的数量显著增加, 如 what'sapp、facebook 和推特, 使人们能够与他人分享他们的经验、观点和知识。遗憾的是, 随着

网络交流被嵌入我们的日常交流中，不文明和不当行为已经占据了许多细微差别，从职业不当行为到职业衰败。一般情况下，燃烧始于粗鲁的信息和评论的交流，而这些信息和评论又会引发更大规模的燃烧。为了防止在线通信降级，必须避免恶意用户进入通信平台。本文提出了一种安全检测模型和一种检测和防止在线燃烧的工具。它检测到在聊天或发布博客时燃烧的存在，审查骂人的话，以及阻止用户从燃烧。少

2017 年 12 月 26 日提交;最初宣布 2017 年 12 月。

127.pers: 个性化和可探索的 poi 推荐系统

作者:[ramesh baral](#), [tao li](#)

摘要: 基于位置的社交网络 (lbsn) (例如 **facebook**) 有许多因素 (例如, 评级、入住时间等) 对兴趣点 (poi) 的建议起着至关重要的作用。与评级不同的是, 评论可以帮助用户阐述自己的意见, 并分享相关利益因素 (方面) 的消费体验程度。虽然现有的一些推荐系统一直在使用用户评论, 但大多数审查的透明度较低, 无法解释。这些原因引起了人们对可解释和解释性建议的极大关注。据我们所知, 这是第一份利用用户评论的论文, 将不同方面的情绪和意见纳入个性化和可解释的 poi 推荐。本文提出了一种称为 pers (个性化的 poi 重组系统) 的模型, 该模型利用深部神经网络对评论方面的类别相关性进行建模, 并将用户方面范畴的双边

关系表述为二部关系。并利用双部分基于核心和基于等级的方法对可解释的推荐进行建模。本文的主要贡献是: (一) 根据用户通过评论发布的方面对用户和地点进行建模; (二) 利用深度神经网络对评论方面的类别相关性进行建模; (三) 规定纳入多个在 poi 建议模型中, 上下文 (例如, 分类、空间等), (iv) 它将用户在方面类别上的偏好表述为双边关系, 将其表示为位置方面类别的二分图, 并对可解释的图进行建模。建议与有序密集子图提取的概念使用二部基于和排名的方法, (v) 它评估生成的建议与三个真实世界的数据集。少

2017 年 12 月 20 日提交;最初宣布 2017 年 12 月。

128. 半监督多式联运哈希

作者:田大勇,龚茂国,周德云,焦石,于磊

摘要: 由于数据量巨大, 以多种方式 (如 facebook 上的图像文本对和 youtube 上的视频标记对) 在相关数据之间检索最近的邻居已成为一项具有挑战性的任务。将数据嵌入二进制代码的多模式哈希方法可以提高检索速度, 降低存储需求。由于无监督多模式哈希方法通常不如有监督的方法, 而监督的方法需要过多的手动标记数据, 本文提出的方法利用部分标签设计了一种半监督多模式散列方法方法。首先计算数据矩阵和标签矩阵的变换矩阵。然后, 利用这些变换矩阵, 引入模糊逻辑来估计无标记数据的标

签矩阵。最后，利用估计标签矩阵学习每种模式中数据的哈希函数，生成统一的二进制代码矩阵。实验表明，所提出的具有 50% 标签的半监督方法可以在比较的受监督方法中获得中等性能，并能达到 90% 标签的最佳监督方法的近似性能。由于标签只有 10%，所提出的方法仍然可以与最差的监督方法竞争。少

2017 年 12 月 9 日提交;最初宣布 2017 年 12 月。

129. 基于脸书的社会情感挖掘技术在脸书后反应预测中的应用

作者: [florian krebs](#), [bruno lubascher](#), [tobias moers](#), [pieter scayap](#), [gerasimos spanakis](#)

摘要: 截至 2016 年 2 月,脸谱允许用户使用五个所谓的 "反应" 来表达他们对帖子的经历情绪。本研究论文提出并评估了预测这些对公司 (如连锁超市) 公共页面上用户帖子的反应的替代方法。为此,我们从大型连锁超市的脸谱网页上收集了帖子 (及其反应),并构建了一个可供其他研究的数据集。为了预测一个新的岗位的反应分布,神经网络结构 (卷积神经网络和递归神经网络) 使用预先训练的单词嵌入进行了测试。通过在每个帖子的评论上引入情感挖掘的引导方法,提高了神经网络的结果。最终模型 (神经网络和基线情感矿工的结合) 能够预测 **facebook** 帖子上的反应分布,平均平方误差 (或错误分类率) 为 0.135。少

