面向新浪微博的社交网络分析

代江海
Beijing Institute Of Technology
School of Computer
3120180985@bit.edu.cn

尚伟
Beijing Institute Of Technology
School of Computer
3220180733@bit.edu.cn

白雪峰
Beijing Institute Of Technology
School of Computer
3120180977@bit.edu.cn

卞西墨
Beijing Institute Of Technology
School of Computer
3120180978@bit.edu.cn

沙九
Beijing Institute Of Technology
School of Computer
3120181023@bit.edu.cn

张昱霖
Beijing Institute Of Technology
School of Computer
3220180771@bit.edu.cn

1 INTRODUCTION

微博，即 microblog，提倡的是随时随地，无处不在的沟通。简单地说，微博客就是因特网用户可以通过短消息（新浪微博为 140 汉字，长微博除外）的形式描述自己感兴趣的事物，表达态度，并与他人分享，讨论。用户可以通过微博融合多种渠道(包括即时通讯、手机、Email、网页等)发布文字、图片、视频、音频等形式的信息。

2012 年起，随着移动互联网的普及，加深了互联网世界的信息爆炸现象的问题。根据中国互联网信息中心（CNNIC）发布的统计报告显示，我国移动互联网普及率已达到 74.1%，智能手机已经成为网民上网的主要终端。网民使用互联网的主要诉求是获取，分享比传统媒体更多的各类信息及碎片化阅读，还有对自己想法的即时表达。因此带动了微博的快速发展。纵观近些年来在微博中经过大量用户讨论而衍生出的重大社会新闻和热门讨论话题，因其所具备的开放性、便捷性及互动性，微博服务正在不断改变社会中的话语体系，帮助打破了传统的新闻媒体与舆论格局，使得话语权从社会精英阶层走向普通网民。目前，国内影响力较大的包括新浪微博、腾讯微博、网易微博、搜狐微博等，上述微博中，2009 年 8 月开始内测的新浪微博是中国最具影响力的微博之一。因此，我们从新浪微博中选择用户样本，分析微博用户的行为特征。

国外在这方面的研究主要有：Qiang，Yan 等[1]通过分析微博客用户发布消息的行为，提出了一个以用户兴趣及社会认证为驱动的分析模型，得到用户自身利益影响微博客用户发布信息的时间间隔等结论；Bruno Goncalve 等人[2]收集了连续 6 个月内 Twitter1700 万个用户的交流数据集，并基于邓巴数字（Dunbar's Number）测试的理论认知稳定社会关系的数量限制，发现用户可以有 100-200 的稳定关系。

国内有关这方面的研究主要有：王晓光[3]以“新浪微博” 为研究样本，较为系统地研究微博客的基本结构、信息传播一般模式，考察微博客用户基本行为特征和关系特征，分析微博客影响力的相关变量，并建立了粉丝数与关注数、博文数回归方程；袁毅和杨成明[4]跟踪微博客用户在某时间周期内关于某一话题的交流数据，发现用户在信息交流过程中形成关注、评论、转发和引用四种社会关系网络，指出四种关系网络有其不同的结构形态，但同时又具有某些共性特征及联系；杨成明[5]以新浪微博客为研究对象，抽取微博客平台提供的各项字段，以统计领域方法从用户性别、地域、影响力等多个角度揭示当前微博客用户的行为特征及存在的问题。平亮、宗利永则基于社会网络理论，结合微博用户之间的“关注”与“被关注”信息传播的网络拓扑关系，从点度中心性、中间中心性和接近中心性 3 个方面对微博社会网络的中心性进行分析[6]。赵文兵、朱庆华等[7]以国内财经网站和讯微博为例，使用计量学方法，对用户特性进行统计分析，并使用可视化软件 pajek 进行可视化分析，并对微博用户进行了分类。

综上，可以看出，国外有关微博客用户行为特征的研究还是要领先于国内研究的。国外研究者主要都在已有模式的基础上加上更加符合微博用户研究的变量或规范，从而验证自己建立的新模式。国内的研究主要集中于通过利用统计学和社会网络的方法和理论，揭示微博

客用户的行为特征及影响因素。国内有关微博客的研究起步较晚，而关于微博客用户的研究更是才开始于这两年，所以存在如下方面的不足：以新浪微博为例，大多数文献抽取的样本数量为几千条，相对这样大的用户总体，样本容量略显不足；国内微博客产生时间大多集中在 2009 年前后，且有关研究成果基本都只从一个方面去研究，缺乏综合性的研究成果；定性研究较多，定量研究较少。因此，本文将以“新浪微博” 用户数据为例，扩大采集的数据集的容量，从统计学、社会网络分析、数据挖掘和情感分析等多角度对微博客用户行为特征及分类进行更加深入、系统地分析。

新浪微博用户群体十分广泛，涉及影视明星、文化名人、企业高管、网络红人、普通大众等各个社会群体。总体而言，新浪微博的用户门槛比较低，有着一定的“草根化”特征。大部分用户都是“草根化”的个人用户。这些个人用户是新浪微博用户的“主力军”，覆盖面广、影响力大。具体而言，新浪微博的个人用户呈现以下几点特征：

