目录

摘要：本项目旨在长文本微博的情感标签预测，不同于传统词向量的文本特征表达，我们提出了一种将长文本分成若干句子，并将句序作为相应的权重的新思想，最后研究了将句序和词向量同时考虑在内的组合模型的预测结果。主要对当前文本分析提出了一种新的特征表示方式，在基于句子顺序会影响句子对长文本的预测重要性的前提下。

1. 项目意义与背景

情感分析一词首次出现即应用于股票市场分析，由 Das 和 Chen 在会议上第一次被提出，在这之前，对于文本信息对股市的影响大多涉及信息量多少或者谣言对股价波动的研究。而他们提出，留言板上的赋有情感的信息也会在一定程度上影响交易行为、市场波幅以及市场效率等，于是他们在雅虎的股票专栏提取留言版上小额投资者的评论短文本，将其分为三类：买入（positive）、卖出（negative）以及中性（neutral）。在实验研究中，他们采用了五个分类器进行训练，最后用投票机制获得分类结果，并对情感分析在实际股票交易中的影响进行了分析，最后发现情感分析这种方法不但可行，而且确实对股票交易量有显著影响。

随着互联网的普及与发展，越来越多的文本信息在网络上产生，其中包括大量的带有感情倾向的主观信息。微博作为我国目前最热门的互联网社交平台之一，因为其发布信息快捷、表达自由、传播速度快等特性被广大网民所青睐，使用者除了普通网民，甚至还包括明星、各行业专家“大 V”和许多自媒体平台。这个基于用户关系的信息分享、获取、传播的平台，包含有数以亿计的大量短文本信息，横跨各类不同领域，对于情感分析任务来说是一个资源丰富的数据集。但由于其操作方便、更新快的一系列特点，也造成了微博文本数据不同于传统文本数据的特性：语言使用不够规范、内容过于简介、含有缩写错写等问题。这些噪声很大程度上阻碍了微博文本情感分析的研究。

在对于微博的文本分析中，最常见的便是情感分类问题，它是指对于自然语言文字中表达的观点、喜好以及态度等进行判别，并用将其应用于实际商业用途等。更具体来说，目前我们更关注于情感极性判断，即是细化到文本所反

映的肯否定、褒贬义的色彩。相对于普通的情感分类，极性判断则是将问题从多分类（例如，喜、怒、哀、乐等）变为了一个二分类或者三分类问题（例如，在投资相关文本中的“利好”、“利空”以及“中性”，电影评论文本中的“好看”和“不好看”），微博平台上就有许多投资者会将带有情感倾向（利好，利空，中性）的投资言论以微博的形式发布于众。我们从中可以挖掘用户的情感、观点，实现股票预测等任务为企业提供决策支持，因此基于投资微博进行情感分析具有重要的商业意义。

2. 项目目标与具体内容

本项目的旨在去探索如何尽可能挖掘微博内容特征，利用机器学习等人工智能的技术实现对于一条投资微博文本的极性分类。不同于传统文本分析，仅仅将。

本项目报告中，我们首先会在第三部分简要介绍进行投资微博情感分类的算法，提出了一种基于句序的情感分类方法；在第四部分，我们会详细介绍在新浪投资微博上的实验过程以及结果分析；最后，我们对于目前的结果做了总结与展望。

3. 项目方法

基于对微博情感文本的分析，我们发现，如图一所示，许多微博虽然单从前几句来看，似乎在表达一个消极/积极的情感，然而，这些可能只是作者对于过去现象的描述，往往在微博的最后作者才会明确的表达出自己的观点与倾向，而这种表述方式的确合理并且常见。

id	content	label
71412	$国元证券(000728)$ 今日早盘该股低开，一方面是受大盘下跌的拖累，一方面是昨日兑现盘今日兑现打压的调整。但从技术面看，该股完全站稳于10日均线之上，5日均线上交10日均线，形成短期的金叉，因而，短期还是看好该股，建议重点关注。	利好

图一 微博样例

因此，基于仅使用传统分类器进行微博情感分类的结果，我们尝试加入句序信息，基本思路为：先分别从投资微博全文中的每一句得到句粒度情感标签，然后将句粒度情感标签利用句子位置给加权组合进来，最终推断出作者在全文表达的最终情感标签。