2017 年 12 月 8 日提交;最初宣布 2017 年 12 月。

130. 网络社交网络中的病毒内容传播

作者: [giannis haralabopoulos](#), [ioannis anagnostopoulos](#) ,
[sherali Zeadally](#)

摘要: 信息流是在线社交网络 (osn) 不断交流的结果。osn 用户全天实时创建和共享不同类型的信息。病毒是作为一个术语引入的, 用来描述在很小的时间范围内接触到广泛受众的信息。例如, 我们测量在 reddit 中提交的信息的传播, 识别不同的模式, 并在 twitter、**facebook** 和图像和多媒体、imgur 和 youtube 的两个托管域上呈现多 osn 扩散分析。我们的研究表明, 正含量是最共享的, 具有最高的病毒概率, 用户创建信息的总体病毒概率较低。最后, 我们强调了 osn 数据访问有限的问题。关键词: 在线社交网络, 病毒, 扩散, 病毒内容, 雷德特, 推特, facebook, imgur, youtube 少

2018 年 1 月 15 日提交;v1 于 2017 年 12 月 5 日提交;**最初宣布** 2017 年 12 月。

131. 零膨胀数据的归一化: 对一个新指标系列的实证分析及其在 altmetrics 数据中的使用

作者:[lutz bornmann](#), [robin haunschild](#)

摘要: 最近, 两个新的指标 (均衡归一比例引用, emnpc;提出了一种基于意义的归一化比例引用, mnp), 用于稀疏的科学计量数

据。这些指标将一个单位（例如研究人员或机构）所提到的论文（例如在**脸书**上）的比例与上述论文在相应领域和出版年份的比例（预期值）进行比较。在本研究中，我们提出了第三个指标（曼特尔–海斯泽尔商, mhq ）属于同一指标家族。 mhq 是以 mh 分析为基础的——这是统计中比较比例的既定方法。我们测试（使用同行的引用和评估, 即 $f1000$ prime 建议), 这三个指标是否可以区分根据同行的评估定义的不同质量级别。因此, 我们测试它们的收敛有效性。我们发现, 在大多数情况下, mhq 指标能够区分质量水平, 而 $mnpc$ 和 $emnpc$ 则不能区分。由于本研究中将 mhq 显示为有效指标, 因此我们将其应用于六种类型的零膨胀的高度排放度量数据, 并测试不同的 `altmetrics` 源是否与质量相关。各种标准指标的结果表明, 标准指标（维基百科、**facebook**、博客和新闻数据）与同行评估之间的关系不如同行引用和评估之间的关系强。实际上, 引用和同行评估之间的关系大约是 `altmetrics` 和对等评估之间的关系的两到三倍。少

2018 年 1 月 26 日提交;**v1** 于 2017 年 12 月 6 日提交;**最初宣布** 2017 年 12 月。

132.yelp 评论中的级联行为

作者:[muhammad raza khan](#)

摘要: 社交媒体改变了营销和消费者研究的格局, 因为企业的采用和推广越来越依赖客户如何在**脸谱**这样的平台上互动和感受企业, 推特、叶利普等像 yelp 这样的社会评论网站已经成为不同企业的重要信息来源。对这些网络平台的社会影响可能导致个人采纳或宣传导致信息级联的想法和行动。在过去几年里, 对信息级联的研究越来越受欢迎, 但大部分研究都集中在推特和**脸谱**等平台上。研究采用或推广使用级联的产品, 有助于确定社会影响的重要潜在模式。在这项工作中, 我们分析了信息的传播, 即在 yelp 的级联在欧洲和北美的不同城市。我们列举并分析了 yelp 社交网络中发生的不同级联拓扑。我们的一些重要发现包括 yelp 评论中存在大量级联, 表明社会影响的重要性、级联的重尾分布以及根据基础准确预测级联大小的可能性初步审查的报告。此外, 我们还发现, 与根节点的属性和级联的根评审相比, 非根节点和非根评审的特征是更重要的特征类型。这些发现可以帮助社会科学家以更系统的方式分析不同城市的客户行为。此外, 它还可以帮助城市的企业找出不同的消费趋势, 从而改进他们的流程和产品。少