（1）以年轻人为主。新浪微博用户中，以 19~39 岁的年轻用户为主体，也就是我们所说的“80 后”“90 后”。处于这个年龄阶段的人热衷于追求各种新鲜事物，接受新事物能力比较强，个人自主意识强烈，喜欢赶潮流，对“互联网+”时代背景下的各种新兴事物大力追捧。

（2）女性用户倾诉欲强烈。新浪微博用户中，女性数量以及活跃度明显高于男性。这是由整个社会环境以及女性自身的性格特征造成的。长期以来，女性在男权社会的话语体系里缺乏话语权，强烈的倾诉欲望被压抑。新浪微博为女性“发牢骚”“表心情”提供了一个平台，她们可以随意宣泄倾诉欲望。

（3）中高等教育用户较多。新浪微博用户的中坚力量是中高等教育用户，比如大学生、白领阶层，他们日常的工作、学习压力比较大，没有时间写博客、看新闻，而新浪微博碎片式的信息形式更加契合他们的生活方式。

2 Object

本项目的旨在对所采集的微博用户进行画像式的分析，对于用户在微博上的行为进行多维度及深层次的研究。在使用微博服务的过程总，微博用户的使用行为主要包括：关注、评论与转发。在微博平台中，用户通过关注其他用户或者被其他用户关注从而建立了属于自己的微博社交关系，并且通过关注其他的用户来获取更多感兴趣的微博信息。同时，通过评论与转发其他用户的微博内容主动参与微博信息的传播及衍生过程。目前，我们针对微博用户的研究主要以这三种行为为出发点，来对微博用户的个人信息（包括关注对象和粉丝），评论转发的行为及方向频率等特征，以及针对微博内容和评论转发内容的进行的情感分析，还有针对根据关注与被关注，转发评论所构建起来的关系网进行的网络分析。

我们使用新浪微博官方提供的 Weibo API 2.0 接口，该接口有针对爬取频率和可使用函数的限制，更高级的函数功能需要向新浪官方进行申请。普通的爬虫接口已经可以满足我们数据分析的需要了。我们使用了两个接口来完成数据采集的工作。两个接口返回的内容不同，可以用来互补。两个接口获取到的内容如下面两张图所示。


```
{ "_id" : NumberLong(5180044582),
   "screen_name" : "山是山",
   "validation" : "知名情感博主",
   "gender" : "男",
   "location" : "江苏",
   "birthday" : "1995-11-26",
   "other" : "知名情感博主",
   "description" : "谁让我拥抱 再一次心动",
   "statuses_count" : 2220,
   "followers_count" : 1269855,
   "follow_count" : 429,
   "verified" : true, "verified_type" : "谁让我拥抱 再一次心动",
   "verified_reason" : "谁让我拥抱 再一次心动",
   "follower" : [],
   "following" : []
```

```
{ "_id" : NumberLong(4362759396998816),
   "pid" : NumberLong(4362756427643383),
   "created_at" : "Fri Apr 19 11:59:19 +0800 2019",
   "text" : "其实现在广东很多人都会说普通话了，尤其是学生之类的说的比较好。而且甚至有的广东小孩子不会说粤语，只会说普通话",
   "uid" : NumberLong(2316045594),
   "like_count" : 703,
   "comments" : [
      { "created_at" : "Fri Apr 19 13:04:22 +0800 2019",
        "text" : "广东不止粤语，还有很多方言，甚至同一种方言又有几种口音，靠近珠三角中心的几个城市的人会三种语音，最精通的就是方言，然后是粤语，最后反而普通话不标准，远离珠三角的几个城市，潮汕地区的人说普通话和客家话，粤语就比较差了，最后就是广州附近的几个城市大多就是粤语精通和二流普通话",
        "uid" : NumberLong(2285132374)
      }]
```


图 2-1 用户数据格式（左），评论数据格式（右）


```
{
   "_id" : NumberLong(4362756427643383),
   "uid" : NumberLong(5180044582),
   "created_at" : "04-19",
   "text" : " #广东人认为自己普通话标准#广东人说普通话时的口音，具体表现是什么，哈哈哈哈哈，真的好好笑 ",
   "source" : "",
   "reposts_count" : 1180,
   "comments_count" : 1315,
   "attitudes_count" : 3458,
   "isLongText" : false,
   "page_info" : {
      "page_pic" : {
         "url" :
"https://wx1.sinaimg.cn/large/7085f3c4ly1g27qtatudij20u00u0god.jpg"
      },
      "page_title" : "#广东人认为自己普通话标准#",
      "content1" : "“你是广东人吧？”\n“里怎么走道？”\n“你是东北人吧？”\n“你咋直道的呢？”\n“你是四川人吧”\n“你哪个晓得勒”",
      "content2" : "3.3万讨论  1.7亿阅读",
      "type" : "topic",
      "object_id" :
"1022:23152232233d45877ca8e9891220ce960e4223"
   }
}
```


图 2-2 微博数据格式

3 Sentiment AnalysisModel

在我们的情感分析实验中，使用的模型是谷歌于 2018 年 10 月发表的 BERT，BERT 本质上是一个语言模型，但是和以往的语言表示模型不太一样，它是通过在所有层左右文本内容联合调节来预训练深层双向表征的模型。BERT 的贡献在于，1.证明了双向预训练对于语言表征的重要性。BERT 使用 Masked（遮盖） 语言模