由此，我们的算法可以大致分为三个板块：训练情感模型、训练句序模型、标签预测

- **训练情感模型**

我们使用原始文本内容作为输入，对应的情感标签作为输出，通过分类训练得到一个情感分类模型（此过程可以用机器学习中的不同分类算法实现），在本文中简称为情感模型；

- **训练句序模型**

我们对每条文本都进行分句处理，首先通过上述情感模型预测出每条句子的情感标签，若句子情感标签与全文本情感标签一致，则记为 1，若不一致，则记为 0，由此可以得到句情感与全文情感的一致性标签。

同时，我们需要得到每一条句子的句序向量。对于每一条句子，我们用一个 10 维向量来记录它在文本中的位置，每一维分别表示为：[是/否是第一句，是/否是第二句，…,是/否是第五句，是/否是倒数第五句，…,是/否是倒数第一句]。例如对于一个有五条句子的微博，第 1 句的句序向量为 [1,0,0,0,0,1,0,0,0,0]；对于一个有 10 条句子的微博，第 3 句的句序向量为 [0,0,1,0,0,0,0,0,0,0]；对于一个有 12 条句子的微博，第 6 句的句序向量为 [0,0,0,0,0,0,0,0,0,0]。

我们将句序向量作为输入，句子的一致性标签作为输出，通过分类训练得到一个句序模型。

- **标签预测**

对于一条投资微博，我们首先通过情感模型预测出其全文情感，然后将投资微博分句，对于每一句通过情感模型预测出每一句的句情感标签，同时，我们使用句序模型预测出每句话的一致性标签，再此可以理解为句子的置信度，并用此置信度对于每个句子进行加权投票得到最终情感标签。

4. 数据预处理

4.1. 初始数据描述

初始训练数据集为约 15000 条由人工标注情感标签的投资微博，情感标签有三类：利空、利好、中性，其中含有利空倾向微博 4076 条，利好倾向微博 6125 条，中性倾向微博 4643 条。比例相对比较平衡，因此不需要进行上下采样等操作平衡数据集。测试集则包含 5000 条数据，缺少情感标签。训练集示例如图二所示，每一条数据包含三个属性：微博 id（id）、微博文本（text）、微博情感标签（label）；而测试集数据唯一与之不同的是缺少 label 属性。

Index	id	text	label
0	5405	【拉菲投资可以买入】近日拉菲价格大跌，上海2003年份拉菲从8660元跌至7740元，去年6月至今，高级葡萄…	利好
1	5409	想向大家推荐一篇分析性很强、概括面很广的文章．$中国银行{601988}$业市场化是大趋势，与非银行业的存…	中性
2	5468	【奇门测股】大盘接近2200点，今天六爻测盘代表官鬼的国家队进场护盘，但还不是买入时机，耐心等候7月7…	利好
3	5488	#开卷有益# 因此，多数散户都没有能力、时间和精神去选股，所以不要奢望跑赢大市，买指数基金即 可有利可…	中性
4	5489	#开卷有益# 葛拉汉《The Intelligent Investor》，曾被巴菲特誉为“最好的投资书”，他学…	中性
5	5493	昨日推荐的书籍《葛拉汉》（The intellegent investor）反应踊跃，@曹晓东_HUST、@理财新室 …	中性
6	5542	#中央大礼#虽自97后，香港已为中国一部分。但因制度上的不同，港人“北上”行商一直有所限制。直至2003年…	中性
7	5548	终于又一次降息了，市场流动资金是增加了，可是，这一举动，意味着什么？意味着6月的经济数据低于预期，意…	利空
8	5552	【丁未月可以入市】明天进入丁未月，之前也讲过这个月股市会上升，沪指能跌到2150点再入货是最理想的。大…	利好
9	237	没错，双赢。沪深两市市值高，但回报率不敢恭维，需要借鉴香港成熟的经验。	中性
10	260	$恒生銀行(000011)$ 香港股市主要指数周涨幅:恒生指数19800.64↗1.85%大市全周成交1849.2亿港元 国…	中性

图二 训练集数据样例

4.2. 数据预处理

对于原始数据，为了方便处理，我们首先去除了其中的部分标点和英文部分，保留了一些重要标点如”#”（在微博中，两个符号”#”中间往往包含了主题等重要信息）。此外，我们利用 jieba 分词对每条微博进行分词处理。最得到的数据样例如图三所示。

