基于数据挖掘技术的个性化微博推荐

成员分工

姓名	学号	分工
彭成	3120191033	算法实现，数据分析
高佳萌	3220190942	算法实现，文档编写
赵嘉旌	3220190923	算法实现，可视化
李家硕	3220190952	提取特征，文档编写
张博	3220190995	构建模型，文档编写

1 问题背景及分析

随着移动互联网的发展，各种移动社交应用不断涌现，微博已经成为当下最热门的社交平台之一。面对庞大的用户生成内容带来的信息过载问题，如何为用户提供准确和高效的知识服务成为当下研究的热点。本文基于数据挖掘技术，进行个性化微博推荐研究。本文采用的模型包括两大部分：待推荐微博获取和微博排序。首先利用三种推荐策略获取待推荐微博，包括：基于用户兴趣标签的推荐、基于用户兴趣话题的推荐、以及基于用户协同过滤的推荐。在微博排序阶段，将以上三种推荐算法的召回结果汇总，将推荐的二分类问题转化为排序问题，综合考虑影响用户对微博感兴趣程度的多种因素，使用深度神经网络进行微博排序。本文通过在新浪微博真实数据上进行实验，证明本文采用的模型在推荐准确率和召回率上取得了较好的效果。

随着移动互联网的发展，各种移动社交应用不断涌现，如 Facebook、Twitter，新浪微博等，社交应用的出现和飞速发展深深影响了人们的交流和娱乐方式，成为人们生活中不可或缺的一部分。用户可以通过 PC、手机等多种移动终端，以文字、图片、视频等多媒体形式，实现信息的即时分享、传播互动。明星和偶像利用社交平台发布自己最新的动态，与粉丝交流，提高自己的人气；商家通过微博发布广告和商品促销信息，吸引用户购买；作为普通用户，可以通过微博发布自己的生活动态，了解亲朋好友的生活，关注最新最热的新闻，同时也可以根据自己的兴趣爱好关注微博作者，与兴趣相投的人进行交流和沟通。微博平台由于其自身内容的丰富性和实时性吸引了大量的用户。根据 2017 年新浪微博用户发展报告，截

止 2017 年 9 月，新浪微博月活跃用户（MAU）共 3.76 亿，日活跃用户（DAU）达到 1.65 亿，微博内容存量已超过千亿。庞大的用户生成内容带来的信息过载问题，使得用户难以从大量的信息中获取自己感兴趣的内容，推荐算法便是解决这一问题的关键。

微博用户主要的信息获取方式来源于主页，也就是通过关注其他用户，接收他们的微博，形成自己的收听列表。一个活跃的微博用户每天在主页能收取到成百上千的微博，用户无法看完所有微博，并且其中很多微博未必是用户感兴趣的内容。所以，准确的挖掘用户兴趣，利用好的推荐算法，优先给用户展现其感兴趣的内容，对提升用户体验起着至关重要的作用。

本文以微博为例，利用数据挖掘技术挖掘用户兴趣，为用户提供准确和高效的知识服务，辅助解决信息过载问题

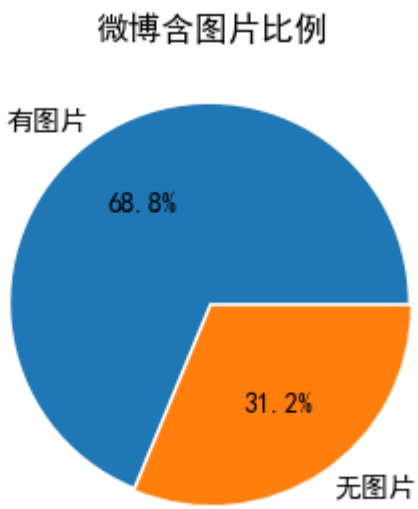
2 问题描述

2.1 数据准备

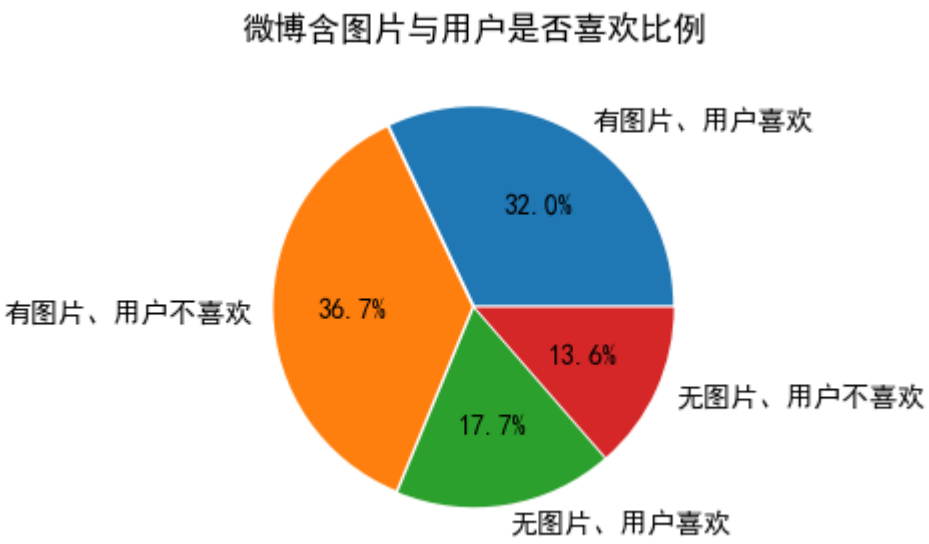
本课题利用从新浪微博 API 爬取的真实用户数据进行实验。随机选择 20 名活跃用户作为目标用户，爬取其一个月内发布的微博，以及目标用户的主页微博，即关注列表的微博信息，共包括 3217 名用户的 216176 条微博。其中每条微博包含 ID、作者 ID、发布时间、内容、点赞数、评论数、转发数等信息。