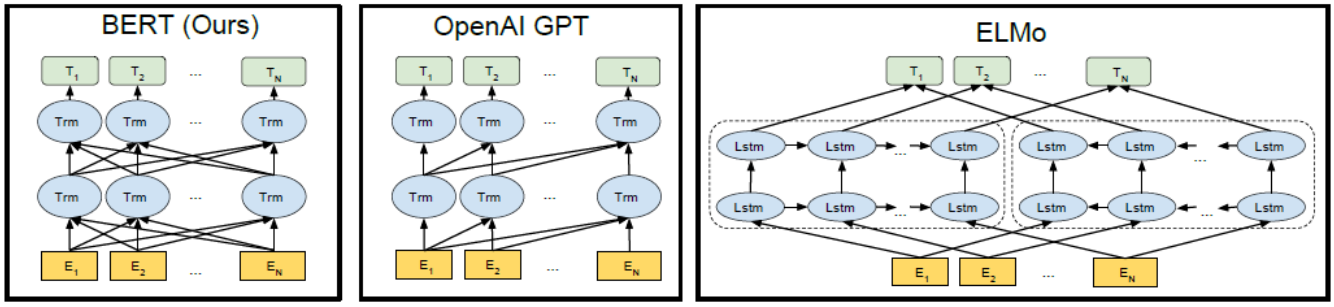


图 3-1 BERT 结构图　BERT 和 OpenAI GPT、ELMO 模型的最大的不同就是它使用了具有双向的 Transformer 作为编码器

型（也称为 “完形填空” 任务）来实现预训练好的深度双向表征。这也与 Peters 等人(发于 2018 年, ELMo 模型)形成了鲜明对比，Peters 等人使用了一种由左到右和从右到左的独立训练语言模型的浅层连接。2.展示了预训练的语言表征消除了许多经过大量工程设计的特定于任务的结构的需求。BERT 是第一个基于微调的表征模型，它在大量的语句级和 token 级任务中实现了最先进的性能，优于许多具有特定任务结构的系统。3.提升了 11 项 NLP 任务的最高水准。

3.1 BERT 的结构

图 3.1 给出了 BERT 和之前的两个文本表征模型的结构图，可以看出，OpenAI 的 GPT 使用了一个从左到右的模型。ELMo 使用了经过独立训练的从左到右和从右到左 LSTM 的连接来为下游任务生成特性。从图中也可以看出是两个相互独立的结构。而 BERT 使用的是一个双向的结构，从图中可以看出是在所有的网络层中都受到左右文本的共同作用。

与 BERT 最具可比性的现有预训练方法 OpenAI 的 GPT 模型，它在大型文本语料库中训练从左到右的 Transformer LM。实际上，BERT 中的许多设计决策都被有意地选择为尽可能接近 GPT，以便可以最小化地比较这两种方法。BERT 的工作的核心论点是其论文 3.3 节中提出的两个新的预训练任务占了大多数经验改进，但 BERT 和 GPT 在如何训练上还存在其他一些差异。

图中最底层的矩形表示文本的输入表示，在下节将会详细描述。中间层每个椭圆都代表一个 Transformer 编码器，其结构与 Vaswani 等人于《Attention is All You Need》[7]描述的原始模型完全相同，结构如下：

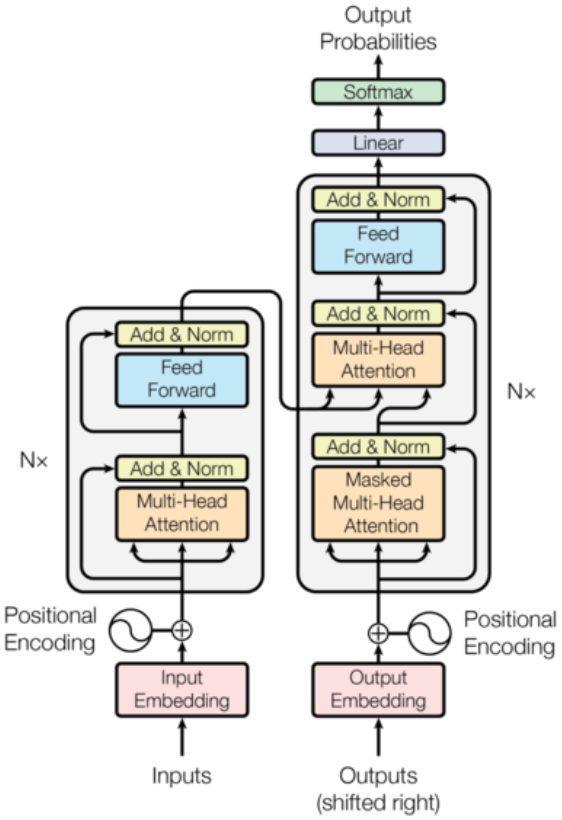


图 3-2 Transformer 结构图

关于 Transformer 的结构由于篇幅问题，本文档不再展开，每个 Transformer 可以将其看做是 LSTM 或 GRU 一样的一个特征提取单元。

3.2 模型的输入表示

BERT 的输入由 token embedding，segmentation embedding 和 position embedding 共同构成。

其 Position Embedding 使用的是传统 Transformer 的位置向量表示，把一个单词的位置表示为一个向量，向不能自主学习位置关系的 Attention 机制中引入句子中单词的位置信息。