Index	Type	Size	Value
0	unicode	1	拉菲 投资 买入 近日 拉菲 价格 大跌 上海 年份 拉菲 从元 跌至 元 去年 高级 葡萄酒 指数 下跌 逾 投资回报 风险 极高 时期 高级 消费品 红酒 ...
1	unicode	1	想 推荐 一篇 分析 性 强 概括 面 广 文章 中国 银行业 市场化 趋势 与非 银行业 存贷 活动 更会 一场 大搏 奕
2	unicode	1	奇门 测股 大盘 接近 点 六爻 测盘 代表 官鬼 国家队 进场 护盘 买入 时机 耐心 等候 日前
3	unicode	1	开卷有益 散户 能力 时间 精神 选股 奢望 跑赢大市 买 指数 基金 即可 获
4	unicode	1	开卷有益 葛拉汉 巴菲特 誉为 投资 书 他学 葛拉汉 选股 思维 学 本书 经是 大市 同步 成绩 买入 指数 基金 跑赢大市 付出 努力 未必 成功 跑赢大 ...
5	unicode	1	推荐 书籍 葛拉汉 踊跃 曹晓东 理财 新室 指出 受益 该书 中 指数 基金定投 人读 好书 受益 投资 知识 请 关注 本微博 张宗永 转发 此条 微博 抽 ...
6	unicode	1	中央 大礼 虽自后 香港 中国 一部分 制度 港人 北上 行商 直至 首次 推出 消除 差别待遇 沪 深港 交易所 设立 合资公司 推出 买卖 基金 视为 中央 ...
7	unicode	1	终于 降息 市场 流动资金 增加 这一 举动 意味着 意味着 经济 数据 低于预期 意味着 经济的 不景气 意料之中 股市 反弹 力度 估计 弱 大盘 处于 探 ...
8	unicode	1	丁未 入市 明天 丁未 讲 股市 沪 指能 跌 点 入货 理想 股市 挑 家电 电子信息 燃气 供水 水力发电 地产 黄金 矿产 做 短线 丁未 戊申 一升 跌 ...
9	unicode	1	没错 双赢 沪深两市 市值 高 回报率 不敢恭维 借鉴 香港 成熟 经验
10	unicode	1	恒生 银行 香港股市 指数 周 涨幅 恒生指数 市 全周 成交 亿港元 国企 指数 红筹 指数 标普 香港 创业板 全周 成交 亿港元 公用事业 恒生 工商 分 ...

图三 数据样例

4.3. 词云展示

我们分别对于三种不同情感倾向的微博文本画出了词云（图四），可以看出，三个不同类别的词云图其实相差不大，占权重最大的词都一样的，因此我们考虑删除出现权值很大的一些词，这些词对分类任务并没有用处。

图四 三类情感标签下的词云

5. 模型

经过第四部分的数据预处理，我们考虑用不同方式和不同模型对微博情感标签进行预测。首先直接将词向量作为训练集进行预测，得到微博标签模型；接着用句序向量作为训练集得到句序模型，结合两个模型进行预测；上述模型都是在考察微博文本情感三分类任务，为了更好的评估模型，我们最后在二分类任务也对模型结果做了检验。

5.1. 微博标签模型

将带标签的数据按照 3：7 的比例分成训练集和测试集，并且保证训练集、测试集和原数据中的利空、利好和中性比例一致。将词向量作为模型最终输入数据，调用 sklearn 包中的多个模型（SVM, Naïve Bayes, Random Forest, Decision Tree），分别对微博情感进行预测，调参之后，准确率如表一所示。此外采用软投票机制集成上述单模型，选取概率最大的标签作为最后预测结果（详见表一中后两行）。可以看到所有模型中，用 SVM+RF 投票准确率最高。

classifier	accuracy
Multinomial NB	0.6108431
SVC	0.6227889
Random Forest	0.6094647
Decision Tree	0.5074661
SVC + RF(soft)	0.6407723
SVC + NB(soft)	0.6253159

表一 模型结果图

5.2. 微博标签模型+句序模型

5.2.1. 三分类任务

- **词序模型的建立**

对于一个带标签的投资微博文本，根据标点符号将投资微博分句，原微博和分句后结果如下表所示，按照上述句序向量表示的规则，我们用十维向量表示每条句子。对于每一句，我们考虑用上述微博标签模型预测句子的情