2017 年 12 月 3 日提交;最初宣布 2017 年 12 月。

133. 基于连续时间事件的动态网络的块点处理模型

作者: [ruthwik r. junuthula](#), [maysam haghdam](#), [kevin s. xu](#), [vijay k. devabhaktuni](#)

摘要: 许多应用程序设置涉及分析一组实体之间的时间戳关系或事件, 例如在线社交网络用户之间的消息。静态和离散时间网络模型通常用作这些设置中的分析工具;但是, 它们通过随着时间的推移聚合事件以形成网络快照来丢弃大量信息。本文介绍了动态网络在不规则时间间隔内以事件形式在连续时间演化的块点过程模型 (bppm)。bppm 受已知的静态网络随机块模型 (sbm) 的启发, 是最近提出的霍克斯无限关系模型 (irm) 的一个更简单的版本。我们展示了 bppm 生成的网络在越来越多的节点的限制下遵循 sbm, 并利用此属性为 bppm 开发有效的推理过程。我们将 bppm 适合于几个真实的网络数据集, 包括一个拥有 3, 500 多个节点和 13 万个事件的 **facebook** 网络, 几个数量级比 hawkes irm 和其他现有的点进程网络模型大几个数量级。少

2017 年 11 月 29 日提交;最初宣布 2017 年 11 月。

134. 通过人群信号检测社交网络中的虚假新闻

作者: [sebastian tschiatschek](#), [atdishsingla](#), [manuel gomez rodriguez](#), [arpit 商人](#), [andedas krause](#)

摘要: 我们的工作考虑利用人群信号来检测假新闻, 并以 **facebook** 最近推出的工具为动力, 这些工具使用户能够标记假新闻。通过聚合用户的标志, 我们的目标是每天选择一小部分新闻, 将其发送给专家 (例如, 通过第三方事实调查机构), 并阻止

专家识别为假的新闻的传播。我们工作的主要目的是通过停止假新闻在网络中的传播,最大限度地减少错误信息的传播。实现这一目标尤其具有挑战性,因为它需要尽快以高度的信心检测假新闻。我们表明,为了有效地利用用户的标志,了解用户的标记准确性至关重要。我们开发了一种新的算法——检测,该算法对检测假新闻进行贝叶斯推理,并随着时间的推移共同了解用户的标记精度。我们的算法采用后采样来积极权衡开发(选择在特定时代最大限度地提高客观价值的新闻)和探索(选择信息价值最大化的新闻,以了解用户的标记精度)。我们通过广泛的实验证明了我们的方法的有效性,并展示了利用社区信号进行假新闻检测的力量。

少

2018年3月2日提交;v1于2017年11月24日提交;**最初宣布** 2017年11月。

135. 随机游走重组系统的调查

作者:[拉克拉纳特学期](#)

摘要: 推荐引擎已成为当今电子商务系统中不可或缺的组成部分。从在亚马逊推荐书籍到在脸谱等社交网络中寻找朋友,它们已经无处不在。通常,推荐系统可分为两大类:基于内容的模型和基于协同过滤的模型。这两个模型都在用户和项之间建立关系,以提供建议。基于内容的系统通过利用从可用上下文中提取的要素来

实现此任务，而协作系统则使用用户项子集之间的共享兴趣。还有另一种相对未探索的方法来提供建议，该方法利用了一个名为随机游走的随机过程。本研究是一项调查，探讨了推荐系统中随机游走的使用案例，并试图对其进行分类。少

2017 年 11 月 11 日提交;最初宣布 2017 年 11 月。

136. 几分钟内的映像网络培训

作者:杨友,赵章,谢超, 詹姆斯·德梅尔,库尔特·基乌策

摘要: 在 nvidia m40 gpu 上完成 90 世纪的成像-1k 培训, 使用 resnet-50 完成培训需要 14 天。此培训总共需要 10^{18} 个单精度操作。另一方面, 世界上目前最快的超级计算机每秒钟可以完成 2×10^{17} 单精度操作 (dongarra 等 2017 年, <https://www.top500.org/lists/2017/06/>)。如果我们能充分利用超级计算机进行 dnn 训练, 就应该能在一分钟内完成 90 年级的 resnet-50 训练。然而, 目前快速 dnn 训练的瓶颈是在算法级别。具体来说, 当前批处理大小 (例如 512) 太小, 无法有效地使用许多处理器。对于大规模的 dnn 培训, 我们专注于使用大批量数据并行同步 sgd, 而不会在固定时代失去准确性。lars 算法 (您, gitman, ginsburg, 2017, arxiv:1708.03888) 使我们能够将批处理大小扩展到非常大的大小 (例如 32k)。我们在 1024 cpu 上的 11 分钟内完成了与 alexnet 的 100 期映像网络训练。大约比

