信息网络专题研究阅读笔记之应用层

一、文献信息

1. 论文题目：Women, politics and Twitter:Using machine learning to change the discourse
2. 论文作者：Lana Cuthberston，Alex Kearney，Riley Dawson，Ashia Zawaduk，Eve Cuthbertson，Ann Gordon-Tighe，Kory W Mathewson
3. 发表途径：AI for Social Good workshop at NeurIPS (2019), Vancouver, Canada.
4. 发表时间：not mentioned

二、问题意义

1. 研究背景

在加拿大的政治制度中，各级民选政府都存在性别不平等。从事政治的妇女，长期饱受网上恶意推文的困扰。针对性别的辱骂性质的文章，会不断助长这种不平等，且大部分从政女性对于网络骚扰问题持不作为态度。

2. 研究问题

对于社交媒体推特上存在的网络骚扰问题，我们希望利用一个基于人工智能的“ParityBOT”Twitter 机器人进行正面干预。本文主要研究，如何训练 ParityBOT 进行文本分析，进而可以对推文分类，有针对性对恶意推文进行“积极推文响应”，提高政治话语。

3. 研究意义

本文主要提供了一个可扩展的模型，用定量和定性评估来分类和响应恶意推文。利用机器学习技术解决妇女在政治制度中所面临的系统性问题。积极改善女性在政治中受到的不平等对待，削弱其在 Twitter 上公平参与政治的障碍，将科学进步与人类进步紧密联系起来。

三、思路方法

1. 研究方法

研究方法概述为：设计→构建→部署→反馈→再评估→改进模型→明确下阶段研究重点。作者将 ParityBOT 分类系统在公共网络骚扰数据集上进行验证，在真实政治选举中部署并在干预期间收集数据，对 ParityBOT 的影响做了进一步分析。

2. 研究思路

1) ParityBOT 实施平台：在作为重要的社交媒体平台的推特上进行实施。政治家们在推特上会进行政治愿景的分享并与选民互动，而女性参政者在这个平台上受到了“倾倒性”的恶意诋毁。