2.2 数据可视化

2.2.1 微博内容含图片的比例

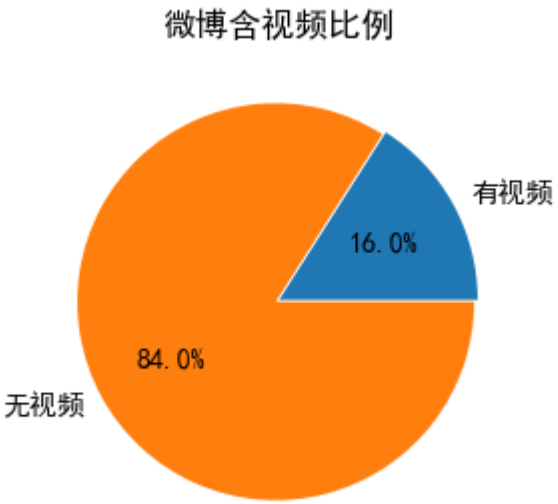


我们可以看出大部分的微博都是带有图片的，但是不含图片的微博也不少，不能直接忽略，所以如何处理好带图片与不带图片这两部分是个值得思考的问题。

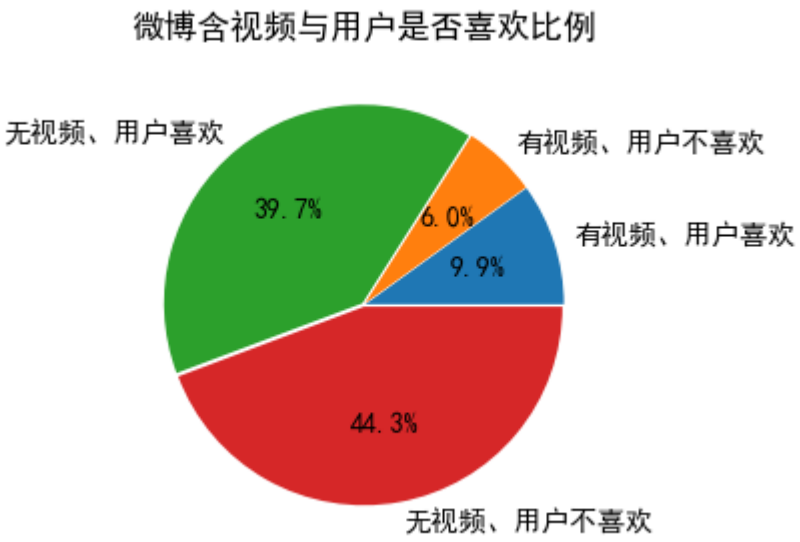


进一步分析，我们可以看出，微博内容有无图片和是否推荐存在一定关系，但关系不是很大，故这个特征起的效果并不是很强。

2.2.2 微博内容含视频的比例

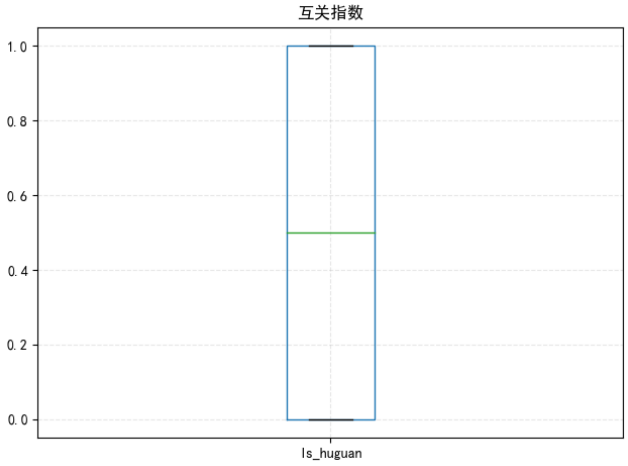


我们可以看到绝大部分的微博是不带有视频的，所以我们如果精力不够的情况下，可以在对其分析的时候，可以忽略带有视频的微博，以减少工作量。

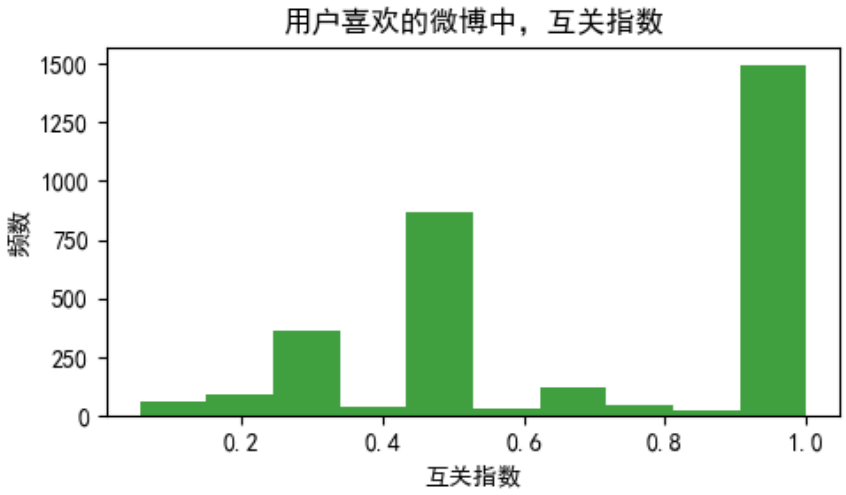


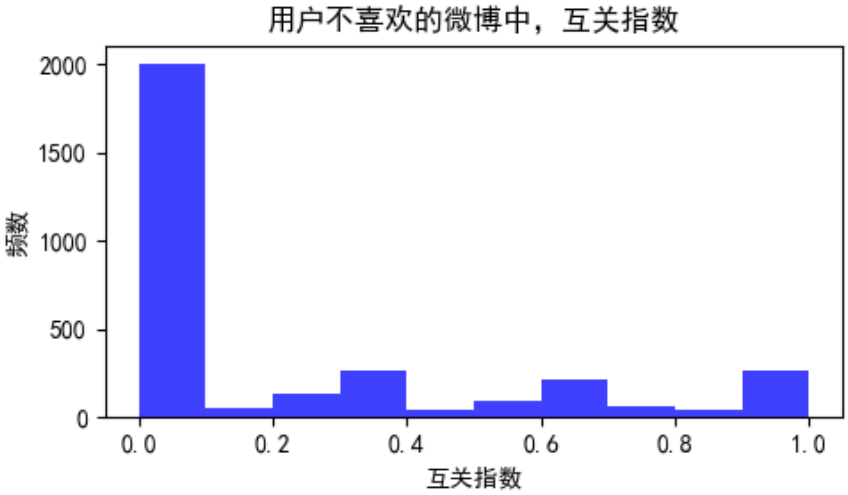
进一步分析，我们可以看出，微博内容有无视频和是否推荐存在一定关系，有视频的微博推荐的可能性会增大。但关系不是很大，故这个特征起的效果并不是很强。