ImageNet 的结果快三倍 (goyal 等人 2017 年, arxiv:170: 170, 006, 02677), 我们在 2048 knl 20 分钟内完成了 90 年的带有 renet-50 的映像网络培训, 但不会失去准确性。使用 renet-50 的最先进的 imagenet 训练速度在 15 分钟内达到 74.9 的最高测试精度。我们在 64 个世纪获得了 74.9% 的前 1 测试精度, 只需 14 分钟。此外, 当我们将批处理大小增加到 16k 以上时, 我们在相应批次大小上的准确性远远高于 **facebook**。我们的源代码可应要求提供。少

2018 年 1 月 31 日提交;v1 于 2017 年 9 月 14 日提交;**最初宣布** 2017 年 9 月。

137. 基于采样的偏好聚合的社会网络中偏好的建模传播

作者:swapnil dhamal, rohith d.valam, y. narahari

摘要: 鉴于人口众多, 收集个人对一套替代办法的偏好, 并得出人口的总体或集体偏好, 是一项密集的任务。我们表明, 可以利用人口背后的社交网络, 通过抽样一小部分具有代表性的节点的偏好, 有效地完成这项任务。我们首先开发一个 **facebook** 应用程序来创建一个由节点和底层社交网络的首选项组成的数据集, 利用该应用程序, 我们开发了一个模型, 用于捕获偏好在典型社交网络中节点之间的分布方式。因此, 我们为选择最具代表性的节点问题提出了一个合适的客观函数。我们设计了两种算法, 即为

广泛的流行投票规则提供性能保证的 Greedy-min 和在实践中表现出优异性能的 Greedy-min 算法。我们将这些算法的性能与随机轮询和流行中心测量进行了比较, 并对得到的结果进行了详细分析。我们的分析表明, 使用社交网络信息选择代表有利于聚合与个人话题(如生活方式)相关的偏好, 而具有合理样本大小的随机轮询则足以进行聚合与社会主题(如政府政策)有关的偏好。少

2017 年 11 月 15 日提交;v1 于 2017 年 8 月 18 日提交;最初宣布 2017 年 8 月。

138. 你的明星故障: android 应用程序评论分析

作者: [rahul aralikatte](#), [giriprasad sridhara](#) , [neelamadhav gantayat](#), [senthil mani i](#)

摘要: 移动应用分发平台(如 google play store) 允许用户以评论和相应星级的形式分享他们对下载应用的反馈。通常情况下, 明星评级从一星到五星不等, 一颗星表示对应用的高度不满, 五星表示高度满意感。不幸的是, 由于各种原因, 用户提供的星级评级往往与评论中表达的意见不一致。例如, 考虑以下对 android 上 **facebook** 应用程序的审查;"真棒的应用程序"。人们可以合理地预计, 这次审查的评级是五星, 但实际评级是一星!这种不一致的评级可能会导致应用的整体平均评级降低(或膨胀), 从而影响用户下载, 因为通常用户在决定下载应用时会查看平均星级评级。

此外, 应用程序开发人员会收到有关应用程序的有偏见的反馈, 这并不代表地面现实。这对于有几千次下载的小应用尤其重要, 因为即使是少数不匹配的评论, 也会大幅降低平均评级。本文对这一评审等级不匹配问题进行了研究。我们手动检查了来自 10 个流行 android 应用的 8600 条评论, 发现我们数据集中 20% 的评级与评论不一致。此外, 我们还开发了三个系统;其中两项是基于传统的机器学习, 一个是基于深度学习, 以自动识别评级与评论中表达的意见不匹配的评论。我们的深度学习系统表现最好, 在确定与特定评价相关的正确星级率方面的准确性为 92%。少

2018 年 8 月 11 日提交;v1 于 2017 年 8 月 16 日提交;**最初宣布** 2017 年 8 月。

139. 谁将分享我的形象? 预测网络社交网络中的内容传播路径

作者:胡文健,克里希纳·库马尔·辛格,肖范义,金英韩金英,陈妮夸,李勇

摘要: 内容受欢迎程度预测由于其**对脸谱**、instagam、推特和 pinterest 等社交媒体网站的用户和主机的重要性的兴趣而得到了广泛的研究。然而, 现有的工作主要集中在使用单一指标(如喜欢或股票的总数)建模受欢迎程度上。在这项工作中, 我们提出了分布式 lstm, 一个基于记忆的深循环网络, 学习递归预测图像的整个扩散路径通过一个社交网络。通过结合用户的社会特征和图