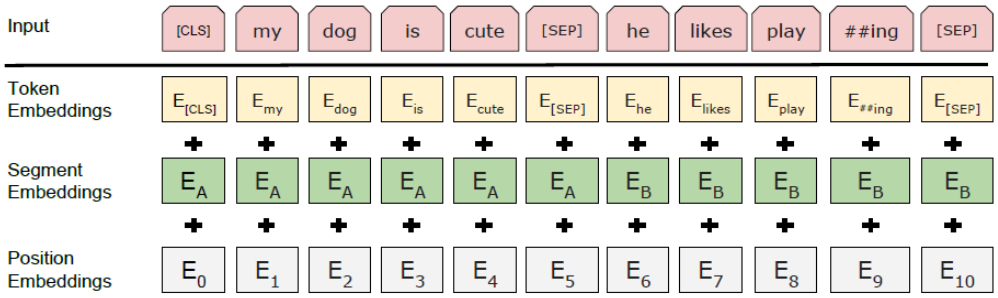


图 3-3 BERT 模型的输入表示，由词向量、段向量、位置向量拼接表示。

Segment Embedding 是一个 0-1 向量，0 和 1 分别代表该位置属于 A 句还是 B 句。

Token Embedding 使用的是预训练词向量。

CLS 标记：每个序列的第一个 token 始终是特殊分类嵌入（special classification embedding），即 CLS。对应于该 token 的最终隐藏状态（即，Transformer 的输出）被用于分类任务的聚合序列表示。如果没有分类任务的话，这个向量是被忽略的。

SEP 标记：用于分隔一对句子的特殊符号。有两种方法用于分隔句子：第一种是使用特殊符号 SEP；第二种是添加学习句子 A 嵌入到第一个句子的每个 token 中，句子 B 嵌入到第二个句子的每个 token 中。如果是单个输入的话，就只使用句子 A 。

3.3 预训练任务

3.3.1 任务 #1 Masked LM

为了训练深度双向表征，作者采用了一个直接的方法，即随机的掩盖一定比例的输入 token ，然后只预测这些被掩盖的 token。作者将这个过程作为 Masked LM，也被称为 “完形填空”。在这个任务中，被掩盖的 token 的最终隐藏向量将被输入到词汇表中的输出 softmax 层，就像标准的语言模型一样。虽然这个方法确实可以获得双向预训练模型，但这种方法有两个缺点。第一个缺点是创建了预训练和微调之间的不匹配内容，因为在微调期间从未看到 [MASK] token。为了缓解这个问题，作者并不总是用实际的 [MASK] token 替换被掩盖的单词。相反，训练一个数据生成器来随机选择 15% 的 token。 比如：my dog is hairy 这个句子中选择 hairy。然后执行以下过程：

- 80% 的时间：用 [Mask] token 掩盖之前选择的单词。例如：my dog is hairy → my dog is [Mask].
- 10% 的时间：用随机单词掩盖这个单词。例如：my dog is hairy → my dog is apple.
- 10% 的时间：保持单词不被掩盖。例如：my dog is hairy → my dog is hairy. (这样做的目的是将表征偏向于实际观察到的单词)

这个转换编码器并不知道哪个单词将被预测，或者哪个单词被随机单词取代。所以，它被迫保持每个输入 token 的分布式的上下文表征。另外，因为随机取代对于所有 token 来说，发生的概率只有 1.5%（15% 中的 10%），所以并不会损害模型的理解能力。

另一个缺点是，由于在每个 batch 中，只有 15% 的 token 需要被预测，这表明模型可能需要更多的预训练步骤才能收敛。在论文 5.3 节中，作者也证明 Masked LM （MLM）的收敛速度略慢于从左到右的模型（预测每个标记），但 MLM 模型的实证改进远远超过增加的训练成本。

3.3.2 任务 #2 Next Sentence Prediciton

很多重要的下游任务，像问题回答（QA）,自然语言推断（NLI）等都是基于理解两个句子之间的关系。这种句子之间的关系不能够被语言模型直接捕获。为了训练理解句子关系的模型，作者预先训练二进制化的下一句子预测任务，该任务可以从任何单词语料库中简单的生成。具体来说，当为每个预训练样本选择句子 A 和 B 时，50% 的时间 B 是跟随 A 的实际下一个句子，50% 的时间是来自语料库的随机句子。如下所示：

- Input = [CLS] the man went to [MASK] store [SEP] he bought a gallon [MASK] milk [SEP]

- Label = IsNext
- Input = [CLS] the man [MASK] to the store [SEP] penguin [MASK] are flight ##less birds [SEP]
- Label = NotNext

作者完全随机选择 NotNext 语句，最终预训练模型在此任务中达到 97%-98% 的准确率。尽管它很简单，但之后在 5.1 节中证明，预训练这项任务对 QA 和 NLI 都非常有帮助。