2.2.3 作者、用户互关与喜欢与否的关系



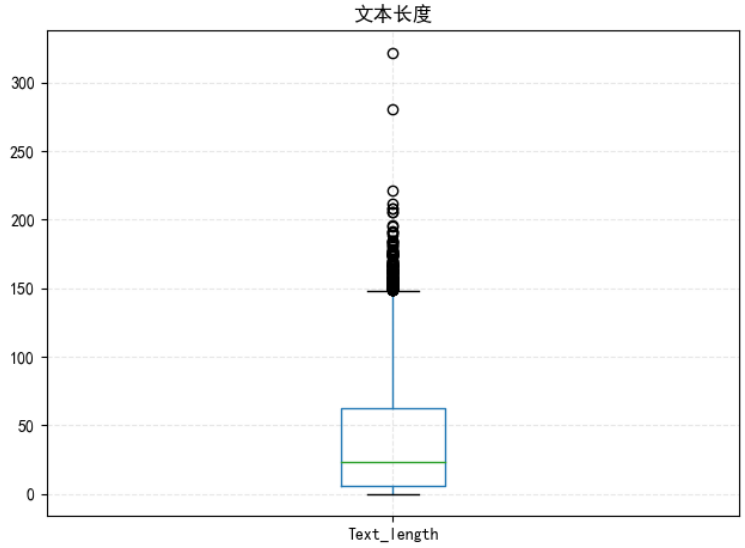
由盒图可以看出，互关指数分布还比较平均，为两头大，中间小的形状，中位数为 0.5，由于两头大，中间小，所以 Q1、Q3 与 min、max 值重合。



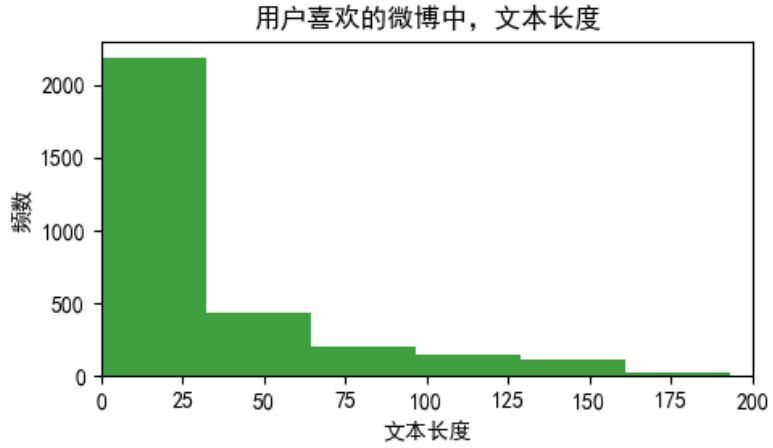


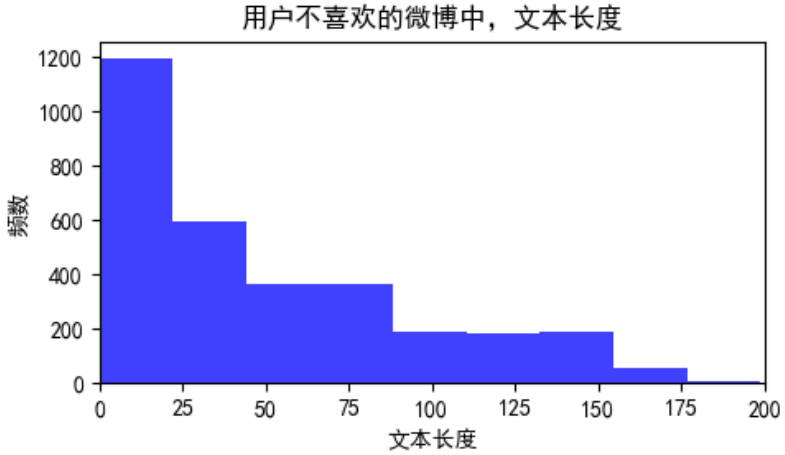
由上面两图我们可以分析得出，互关指数越高，用户越趋近于喜欢这套微博内容，相反，互关指数越低，用户越不喜欢该内容。

2.2.4 文本长度与该微博内容受不受欢迎的关系：



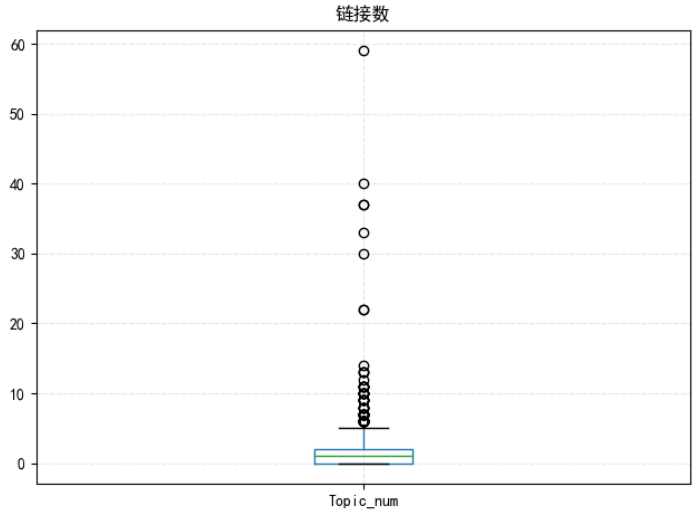
由盒图可得知，大部分微博文本内容长度在 150 字之内，且一般在 50 字之内。