像特征, 并使用显式存储单元对迄今所采取的扩散路径进行编码, 我们的模型比只对图像编码的备用基线更准确地预测图像的扩散路径。社会特征, 或缺乏记忆。通过将单个用户映射到用户原型, 我们的模型可以推广到培训期间看不到的新用户。最后, 我们证明了模型生成扩散树的能力, 并证明了生成的树与地面真理树非常相似。少

2017年11月29日提交;v1 于 2017年5月25日提交;**最初宣布** 2017年5月。

140. 自动进行弱内存程序的演绎验证

作者:[亚历山大·萨默斯](#),[彼得·穆勒](#)

摘要: 为弱内存模型 (如 c11 内存模型) 编写正确的程序具有挑战性, 因为这些模型提供的弱一致性保证。最近提出了验证此类程序的第一个程序逻辑, 但迄今为止, 它们的使用仅限于手动证明。由于推理功能 (如高阶断言、模式和丰富的权限资源), 通过一阶求解器在这些逻辑中自动进行证明并非微不足道。在本文中, 我们提供了一个使用现有的演绎验证工具的弱内存程序逻辑的第一个实现。我们处理了最近的三个程序逻辑: 宽松分离逻辑和两种形式的围栏分离逻辑, 并展示如何使用 viper 验证基础结构对这些逻辑进行编码。在此过程中, 我们演示了几种可用于其他逻

辑的新编码技术。我们的工作得到了实施, 并根据现有报纸以及脸谱开源文件夹图书馆的例子进行了评价。少

2018 年 2 月 19 日提交;v1 于 2017 年 3 月 18 日提交;最初宣布 2017 年 3 月。

141. 脸谱的使用、个性特征和学习成绩: 一项相关研究

作者:[georgia sapsani](#), [ni 油奥斯·塔塞柳斯](#)

摘要: 本文考察了学生个性、社交媒体使用与学习成绩和敬业度的关系。具体而言, 本研究的目的是利用五大个性测试, 考察学生脸谱(fb) 的使用和个性特征与 (a) 学生参与、(b) 为上课做准备的时间、(c) 花在课堂上的时间之间的关系。课外活动和 (d) 学习成绩。结果表明, fb 时间与学生参与度和备课时间呈显著正相关。最后, 一些脸谱活动与学生的个性特征呈显著正相关, 还有一些与研究的每个因变量呈显著负相关。并对这项研究对教师和学生影响进行了讨论。少

2018 年 6 月 19 日提交;v1 于 2017 年 3 月 12 日提交;最初宣布 2017 年 3 月。

142. 从子空间构造的数组代码的位置和可用性

作者:[natalia silberstein](#), [tuvi etzion](#),[moshe schwartz](#)

摘要: 在谷歌、脸谱和亚马逊等互联网规模的公司中, 创建和处理的数据数量不断增加。因此, 如此大量数据的有效存储已成为现代计算中的一个根本问题和尖锐问题。没有一台机器能够满足如此巨大的存储需求。因此, 依赖于数万个存储节点的分布式存储系统 (dss) 是唯一可行的解决方案。这类系统在所有现代互联网规模系统中都得到了广泛的应用。然而, dss 的设计带来了一些关键的挑战, 与单用户存储系统明显不同。此类系统必须能够有效地重建数据, 克服服务器故障, 纠正错误等。在过去几年里, 为应对这些挑战进行了大量的研究, 随着存储数据量的增加, 研究也在不断增加。本文的主要目的是考虑具有分布式存储系统两个最重要特征的代码, 即局部性和可用性。我们的代码是阵列代码, 它基于有限域上线性空间的子空间。我们提出了几个这样的代码的结构, 这是问-类似于一些已知的块代码。其中一些守则具有独立的智力价值。我们检查构造代码的位置和可用性。特别是我们区分了两种类型的局部性和可用性, 节点与 \sim 符号、局部性和可用性。据我们所知, 这是文学中第一次给予这样的区分。少