3.4 将BERT用于情感分类实验

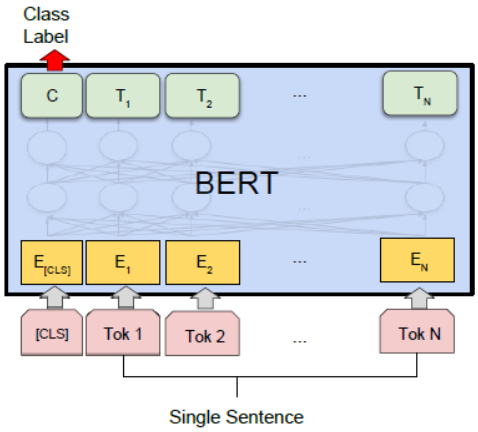


图 3-4 将 BERT 用于分类任务

我们的目标是给定一条微博正文，得到它的情感分类（3 分类），因此我们使用了一个外部数据集 SMP2019 的数据进行训练。我们将一条微博作为单句输入 BERT，将最后一层第一个 CLS 标记的输出作为句子表示，外接 softmax 分类器得到类别的概率分布。

4 Dataset

4.1 数据获取

新浪微博拥有 API2.0，虽然使用简单，但有要求频率和功能限制。因此，若想完成本项目的目标，还需要 API 之外更高级功能。此时，需要自行编写网络爬虫，我们自定义的爬虫可以根据本项目的目标完成几乎所有的要求。在这种情况下，我们使用两种方法一起完成数据采集任务。

因此，本项目数据采集流程图如下所示：

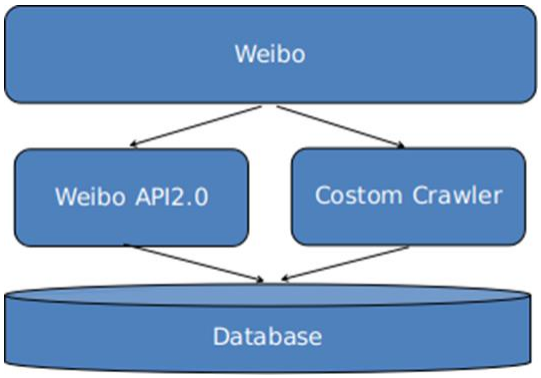


图 4-1 数据采集流程图

另外，针对本项目情感分析部分，其所用的数据集来源于 SMP2019，可以直接下载。

4.2 数据说明

通过爬虫方式采集到的微博数据汇总如下表所示：

数据类型	数据条数	数据细节
用户	186	100 关注者，100 粉丝
转发	15013	80 转发/用户
评论	300345	20 评论/转发

用于情感分析训练的数据说明：

情感分析模块使用的数据来自 SMP 2019 的测评，数据来源也是微博，每条数据都被标注为正向、负向、中立三类情感，我们的训练、验证、测试数据集划分在这里，然后数据集的平均字符数和平均单词数分别是 17 和 9。

数据集	分类数	训练/开发/测试集	AVG.CHARS	AVG.WORDS
SMP 2019	3	10353/1478/2957	17.45	9.77

其原始格式为：

图 4-2 情感分析数据集原始格式

5 Setting

对于情感分析的训练，超参数设置如下：

超参类型	超参值
LEARNING RATE	2E-5
Batch size	20
EPOCH	5
Max seq length	50
OPTIMIZER	ADAM
Dropout	0.4

情感分析使用的模型是谷歌去年发布的 BERT，将一条微博作为单句输入 BERT，将最后一层第一个 CLS 标记的输出作为句子表示，外接 softmax 得到类别的概率分布。

我们使用的是 BERT 的 pytorch 版本，其模型定义如下：

```
import torch.nn as nn
class BertForSequenceClassification(nn.Module):
    def __init__(self, bert, opt):
        super(BertForSequenceClassification, self).__init__()
        self.num_labels = opt.polarities_dim
        self.bert = bert
        self.dropout = nn.Dropout(opt.dropout)
        self.classifier = nn.Linear(opt.bert_dim, self.num_labels)

    def forward(self, inputs):
        input_ids, token_type_ids, attention_mask = inputs[0],inputs[1],i
        _, pooled_output = self.bert(input_ids, token_type_ids, attention
        pooled_output = self.dropout(pooled_output)
        logits = self.classifier(pooled_output)
        return logits
```

图 5-1 pytorch 版 BERT 模型定义

对于用户网络分析，功能设置包括：Degree Centrality、Closeness Centrality、Betweenness Centrality、Clique Analysis；