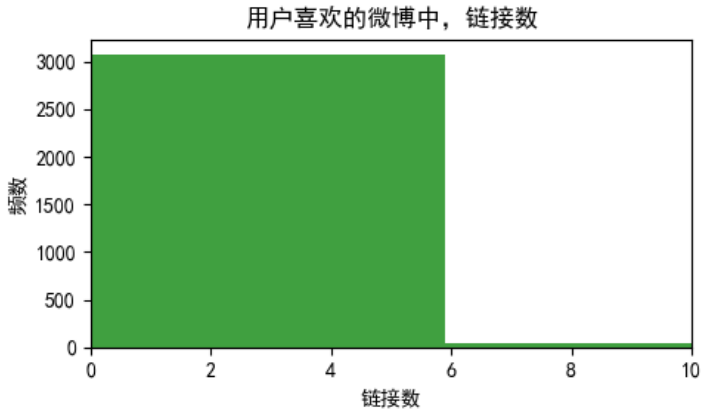


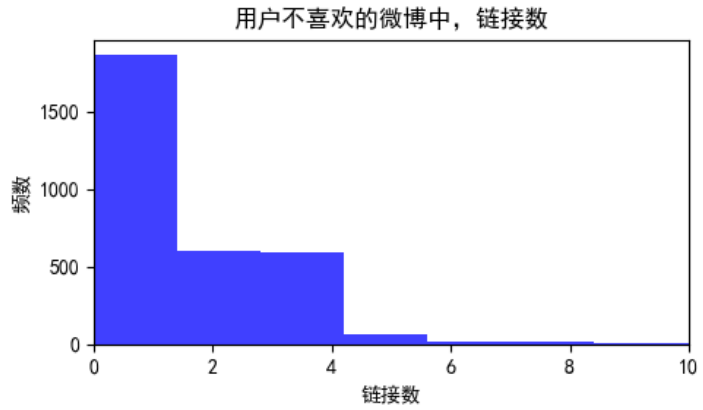
可以从上图得知，文本的长度与用户喜欢与否没有太大联系，但明显的是，短文本看上去更容易受到喜欢。

2.2.5 链接数量与该微博内容受不受欢迎的关系：



由盒图可得知，大部分微博不带有链接或带有少数几个链接。

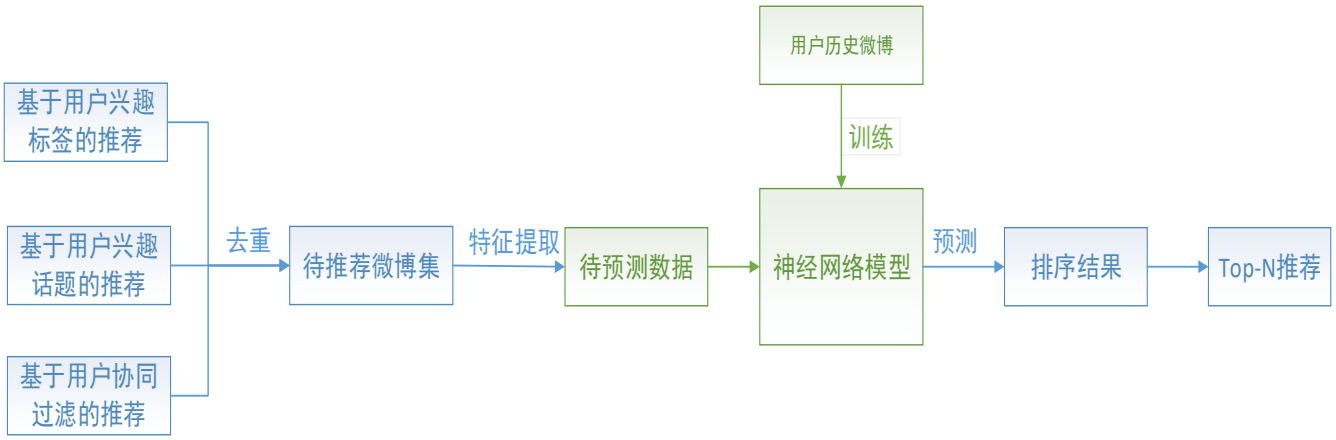




可以由上图得知，微博内容中含有的链接数越多，其不受喜欢的可能性越大。

3 微博推荐模型

本文采用混合型微博推荐算法进行个性化微博推荐。该模型分为两大部分：待推荐微博获取和微博排序。待推荐微博获取模块从三种不同的角度，采用数据挖掘技术挖掘用户兴趣，并根据用户兴趣召回微博，形成待推荐集。三种策略分别为基于用户兴趣标签的推荐、基于用户兴趣话题的推荐、基于用户协同过滤的推荐。微博排序模块对待推荐微博集提取分类特征，输入到深度神经网络模型，并根据输出值进行微博排序，实现 Top-N 推荐。模型总体的流程图如图所示。



3.1 待推荐微博获取

3.1.1 基于用户兴趣标签的推荐

用户标签是一种个性化信息，描述了用户的个人兴趣，对于推荐系统来说是构造用户画像挖掘、用户兴趣的重要信息。但实际上很少有用户主动为自己添加兴趣标签，且微博中的用户简介等信息往往长度很短且结构不统一，严重的数据稀疏使得难以进行基于标签的推荐。本文利用 TextRank with TF-IDF[11]关键词提取算法，从用户历史微博内容中提取关键词，并针对单个关键词描述不准确的问题，采用关联规则 Apriori 算法对其进行扩展，得到用户兴趣标签。

1、采用 TextRank with TF-IDF 关键词提取算法挖掘关键词

首先将文本构造成一个无向图 G(V,E)，V 为顶点，即文本中的单词。E 为边，连接图中的顶点，权重为顶点单词的共现次数。每个顶点 V_i 的权重得分可通过公式(1)迭代得出：

$$S(V_i)=(1-d)+d*(1+TFIDF(V_i))*\sum_{V_j\in E(V_i)}\frac{w_{ij}}{\sum_{V_k\in E(V_j)}w_{jk}}*S(V_j) \tag{1}$$