2018年6月23日提交;v1 于 2017年1月25日提交;**最初宣布** 2017年1月。

143.我们真的需要把他们都抓起来吗? 一种新的用户引导社交媒体爬网方法

作者: [弗雷德里克·erlandsson](#), [piotr bródka](#), [martin boldt](#), [henric johnson](#)

摘要: 随着脸谱和推特等流行社交媒体服务的使用日益增多, 在无法进入核心基础设施或支付费用的情况下, 从网络中收集所有内容是一项挑战。因此, 如果无法收集所有内容, 必须考虑哪些数据最重要。在这项工作中, 我们提出了一个新的用户引导的社交媒体爬网方法 (usmc), 它能够从社交媒体收集数据, 利用人群的智慧来决定收集用户生成的内容的顺序, 以覆盖尽可能多的用户交互, 因为可能。usmc 通过对 160 个公共 **facebook** 页面进行爬网进行验证, 其中包含来自 3.68 亿用户的内容, 其中包括 13 亿次交互, 并与其他两种爬网方法进行了比较。结果表明, 通过对其 20% 的帖子进行采样, 可以覆盖 **facebook** 页面上约 75% 的互动, 同时将抓取时间缩短 53%。此外, 与所有帖子创建的社交网络相比, 20% 样本构建的社交网络包含 75% 以上的用户和边缘, 并且具有类似的程度分布。少

2017 年 12 月 13 日提交;v1 于 2016 年 12 月 6 日提交;**最初宣布** 2016 年 12 月。

144. 情绪分析的演变——研究主题、场地和热门论文的回顾

作者:[mika viking mäntylä](#), [daniel graziotin](#), [miikka kuutila](#)

摘要:和..。近年来, 情绪分析已从分析在线产品评论转向来自推特和脸谱的社交媒体文本。除了产品评论之外, 许多主题都能扩展情绪分析的使用范围.

2017 年 11 月 21 日提交;v1 于 2016 年 12 月 5 日提交;**最初宣布** 2016 年 12 月。

145. 矩阵网络的卷积归因

作者:孙庆云,严大伟,斯蒂芬·博伊德

摘要: 矩阵网络是矩阵的一个家族, 由加权图建模相关性。我们考虑完成部分观测矩阵网络的任务。我们假设了一个新的采样方案, 其中一小部分矩阵可能是完全没有观测到的。我们如何才能从不完全观测中恢复整个矩阵网络? 这个数学问题出现在许多应用中, 包括医学成像和社交网络。为了恢复矩阵网络, 我们提出了矩阵具有低阶的图傅立叶变换的结构假设。我们提出了一个凸优化问题, 并证明了优化问题的精确恢复保证。此外, 我们还对不同等级和采样率的精确恢复系统进行了数值描述, 发现了一种新的相变现象。然后给出了迭代归因算法, 有效地解决了优化问题, 完成了大规模矩阵网络。我们用 mri 和 facebook 用户网络等多种应用演示了该算法。少

2018 年 6 月 7 日提交;v1 于 2016 年 6 月 2 日提交;**最初宣布** 2016 年 6 月。

146.焦虑、酗酒和学术：学生脸谱自白页面的考证

作者:soubhik barari

摘要: 大学生在完全匿名的情况下，在社交媒体上向同龄人透露了什么？他们的校园环境是否与他们的披露主题有关？为了回答这些问题，我分析 **facebook** 的告白页面。这些网页在数以百计的大学校园中很受欢迎，允许学生在公共社区论坛上匿名发布个人供词。在这一初步的研究报告中，我分析了网上学生忏悔行为的几个解释因素。我将跨越 3 年的近 200,000 个告白帖子结合起来，通过机械土耳其人将潜在的 dirichlet 分配 (lda) 与人的验证结合起来，以可缩放地识别这些在线告白中的主题。在可能的情况下，我还将帖子链接到从推特上分析的现实世界新闻事件。我发现，提到社会经济学以及身心健康的供词更多发生在一流、昂贵的民办高校。虽然与事件有关的供词往往提到及时的与学校有关的活动，但许多人提到了当地校园范围以外的全球和国内事件。结果显示，来自不同校区的本科生披露的主题不同，如种族、社会经济学和政治，但总体上，随着时间的推移，以类似的模式发布。此外，研究结果证实，匿名 **facebook** 告白者在健康和社会经济地位等重要但禁忌的话题上得到了供词的支持。少

2018 年 8 月 19 日提交;v1 于 2015 年 6 月 16 日提交;**最初宣布** 2015 年 6 月。