对于主题网络分析，功能设置包括：Degree Centrality、Closeness Centrality、Betweenness Centrality。

6 Results and Analysis

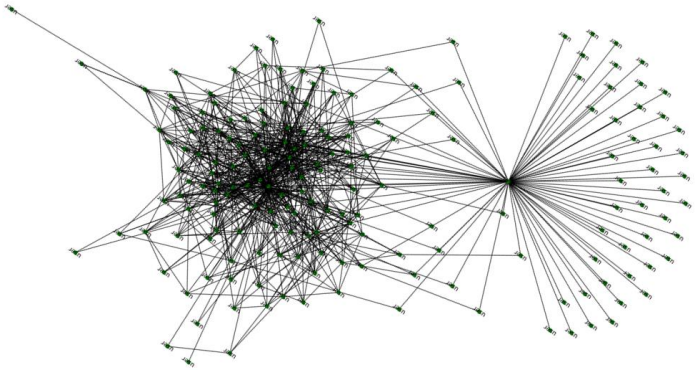
6.1 实验结果

1. 情感分析实验结果如下表所示：

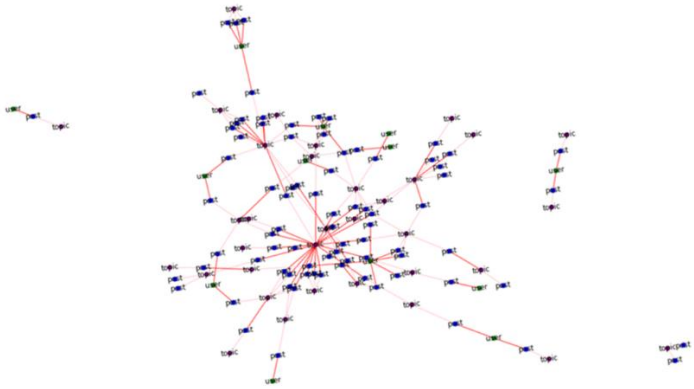
数据集	ACCURACY	F1
TRAINING SET	0.9088	0.8534
DEV SET	0.8778	0.8322
TEST SET	0.8545	0.8267

2. 社交网络分析结果

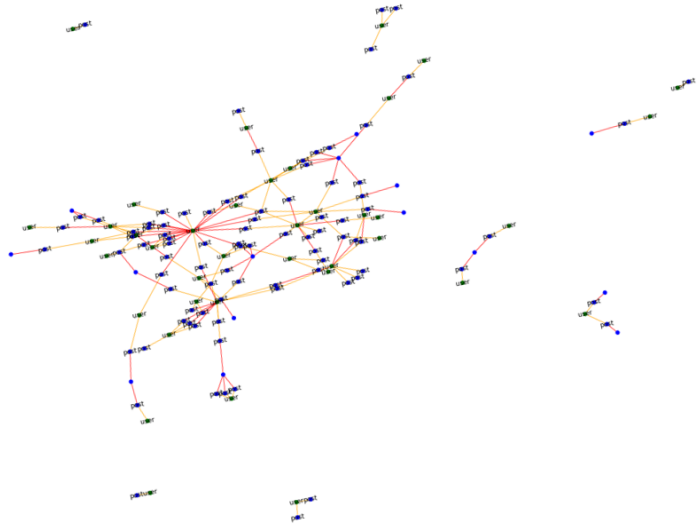
1）User-following-User



2）Post-focus-Topic



3） Post-@-User



6.2 实验分析

在情感分类方面，我们的模型经过验证集调参，在 3 分类的情况下可以得到令我们满意的性能，且实

验使用的数据集来源也是微博，因此我们认为由此训练的模型可以用来预测我们爬取微博数据的情感倾向。

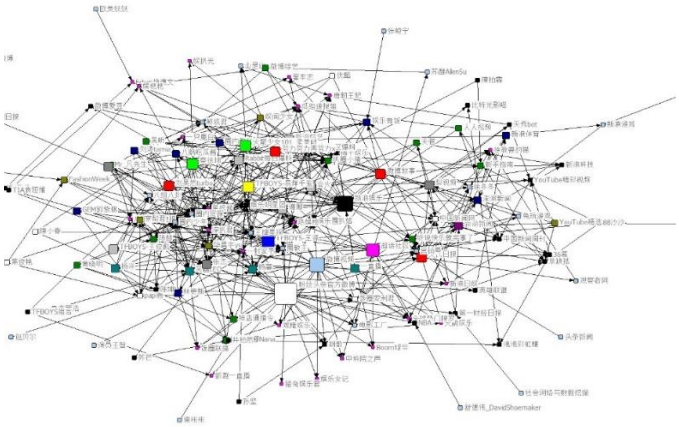
网络分析工具使用汇总如下：

Name	Version	Function
UNICENT	V6.645	NETWORK ANALYSIS
NETDRAW	V2.161	VISUALIZATION