公式(1)中，$S(V_i)$表示顶点 V_i 的权重得分；d 为阻尼系数，表示图中某一顶点指向其他任意顶点的概率，一般取值 0.85；$E(V_i)$表示以 V_i 作为顶点的边的集合；w_{ij} 表示由顶点 V_i 和 V_j 所连接的边的权重。公式表明，每个单词 V_i 的权重由与其共现的其他单词 V_j 的权重和它们边的权重占比决定。

基于公式(1)进行迭代，最终求得每个顶点即单词的权重，取 Top-20 作为用户的关键词。

2、使用关联规则 Apriori 算法对关键词进行扩展

利用传统的关键词提取算法虽然能够挖掘到用户兴趣，但存在一些问题。下面以某真实用户 A 举例说明，利用上节的方法，得到 A 的关键词为：

“奥斯卡，抽奖，sophie，funko，权力，游戏，粉丝，漫威，凤凰，turner”

其中“sophie”和“turner”是指一位英国女演员 Sophie Turner，“权力”和“游戏”是指一部有名的美国电视剧《权力的游戏》，这些特有的名词在分词这一步骤被拆分，因为算法无法对这些名词组进行识别。并且，一个单词在不同语境下可能会有不同的含义，比如，“苹果”和“香蕉”在一起出现时表示的含义明显与和“手机”或“产品”一起出现时表示的含义不同。若直接将挖掘得到的关键词作为兴趣标签，会导致两个问题：

①兴趣标签不能准确表达用户兴趣，产生歧义（比如在对用户 A 进行基于内容的推荐时，可能会给他推荐“游戏”相关的微博，但事实上用户对游戏并不感兴趣）。

②兴趣标签只能涵盖用户的一部分兴趣点。

本文利用关联规则 Apriori 算法，对关键词进行扩展，解决上述问题。步骤如下：

①选取权重 Top-20 的关键词作为项集，遍历目标用户历史微博，对于每一条微博，记录出现在该条微博中的关键词，形成一条事务。仍然以用户 a 举例说明：

用户 A 的 Top-20 关键词为：

“奥斯卡，抽奖，sophie，funko，权力，游戏，粉丝，漫威，凤凰，turner，姐妹，战警，jordyn，奥妹，kim，eminem，钱，khloe，品牌，休息室”

对于用户 A 的曾经转发的这条微博：

“惹//【权游将推出美妆周边】著名美妆品牌 urbandecay 在今天宣布将与权力的游戏合作，在四月共同推出新款美妆产品。美剧权力的游戏#权力的游戏第八季##ForTheThrone#”

生成一条事务为：[“品牌”，“权力”，“游戏”]

②设置最小支持次数为 2，最小置信度为 0.7，利用 Apriori 算法生成频繁项集和关联规则，记录所有形式为 X->Y 的关联规则：X 为一个关键词，Y 为关键词的集合。用户 A 的关联规则如下：

{'turner'}-->{'sophie'}

{'sophie'}-->{'turner'}

{'凤凰'}-->{'战警'}

{'游戏'}-->{'权力'}

{'权力'}-->{'游戏'}

③为每个关键词维护一个扩展集合，遍历所有关联规则，将每条关联规则中 Y 的每个关键词加入关键词 X 的扩展集合。扩展集合相同的只计一次，分数累加。按扩展后的关键词集合进行排序，取 Top-10 作为用户兴趣标签。

3、基于用户兴趣标签的微博推荐

得到用户标签后，首先对兴趣标签得分进行归一化处理，对于每一条待推荐微博，遍历用户兴趣标签，若某个标签出现在微博文本内容中，则微博兴趣度得分加上对应标签得分。注意对于多单词组成的标签，必须每个单词均出现在微博文本中才能加分。

设目标用户为 u，待推荐微博为 w，用户兴趣标签为 tag={t_1,t_2,...t_{10}}，用户 u 对微博 w 的兴趣度计算方式为公式(2)：

$$TagInterest(u,w)=\sum_{i=1}^{10} In(w,t_i)*Score(u,t_i) \tag{2}$$

In(w,t_i)表示标签 t_i 是否在微博 w 中，是为 1，否为 0；Score 为标签的 TextRank 得分。将待推荐微博按兴趣度降序排序，取 Top-N 进行推荐。

基于用户兴趣标签推荐的总流程如算法 1 所示。

算法 1：基于用户兴趣标签的推荐
输入：训练集微博、待推荐微博、目标用户 u、推荐个数 N
输出：推荐微博列表
1、将每个用户的历史微博合并为一个微博集
2、对每条微博进行分词，过滤停用词，只保留名词
3、对目标用户 u 的微博集中每个词，计算 TF-IDF 得分
4、根据公式（1）迭代，计算 Textrank with TF-IDF 得分
5、根据得分降序排序，取前 20，使用 Apriori 算法进行扩展
6、按扩展后的得分排序，取 Top10 作为用户标签
7、根据公式（2）计算用户对待推荐微博的兴趣度，取 top-N 推荐

3.1.2 基于用户兴趣话题的推荐

传统的微博推荐只利用了微博中的文本信息，而忽略了微博中包含的丰富的链接信息。本文利用微博中超话和话题链接等，发现用户感兴趣的话题，并为用户推荐内容上相似的微博。

1、用户兴趣话题挖掘

微博中包含了丰富的链接信息，可以很好的表明一条微博的主题，例如图所示微博：

微博中的蓝色字均为链接，两个“#”中的为微博话题链接，类似超人符号开头的为微博超级话题链接，点击链接后进入话题界面，可以查看相应话题下的各种微博。目前很多用户在发微博时喜欢添加上对应的话题，一方面明确标识微博的主题，另一方面能够提高微博的曝光度。针对推荐系统来说，微博用户对话题的行为能够表明用户的兴趣，有助于构建用户画像，并且微博话题往往结构统一，利于分析。

本文利用微博话题链接信息，采用 TF-IDF 算法挖掘用户兴趣话题。对于每一个微博用户，将其历史微博话题列表视为一个文档，目标用户以及关注用户的微博话题列表形成一个文档集，根据 TF-IDF 公式计算目标用户话题列表中每个话题的 TF-IDF 得分，取 Top-10 作为用户兴趣话题。