2) ParityBOT 机器人设计：如下图 1 所示的框图形式来直观展现。

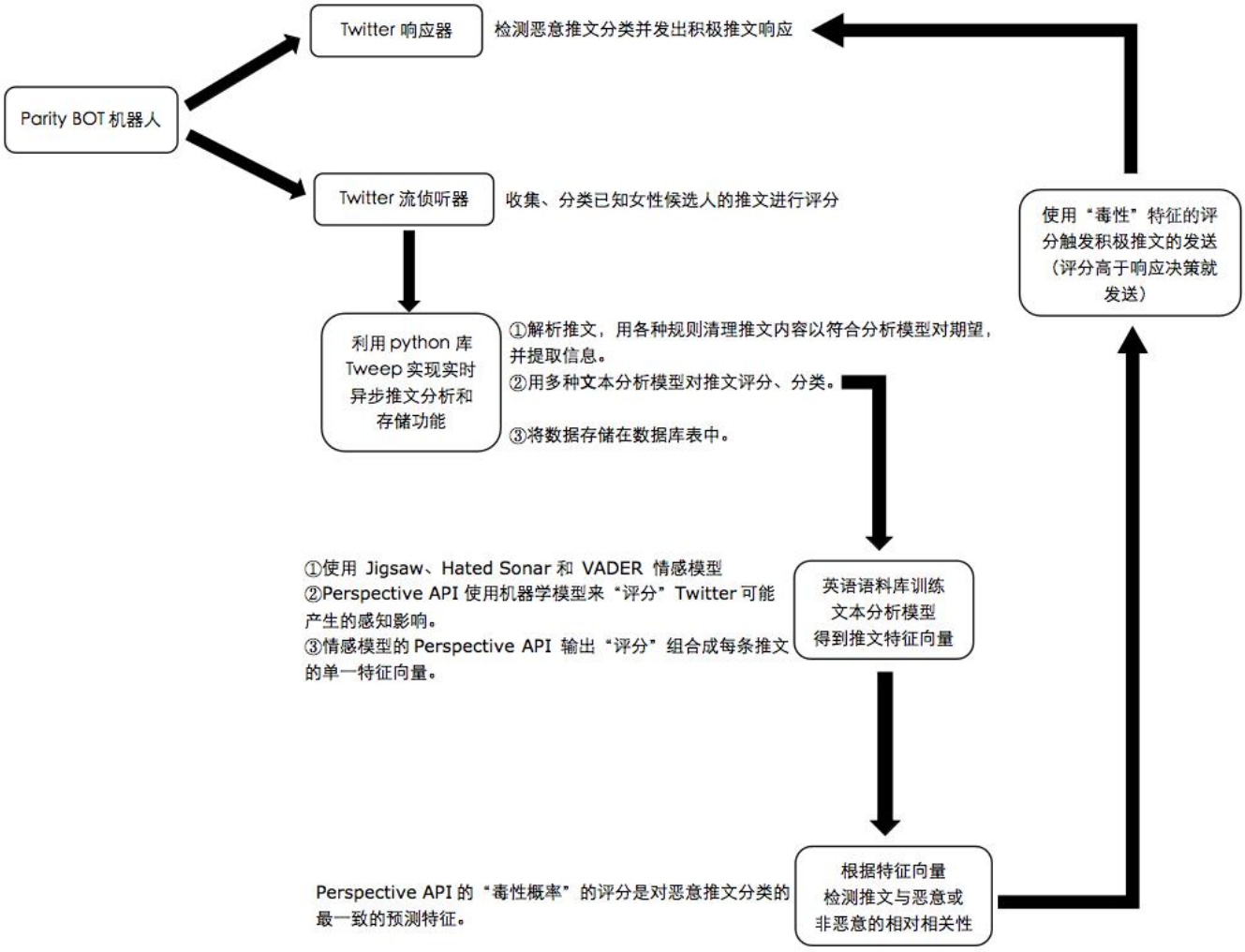

图 1 Parity BOT 技术细节

- 推特清理：tweet 的清理主要使用正则表达式规则，其允许训练、验证和测试数据集之间的一致性和泛化。
- 推特特征化：每条推特文本与单一特征向量（17 个 Perspective API，3 个来自 Hate Sonar 模型，4 个来自 VADER 模型）相关联。
- 验证和消融实验：特征化数据集中的每个条目由 24 个特征和一个恶意或非恶意的类标签组成。特征化数据集被随机分成与完整数据集的类平衡相匹配的训练集和测试集。并使用 ADASYN 来重采样和平衡数据集中的类别比例。通过消融实验（Ablation Experiments）测试从各种文本分类模型中导出的特征相对影响。

3) 收集 Twitter 处理信息、预测性别：在部署 ParityBOT 时需要使用在线资源创建选举中所有候选人的数据库。通过使用 Python 库中的“性别猜测”来根据候选人姓名预测性别，并手工验证；ParityBOT 发送的积极推文由志愿者提交在线表格提供，且所有志愿者未经筛选，任何人都能够访问 Positivitweet 提交表格。

4) ParityBOT 项目评估：作者设计了一个基于用户体验研究访谈标准的讨论指南。参与访谈人员对 ParityBOT 有不同程度的预先认知。通过与参与者的电话访谈，获取以下信息：ParityBOT 对于妇女参政的影响、与 ParityBOT 互动过的用户的反馈、没有与 ParityBOT 互动的用户的初步印象、ParityBOT 潜在发展平台。

四、实验结论

1) 实际部署成果：在 2019 念艾伯塔省选举、2019 年加拿大联邦选举中部署了 ParityBOT。根据不同情况来设置不同的响应决策阈值 (毒性特征评分)，高于阈值则发送积极推特。在具体的定量分析后，可以发现 ParityBOT 并没有对每一条被分类的恶意推文都发送一个积极回应，反映了能够人为限制 ParityBOT 的决策率。
2) 价值和局限性：作者为 ParityBOT 编写了指导方针来支持 BOT 项目的持续发展。
   - ParityBOT 对于推特的分类可能产生错误，但是能通过选择决策阈值来减轻其影响。
   - ParityBOT 的发展在社交媒体和政治应用中存在一定风险。
   - 在收集 Twitter 处理数据确定候选人性别方面存在限制。
3) 用户体验成果：通过对 BOT 研究方案的评价和讨论，发现利用机器学习技术的 ParityBOT 在改变话语中确实起着一定的作用，也鼓励了更多样化的候选人参与政治；但 ParityBOT 也存在缺点，会带来一定的负面性，使与会者因此而注意到那些恶意推文。

五、启发思考

这次的专题学习和论文阅读，主要围绕“机器学习”展开。因为我之前对于机器学习的认知一直都停留在“耳熟能详”，对具体实现和应用细节并不清晰。在论文学习和探索的过程中，会遇到很多生僻的概念。通过查找资料可以加深理解，有时候搜索的一个小概念背后就能够衍生出一类经典的机器学习算法。因为本篇论文主要是基于人工智能对于社会发展的推动作用，不难发现它带给我们的益处是方方面面的。技术不能仅停留在数据和实验里，真正应用到生活中才是最好的研究成果。机器学习的部署不需要过多的人为干预，只要建立起训练模型，就能够使其自行在数据库中进行迭代收敛。通常还会加入一些测试数据集来检验模型的训练成果，然后作出进一步的调试以得到良好的性能输出。

作为通信专业的同学，我感觉以前对于这种热门技术的关注是不够的。老师给我们推荐了很多机器学习开放资源的网站，可以通过有选择性的筛选和学习，拓宽自己的专业眼界。以前总会认为机器学习很抽象，但是真的接触后，不管是从小的训练模型的建立，还是到庞大的神经元网络的设计，都是非常有乐趣的。自己作为一个 Python 的初学者虽然还不能很好的编写基于机器学习的代码，但在跑完一些基础源码后，可以理解其中的核心思想。通过一点点提升代码储备，我相信以后肯定会越来越驾轻就熟。在日后的专业学习当中，我也会不断关注机器学习方面的前沿发展，使自己具备一个通信人应该有的对前沿技术的热忱和兴趣!

【注】：本论文没有代码，机器学习代码运行的结果将在大作业中体现。