模式识别实验报告

社交媒体机器人识别

符兴 7203610316

1 选题背景

随着社交媒体的迅速发展，人们的信息获取和交流方式发生了巨大的变化。许多大型社交媒体平台中成为了许多人获取新闻、表达观点和与他人互动的重要渠道。然而，随着这些社交媒体平台的用户数量不断增加，其中也涌现出了大量的机器人账户，也被称为“社交媒体机器人”或“社交媒体僵尸”，它们被自动化脚本控制，具有自动发布、转发和评论的能力，可以在社交媒体上大量发布操纵者预设的内容。

社交媒体机器人往往被用于传播虚假信息和恶意内容，它们可以迅速扩散谣言、假新闻和仇恨言论，给公众带来误导和混乱。此外，社交媒体机器人还可能被用于营销目的，通过盲目灌水、发帖来达到扩大产品或服务曝光度的目的，这一行为破坏了市场的公平竞争；同时，用户也被此类毫无营养的信息淹没，严重影响了用户的正常体验。

因此，识别和应对社交媒体机器人的问题显得尤为重要。然而，由于技术的不断发展，社交媒体机器人的操纵者们也越来越善于隐藏它们的真实身份。许多社交媒体机器人具有高度智能化的特征，其可以通过先进的自然语言处理技术生成更加真实的文本，这也使得传统基于规则或简单机器学习方法在社交媒体机器人识别中效果不尽如人意，因此需要更深入地挖掘机器人和人类之间不同的行为特征和文本特征。

近年来，学术界和工业界都进行了一系列相关研究。研究人员尝试使用各种技术和方法，包括深度学习、文本挖掘和社交网络分析等，以区分真实用户和机器人账户。本文在前人工作的基础上，进一步针对社交媒体的网络结构特征，充分利用社交媒体用户的行为特征和语言特征，提出一种基于 GCN 的社交媒体机器人识别模型，并且该模型在 2022 人民网智能挑战赛的赛道五中获得 0.8256 的分数。

2 研究内容

2.1 数据分析

该任务是一个二分类任务，最直观的想法就是构建表示用户的特征向量，然后用一个二分类器对该特征向量进行分类。在构建用户的特征向量时，如果机器人和正常用户的特征向量在整个向量空间中具有明显的分布差异时，分类器就能够很好地将其进行分类。因此，需要选择那些机器人和人类具有分布差异的信息去构建用户的特征向量。

在官方提供的数据集中，每个用户共有 16 种的用户信息，如表 1 所示。

表 1 用户信息类别

ID	followers_sount	default_profile
tweet	friends_count	geo_enabled
name	listed_count	contributors_enabled
screen_name	created_at	default_profile_image
location	favourites_count	protected
description	verified	statueds_count