2、基于用户兴趣话题的推荐

对兴趣话题得分进行归一化处理，每一条待推荐微博，遍历用户兴趣话题，若某个话题出现在微博中，则微博兴趣度得分加上对应话题得分。设用户为 u，待推荐微博为 w，用户兴趣话题为 topic=$\{t_1,t_2,...t_{10}\}$，公式(3)为用户对微博的兴趣度计算公式：

$$TopicInterest(u,w)=\sum_{i=1}^{10} In(w,t_i)*Score(u,t_i) \tag{3}$$

$In(w,t_i)$表示话题 t_i 是否在微博 w 中，是为 1，否为 0；Score 为话题的 TF-IDF 得分。将待推荐微博按兴趣度降序排序，取 Top-N 进行推荐。

基于用户兴趣话题推荐算法的总流程如算法 2 所示

算法 2：基于用户兴趣话题的推荐
输入：训练集微博、待推荐微博、目标用户 u、推荐个数 N
输出：推荐微博列表
1、获取训练集微博中的话题链接部分。

2、对部分无意义话题进行过滤，低频话题进行过滤

3、对于目标用户微博中的话题链接，计算 TF-IDF 得分

4、按得分排序，取 Top10 作为用户兴趣话题

5、根据公式（3）计算用户对待推荐微博的兴趣度，取 Top-N 推荐

3.1.3 基于用户协同过滤的推荐

基于协同过滤的推荐是目前个性化推荐算法中研究和应用最为广泛的推荐算法，也是推荐效果最好的算法之一。算法的主要思想是：根据目标用户的历史行为，找到与其兴趣相似的用户集合，并将这个集合中的用户喜欢且目标用户没有见过的物品推荐给目标用户。本文使用基于用户的协同过滤算法，利用用户历史交互行为计算用户间相似度，为目标用户推荐其相似用户的微博。

1、用户相似度计算

交互行为指用户的发布、转发微博行为。交互行为一方面能够表现用户的兴趣，即用户会发布或转发感兴趣内容的微博；另一方面能够体现用户的社交偏好，反映社交关系的强弱：用户会转发自己喜欢的博主的微博。本文通过统计目标用户与其他用户的微博转发行为，和用户对原创作者的转发次数计算交互行为相似度。

用户 u 的历史微博集为 W(u)，用户 v 的历史微博集为 W(v)，则用户 u 与用户 v 的交互行为相似度计算方式为公式(4)：

$$Sim(u,v)=\frac{\left|W(u)\cap W(v)\right|+\sum_{w\in W(u)\cap W(v)} isOriginal(v,w)}{\sqrt{\left|W(v)\right|}} \tag{4}$$

$W(u)\cap W(v)$ 表示用户 u 和用户 v 同时发布或转发过的微博集，isOriginal(v,w)表示用户 v 是否为微博 w 的原创作者，是为 1，否为 0。

2、 基于协同过滤的推荐

得到用户相似度后，基于用户的协同过滤通过公式(5)计算用户对物品的兴趣度：

$$Score(u,w)=\sum_{v\in Like(w)} Sim(u,v) \tag{5}$$

Like(w)表示对微博 w 产生行为的用户集合，Sim(u,v)为用户 u 和用户 v 的相似度。直观来讲，用户 u 对微博 w 的兴趣度为喜欢物品 w 的所有用户与用户 u 的相似度得分之和。

3.2 微博排序

传统的推荐算法，如基于内容或协同过滤的推荐，往往只考虑了内容或社交关系等一种或两种影响因素，而用户对一条微博感兴趣程度的影响因素有很多，比如用户对微博内容的感兴趣程度、对微博作者的喜好或熟悉程度、微博本身内容是否丰富、吸引人等。本模型将3.1 三种推荐策略获得的待推荐微博混合去重，并充分考虑影响用户对微博感兴趣程度的多重因素，从微博中提取特征。将用户历史微博作为训练样本，类标签为用户是否对微博产生行为（0/1），将排序看作机器学习的二分类问题，训练深度神经网络模型，预测用户对待推荐微博的感兴趣程度，按得分进行 Top-N 推荐。

本模型通过主要的三大影响因素：个人兴趣、作者偏好和微博质量，选取下表 1 中八个特征用于构建神经网络模型。

表 1 模型特征

Table1 The model features

影响因素	特征名	含义
个人兴趣	Tag_score	微博内容与目标用户兴趣标签匹配得分，计算方式为公式(2)
	Topic_score	微博话题与目标用户兴趣话题匹配得分，计算方式为公式(3)
作者偏好	Author_score	目标用户与微博作者的相似度得分，计算方式为公式(4)
	Is_huguan	目标用户与微博作者是否互相关注，是为 1，否为 0
微博质量	Has_pic	微博是否包含图片，是为 1，否为 0
	Has_video	微博是否包含视频，是为 1，否为 0
	Text_length	微博文本内容长度
	Topic_num	微博中包含的话题链接数目

4 实验

4.1 实验设置

本文实验数据来源于新浪微博真实数据。随机选择 20 名用户作为目标用户，爬取其 200 条微博，以及目标用户的关注用户微博。共获得 3217 名用户的 216176 条微博，并划分训练集和测试集。得到训练集：146908 条微博；测试集：69268 条微博。

4.2 实验结果

4.2.1 基于用户兴趣标签的推荐

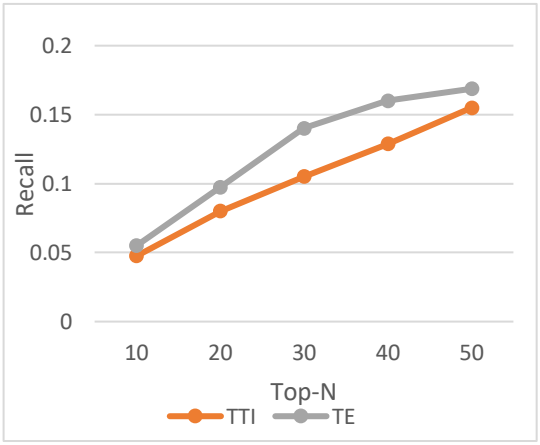
本节实验实现了 3.1.1 节中基于用户兴趣标签的推荐（TE），为了证明基于关联规则的标签扩展的有效性，与 Textrank with TF-IDF（TTI）算法作为对比，采用推荐准确率和召回率作为评价指标，分别进行 Top-10、20、30、40、50 推荐，实验结果如下。