针对用户网络和话题网络分析，包括度中心性、紧密度中心性、介数中心性和团。使用了 UNICENT 网络分析工具以及 NetDraw 进行可视化。

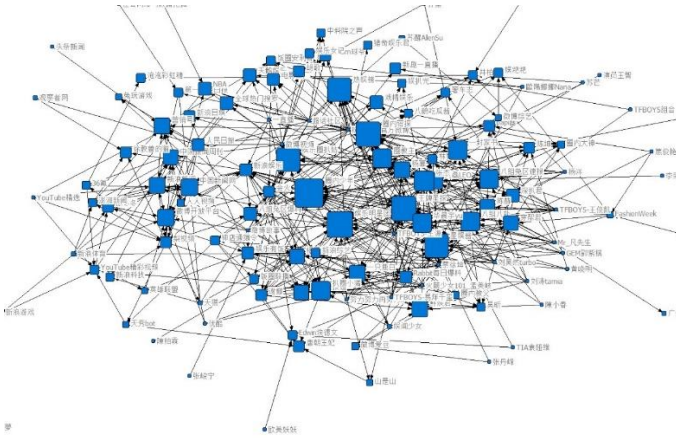
1．用户网络：

（1）度中心性



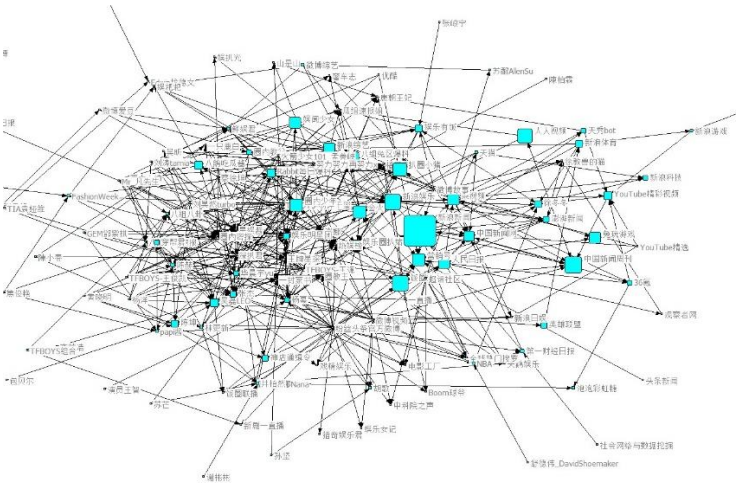
分析：度中心性由相邻节点的数量决定，在现实网络中只有少部分节点具有较高的度。在整个用户网络中“粉丝头条官方微博”、“新浪娱乐”拥有的信息资源掌控能力和信息交流能力强，在众多节点之间存在交流。

（2）紧密度中心性



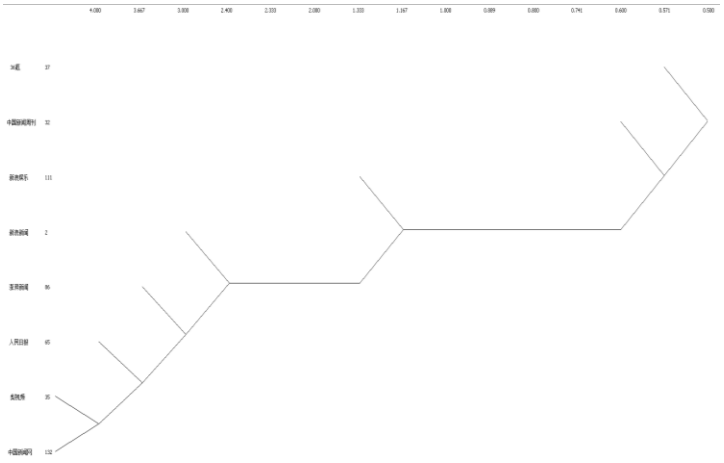
分析：紧密度中心性表示到其他节点的紧密程度，反应该节点的重要性。在用户网络中“圈内少年”、“娱乐明星团”这两个博主在这个用户网络中具有较大的影响力。

（3）介数中心性



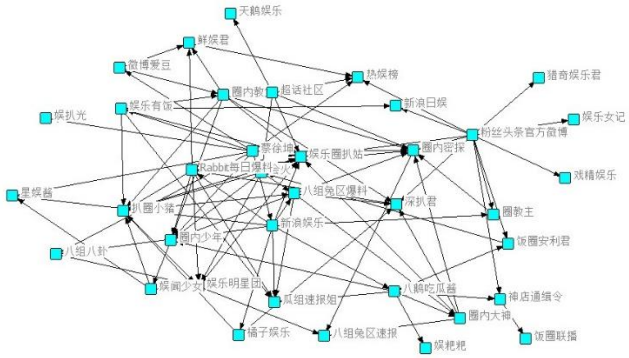
分析：介数中心性表示经过某一节点最短路径的数量，具有高介度的节点在网络中的信息传播中起着重要作用。在这个用户网络中“新浪新闻”具有较高的信息传播能力。