对每个信息类别进行统计，可以发现在如下信息类别中，机器人和人类具有较为明显的分布差异。

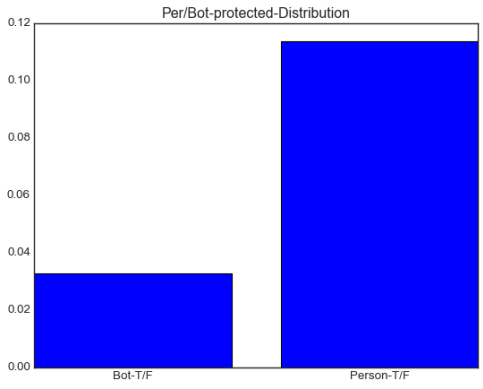

图 1a 在 Protected 上人/机器的分布差异

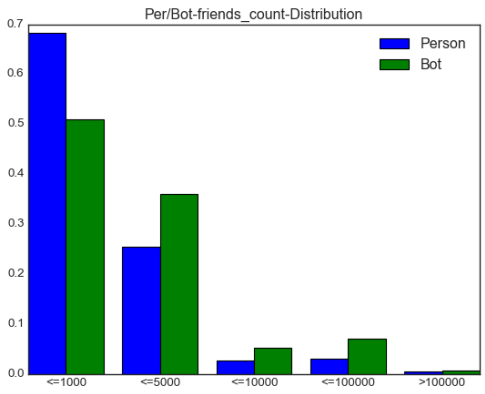

图 1b 在关注数上人/机器的分布差异

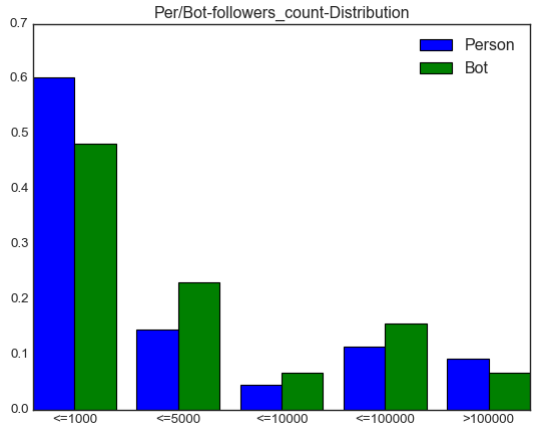

图 1c 在粉丝数上人/机器的分布差异

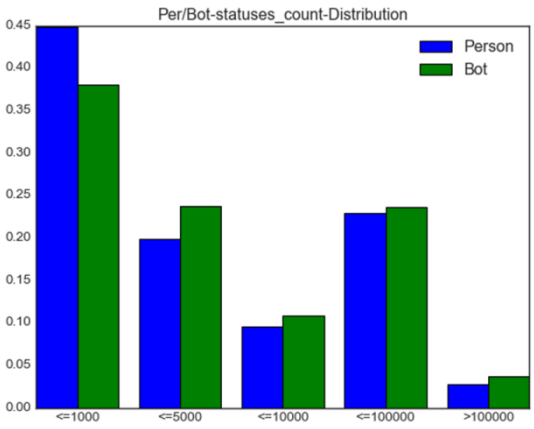

图 1d 在发帖数上人/机器的分布差异

从上面的数据我们可以发现，在某些方面，机器人和人类具有较为明显的差异；如在关注数上，正常用户关注数<10000 人的比例高达 90%，而机器人关注数在>10000 人的区间上仍占有不小的比例；同样地，在粉丝数和发帖数等信息类别上，人类和机器人都有着显著地差别，即机器人账户会关注非常多地账号，同时其发帖量也非常大，这一特点十分贴合机器人灌水、大量输出信息的目的。

通过可分性判据分析后，最终选出了参与构建用户特征向量的信息类别，如表 2 所示。

表 2 选中的用户信息类别

ID	**followers_sount**	default_profile
tweet	**friends_count**	geo_enabled
name	**listed_count**	contributors_enabled
screen_name	created_at	default_profile_image
location	**favourites_count**	**protected**
description	**verified**	**statueds_count**

2.2 数据处理

在 2.1 节选中的信息类别中，Tweet、Description 是用户的文本信息，需要对其进行编码。在本文中，选用 RoBERTa 模型构建文本特征向量。构建用户初始特征向量的具体步骤为：

(1) 判断用户的推文数量是否达到选用阈值；由于数据集中每个用户的推文数量分布差异较大，在训练集中会舍弃掉推文数量较少的用户；

(2) 由于数据集中的文本是直接抓取社交媒体上的文本，其中充斥了许多预训练模型 RoBERTa 不能编码的字符，如颜文字、Emoji 符号、短连接等等，在这些字符上可以根据需求做出不同的策略；例如删掉短连接，替换 Emoji 符号为相关情绪词符号等等，后续会具体阐述。

(3) 编码用户的推文特征时，首先构建出每一条推文的特征，最后取所有特征的平均值作为用户的总推文特征向量。用户的 Description 特征单独为一个特征向量。

(4) 在选中的信息中，有些信息是 0、1 分布的，如 Protected、Verified 信息；还有些信息不是 0、1 分布的，如粉丝数、关注数等，需要将他们拆开为两个特征向量；后者需要对其数值进行归一化操作，然后再填入相应地信息槽中。

(5) 综上 4 个部分的特征向量在经过全连接层后得到用户的初始特征向量。

2.3 模型结构

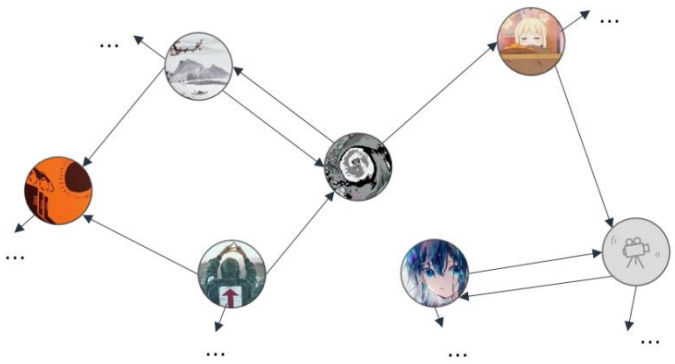

图 2 社交网络结构

如图 2 所示，社交网络是一个图结构，在社交网络中，人与人之间都有一定的联系，例如所关注人和自己有相同的喜好，所关注的人在现实生活中和自己有所联系等等，因此可以通过图卷积神经网络 GCN 去捕获用户和用户之间的行为特征。

GCN 相较于 GNN 而言，在聚合方式上有较大差异，具体体现在 GCN 的邻接矩阵加上了单位对角阵；GCN 的度矩阵对行和列进行归一化的操作，行归一化系数代表节点自身的一个变化程度，关联的节点越少，系数越大，越容易随波主流，更易受关联节点的影响。而列归一化系数，代表关联节点对当前节点的影