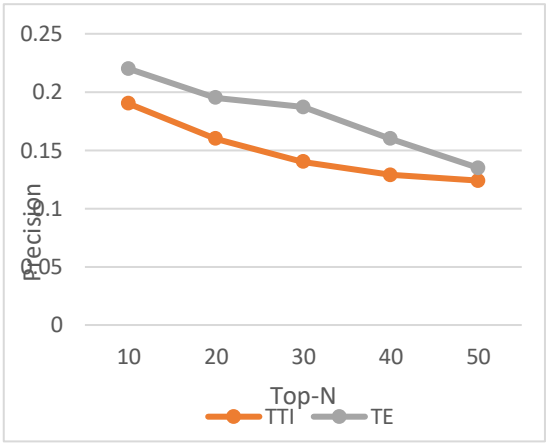
Recall

Top	10	20	30	40	50
TTI	0.048	0.08	0.105	0.129	0.155
TE	**0.055**	**0.098**	**0.14**	**0.16**	**0.169**

Precision

Top	10	20	30	40	50
TTI	0.19	0.16	0.14	0.129	0.124
TE	**0.22**	**0.195**	**0.187**	**0.16**	**0.135**





从实验结果可以看出，与 Textrank with TF-IDF（TTI）算法相比，改进后推荐效果有明显提升。

本文选取了两个真实微博用户，分别用 Textrank with TF-IDF 和基于关联规则的标签扩展方法挖掘用户兴趣标签，部分兴趣标签结果如下表。根据扩展后的标签，可以明显看出用户 A 对美剧权利的游戏、电影 x 战警等相关内容感兴趣，而用户 B 对电影神奇动物在哪里感兴趣，标签 1、2 都是电影中的角色。通过对比可以发现，经过扩展后的标签能够将相关的单词组合，能够更加准确表达用户兴趣，直观表明了算法的有效性。

	User A(TTI)	User A(TE)	User B(TTI)	User B(TE)
1	oscar	Game,thrones	grindelwald	grindelwald
2	lottery	sophie,turner	geric	grindelwald,dumbledore
3	sophie	phoenix,x-men,marvel	ggad	geric
4	game	oscar	beast	ggad,beast
5	throne	lottery	fan	marvel,fan

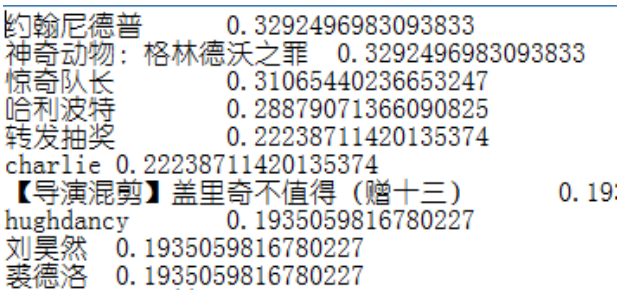
4.2.2 基于用户兴趣话题的推荐

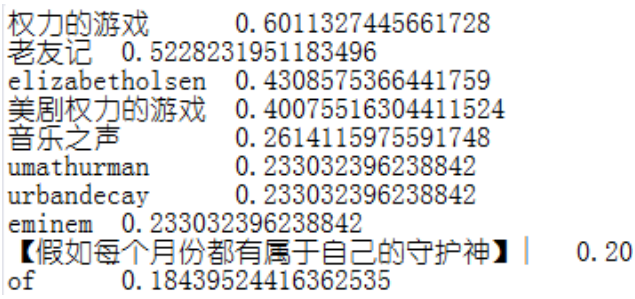
在训练数据集上进行用户兴趣话题挖掘，测试集作为待推荐微博集，分别进行 Top-10、20、30、40、50 推荐，实验结果如下：

Top	10	20	30	40	50
Recall	0.068	0.079	0.093	0.111	0.116
Precision	0.27	0.158	0.124	0.111	0.093

从上面结果可以看出，基于用户兴趣话题的召回率在 Top-10 推荐时高于基于用户兴趣标签的推荐，但随着推荐个数的增加，用户兴趣话题的召回率的增长并不十分明显，原因是并不是每条微博都包含话题链接，有些用户的微博包含的话题链接较少，实际召回个数并没有达到设定的阈值。

部分用户兴趣话题展示：





4.2.3 基于用户协同过滤的推荐

根据用户历史行为计算用户间相似度，并实现基于用户协同过滤的推荐，实验结果如下：

Top	10	20	30	40	50
Recall	0.085	0.145	0.205	0.25	0.275
Precision	0.34	0.29	0.273	0.25	0.22

可以看出基于用户协同过滤的推荐效果比以上基于用户兴趣标签、兴趣话题的推荐效果要好。

部分用户相似度结果展示：

1942793263	0.16012815380508713	6192453282	0.9071067811865474
3264015433	0.0	5446959152	0.35355339059327373
2850809427	0.0	6408711245	0.636852028330519
5995275735	0.31622776601683794	5628587728	0.09205746178983235
6536169960	0.5	5087749666	3.3666666666666667
5661357901	0.5207556439232955	6504766793	1.179535649239177
5958121851	1.3504474832710556	6727846830	0.832455532033676
5908248367	0.15249857033260467	6403830952	2.1824814143238607
6808300044	0.0	1395738842	0.0
2520975127	0.0	5279463384	0.0