（4）团分析



分析：我们对整个用户网络进行社区发现分析，共得到 18 个团。我们认为在这个用户网络中“梨视频”和“中国新闻网”更相似，其成团分值为 4。

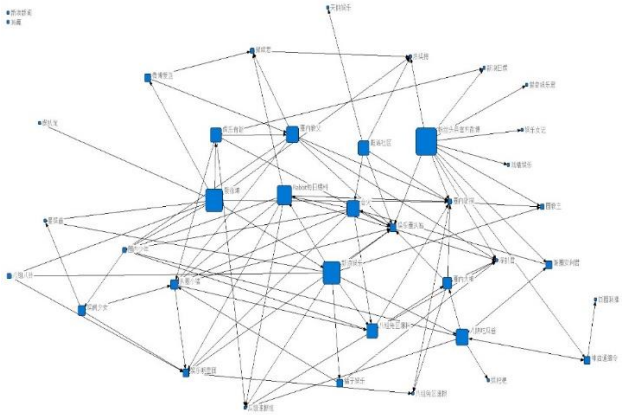
2. 话题网络分析



#蔡徐坤给 B 站发律师函#

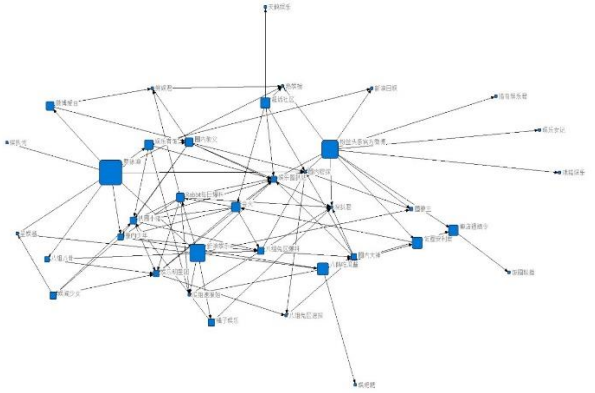
我们进行话题网络分析，选择的话题为“蔡徐坤给 B 站发律师函”，在我们关注的用户里面，共有 36 个用户对这个话题进行了讨论，建立了话题网络。

（1）度中心性



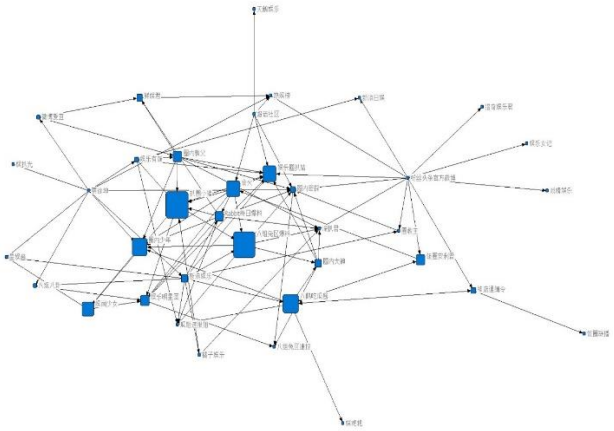
分析：度中心性最大的为“粉丝头条官方微博”，说明在这个话题网络中“粉丝头条官方微博”与众多节点之间存在交流，具有很高的积极性。

（2）紧密度中心性



分析：对整个网络进行紧密度中心性分析，我们发现“蔡徐坤”在这个话题网络中具有很高的“辐射力”。

（3）介数中心性



分析：介数中心性最大的是“八组兔区爆料”、“扒圈小猪”，作为娱乐方向的大 V，他们在“蔡徐坤”话题中起到非常重要作用，很多用户通过他们来获得话题的信息。如果失去该节点，那么经过“八组兔区爆料”、“扒圈小猪”的所有最短路径就会改变。

7 Conclusion

本项目的主要工作如下：

1. 我们使用新浪微博官方提供的 Weibo API 2.0 接口和一个传统的自定义爬虫，完成了数据采集的工作，共得到 186 名用户的 15013 条博文和 300345 条评论数据。
2. 对采集到的数据进行一些统计与挖掘，构建词云、用户与博文的关系、博文长度关系等条目，作为对数据集的初步探索。
3. 我们使用 SMP2019 数据集和 BERT 模型进行训练，得到了一个情感分类模型，对我们采集的数据进行情感分类。
4. 我们使用网络分析和相关可视化技术，对用户网络和话题网络进行建模和分析，并且得到了一些和现实直觉相符合的结论。
5. 我们基于 python 构建了一个可视化系统，将上述所有功能集成，并且考虑了很多实际应用上的问题，使我们的可视化系统具有可重用性，具有一定的实际应用价值。同时对我们分析结果进行展示，让受众可以得到更直观的展示。

在这些工作的基础上，我们进行了合理的分工，从题目确定、问题分析、数据爬取、模型训练、网络分析，到最后的可视化展示，每个成员都展示了很高的积极性与合作性，同时将平时课上老师所讲的内容活学活用，令我们受益匪浅。

在课程之后，我们打算继续完善这个项目，作为实验室内部的一个长期维护项目，未来会加入微博的语义分析、相似度分析，微博和话题的归属度分析等功能，并考虑开源供大家学习与探讨。

REFERENCES

[1] Qiang, Yan, Lanli, Yi, Lianren, Wu. Human dynamic model co-driven by interest and social identity in the MicroBlogcommunity[J]. Physica A, 2012 ,(391): 1540 -1545

[2] Bruno Goncalves, Nicola Perra, Alessandro Vespignani. Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number[J]. Plos One, 2011, 6(8):1 - 5

[3] 王晓光. 微博客用户行为特征与关系特征实证分析——以新浪微博为例[J]. 图书情报工作,2010,54(14):66-70.

[4] 袁毅, 杨成明.微博客用户信息交流过程中形成的不同社会网络及其关系实证研究[J]. 图书情报工作,2011,55(12):31-35.

[5] 杨成明. 微博客用户行为特征实证分析[J]. 图书情报工作 2011,55(12):21-25.

[6] 平亮, 宗利永. 基于社会网络中心性分析的微博信息传播研究——以 sina 微博为例图书情报知识,2010,138(6):92-97.

[7] 赵文兵, 朱庆华, 吴克文, 黄奇. 微博客用户特性及动机分析——以和讯财经微博为例[J]. 现代图书情报技术，2011,202(2):70-75.