响程度，关系网越复杂的节点，它对其他节点的作用就越小。

如图 3 所示，GCN 中的每个节点经过广播、接收、变换后得到当前节点的特征向量表示，然后经过分类器对其进行分类。

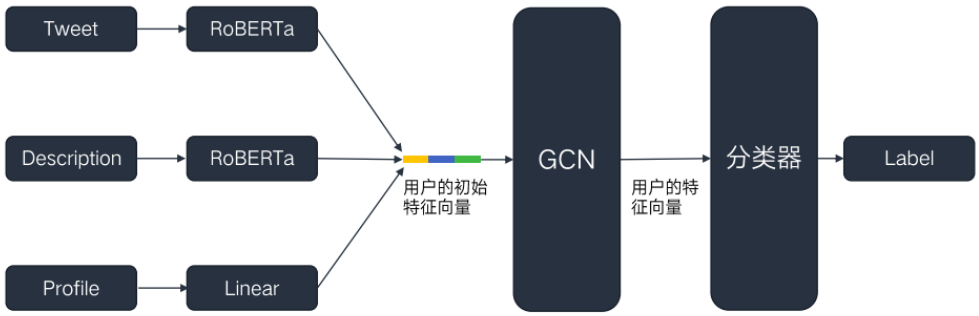

图 3 模型结构

3 实验结果

3.1 Emoji 符号转换

在 2.2 介绍的构建用户初始特征向量步骤中，在提取用户推文特征时，受限于预训练模型的编码字符，推文中表示情绪的 Emoji 符号不能够被正确编码，因此尝试将 Emoji 符号解码为相对应的英语表示，如图 4 所示；然后使用 RoBERTa 对整个推文进行编码。

但是，模型整体的 F1 值有所下降，反复多次均为超过不加 Emoji 解码的模型。根据对数据集的分析，猜测可能由于在社交媒体上，Emoji 符号的具体含义往往和其对应的英文表示有所出入，例如人们在使用“鼓掌”的 Emoji 时除了表示对事物的赞扬，也可能表示人们对某件事情的落井下石，亦或是商家文案最后所添加的一系列表情符号，但实质上所表示的语义信息并不充足，模型难以学习到此类信息；

```
{ home 桌面 Pattern_Recognization lab5 } python
Python 3.9.16 (main, Mar  8 2023, 14:00:05)
[GCC 11.2.0] :: Anaconda, Inc. on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import emoji
>>> print(emoji.demojize('😀'))
:grinning_face:
>>>
```

图 4 Emoji 符号解码

3.2 数据采样

在 2.2 节中介绍的具体步骤中，会将推文数量较少的用户从训练集中剔除，将其放入验证集，我们期望模型能够去学习到更具有一般性的特征，而不是某些特例。除此之外，尽管在数据集中机器和人类的样本分布大约为 1:4，但是在现实生活中，机器人的数量相对于正常用户数量而言是较少的，因此在构建验证集时会有意扩大这个分布比例，使得验证集能够更好地反应模型参数带来的影响。

3.3 不同图卷积模型、不同的预训练模型编码的实验结果对比

表 3 实验结果

模型	F 值
BERT + GCN	0.7851
ROBERTA + GCN	0.8040
ROBERTA + GCN + 数据采样	**0.8256**

最终名次	队伍	分数
冠军	603	0.8472
亚军	gogoHH	0.8296
季军	今年想去春茧咧	0.8256
	吃个面包	0.8203
	你先别急	0.8002
	猛追湾三道拐work-life balance	0.7974
	不如吃茶去	0.7476
	ZZZero	0.7112

图 5 模型测试结果

4 总结

本文阐述了基于 GCN 的社交媒体机器人识别模型，其利用图卷积神经网络捕获用户的行为特征，并通过分类器对用户进行分类。未来，还可以使用 Attention 机制去捕获用户文本中的情感信息，进一步挖掘用户之间的互动行为对情绪的传播，从而更好地区分社交媒体机器人和正常用户。

参考文献

[1] Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. InProceedings of the 26th ACM SIGKDD international conference on knowledge discovery & datamining, pages 492–502, 2020.

[2] Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. Fake news detection via knowledge-driven multimodal graph convolutional networks. In Proceedings of the 2020 International Conference on Multimedia Retrieval, pages 540–547, 2020.

[3] Yi-Ju Lu and Cheng-Te Li. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. In Proceedings of the 58th Annual Meeting of the Association forComputational Linguistics, pages 505–514, 2020.

[4] Philip N Howard, Bence Kollanyi, and Samuel Woolley. Bots and automation over twitter during the us election. Computational propaganda project: Working paper series, 21(8), 2016.

[5] Samantha Bradshaw, Bence Kollanyi, Clementine Desigaud, and Gillian Bolsover. Junk news and bots during the french presidential election: What are french voters sharing over twitter? Technical report, COMPROP Data Memo, 2017.