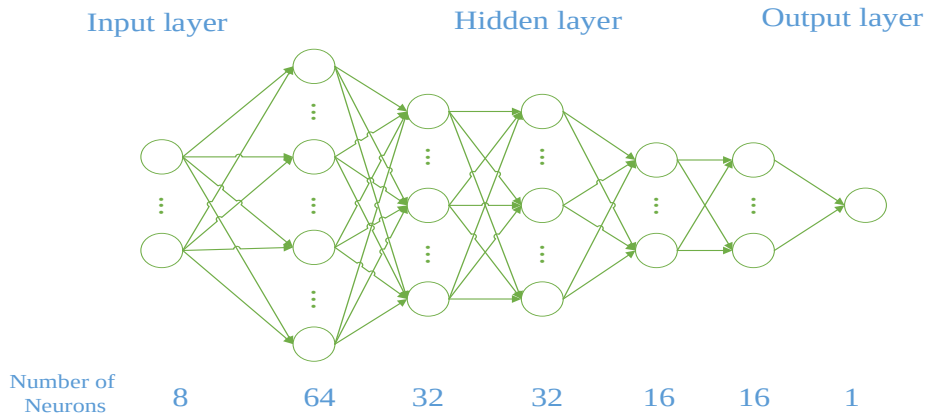
4.2.4 深度神经网络模型实现细节

本文将微博训练集中，目标用户的历史微博作为正样本，将关注用户微博中，出现三次及以上且目标用户没有行为的微博作为负样本。最终的得到正样本 3160 条，负样本 3120 条，共 6280 条，其中 80%作为模型训练集，剩余 20%作为观察集，用于观察模型效果和调参。

经过多次调参训练，得到的最优神经网络模型结构如图。

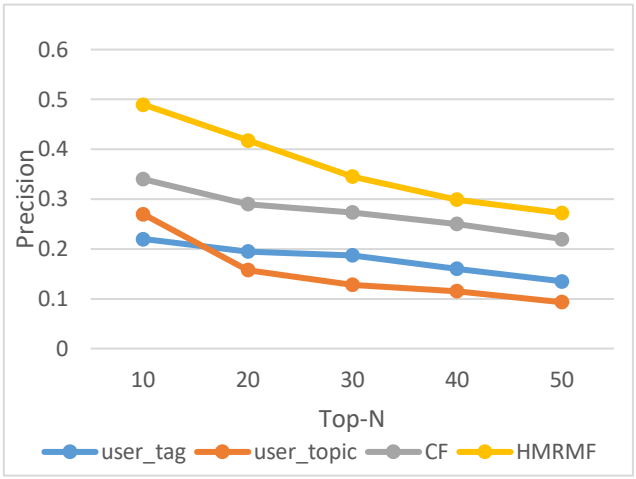
模型包含输入层，五层隐含层和输出层，隐含层结点个数分别为 64,32,32,16,16。模型使用 Relu 作为激活函数，使用 adam 进行权重优化，初始学习率为 0.001，最大迭代次数为 300

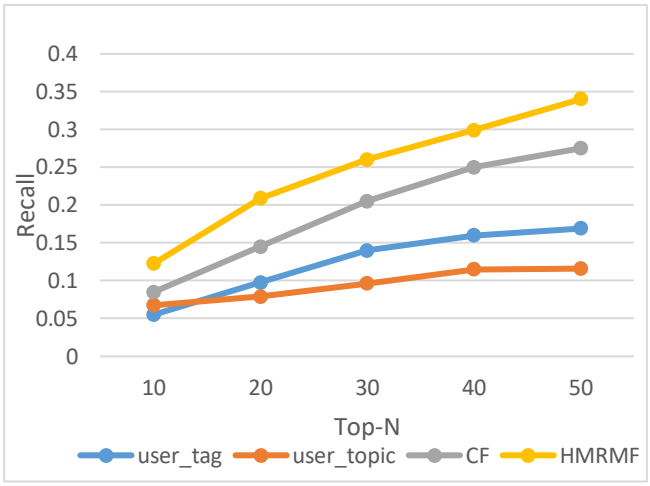
。



4.2.5 混合型推荐算法与单个推荐算法推荐效果对比

本节实验分别将单个推荐算法与总体模型推荐效果进行对比，推荐召回率和准确率结果如下图所示。





Recall

Top	10	20	30	40	50
user_tag	0.055	0.098	0.14	0.16	0.169
user_topic	0.068	0.079	0.093	0.111	0.116
CF	0.085	0.145	0.205	0.25	0.275
HMRMF	**0.135**	**0.214**	**0.264**	**0.304**	**0.346**

Precision

Top	10	20	30	40	50
user_tag	0.22	0.195	0.187	0.16	0.135
user_topic	0.27	0.158	0.124	0.111	0.093
CF	0.34	0.29	0.273	0.25	0.22
HMRMF	**0.54**	**0.428**	**0.352**	**0.304**	**0.277**

由上面结果可以看出，三种单独的推荐算法中，推荐效果：协同过滤>用户兴趣标签>用户兴趣话题。基于用户兴趣标签的推荐与基于用户兴趣话题的推荐虽然都是利用微博内容，但通过分析推荐结果发现，两种推荐算法召回的推荐集重复度较低，说明基于用户兴趣话题的推荐能够从不同角度挖掘用户兴趣，证明了算法的有效性。

从结果中可以看出，模型总体推荐效果与三种单独的推荐算法相比有很大提升，说明模型能够综合每种算法的优点，将用户感兴趣的微博赋予更高的得分，推荐召回率和准确率有明显的提升。

5 结论

本文基于数据挖掘技术，进行个性化微博推荐。模型包括两大部分：待推荐微博获取和

微博排序。首先通过用户历史微博内容挖掘用户兴趣，并且针对用户标签描述不准确问题，采用关联规则进行扩展；并利用微博中包含的丰富的链接信息，发现用户兴趣话题，进行基于用户兴趣话题的推荐；基于用户的协同过滤的推荐利用用户历史交互行为计算用户间相似度。微博排序模块将以上三种推荐算法的召回结果汇总，将推荐的二分类问题转化为排序问题，综合考虑影响用户对微博感兴趣程度的多种因素，使用深度神经网络进行微博排序。本文通过在新浪微博真实数据上进行实验，得出以下结论：采用关联规则 Apriori 扩展用户兴趣标签扩展能够有效提高推荐准确率，使兴趣标签描述更加准确；基于用户协同过滤的推荐算法推荐效果最好；与单个算法相比，将待推荐微博混合并经过神经网络排序后，推荐准确率和召回率有明显提升。