感结果，将此结果与原微博结果进行比较。如果相同的话，说明此句子对原微博有较大贡献，将其标签设为 1；如果不同的话，说明此句子对原微博没有贡献，将其标签设为 0。此时我们得到了一个新数据集，十维的词序向量和情感贡献标签（0、1），将数据集划分成训练集（37729 条）和测试集（11934 条），利用此训练集得出词序模型。

显然我们此时有两个模型，微博情感标签模型和句序模型，接下来会详细说明我们如何将两个模型结合进行标签预测。对于新的微博文本测试集，我们对每条微博进行分句，首先对分好的句子进行情感标签预测（0，1，2），然后用句序模型对每条句子预测出贡献标签（0，1），下面对于具体一条微博进行分析。对于 id 为 5405 这条微博，初始文本如表二所示，可以分成三条句子，经过上述步骤处理之后，得到如表三所示结果，第四列是句子情感标签，第五列是表示句子贡献的概率，第六列是句子贡献标签，利用第五列概率数据作为权重进行软加权，利用第六列标签数据作为权重进行硬加权，最终结果见表四。

id	content	label
5405	【拉菲投资可以买入】近日拉菲价格大跌，上海2003年份拉菲从8660元跌至7740元，去年6月至今，高级葡萄酒50指数已下跌逾25%。现在各种投资回报都是风险极高的时期，高级消费品如红酒，应该有优于其他投资工具的回报。拉菲这年间已跌了很多，当它和其他一级酒庄价钱进一步拉近，应是可以买入的时机。	1

表二 分句前的微博文本

id	vector	content	label	weight	sen_label
5405	[1, 0, 0, 0, 0, 0, 1, 0, 0, 0]	【 拉菲 投资 可以 买入 】	1	[0.99616229 0.00383771]	1
5405	[0, 1, 0, 0, 0, 0, 0, 1, 0, 0]	近日 拉菲 价格 大跌 ， 上海 年份 拉菲 从元 跌 至 元 ， 去年 月 至今 ， 高级 葡萄酒 指数 已 下跌 逾 。	3	[0.05049965 0.94950035]	1
5405	[0, 0, 1, 0, 0, 0, 0, 0, 1, 0]	现在 各种 投资 回报 都 是 风险 极高 的 时期 ， 高级 消费品 如 红酒 ， 应该 有 优于 其他 投资 工具 的 回报 。	1	[0.81859865 0.18140135]	1
5405	[0, 0, 0, 1, 0, 0, 0, 0, 0, 1]	拉菲 这 年间 已跌 了 很多 ， 当 它 和 其他 一级 酒庄 价钱 进一步 拉近 ， 应是 可以 买入 的 时机 。	1	[0.94100793 0.05899207]	1

表三 分句后的结果

加权投票预测				
label	1	2	3	最终预测标签
硬加权	3	0	1	1
软加权	2.756	0	0.95	1

表四 加权投票结果

- **考虑词序的预测结果分析**

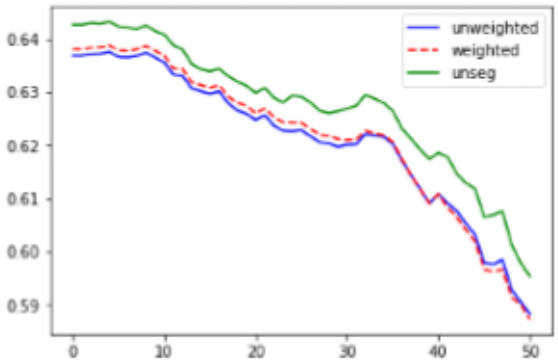
为了更好的研究加入词序向量后的预测结果，我们考虑多种模型的组合结果。对每种模型，我们考虑软加权，不加劝直接计数和不分句三种情况下的预测结果，如表五所示。可以看到在情感标签模型为 SVM，句序模型为 RF 的时候，加权结果高于不加劝结果高于不分句结果，符合我们预期结果；但是在其他组合模型的情况下，结果并不理想，特别是在(RF+SVM)/RF 模型组合中，不分句的准确率是最高的。

三分类(情感/句序)	(RF+SVM)/RF	SVM/RF	RF/(RF+SVM)	NB/RF	SVM/logistic	SVM/GBTree
weighted	0.638055222	**0.62284914**	0.614045618	0.629251701	0.621648659	0.621448579
unweighted	0.636854742	**0.620648259**	0.613445378	0.628451381	0.62344938	0.62344938
unsegment	0.642657063	**0.615246098**	0.633853541	0.629451781	0.615046018	0.615046018

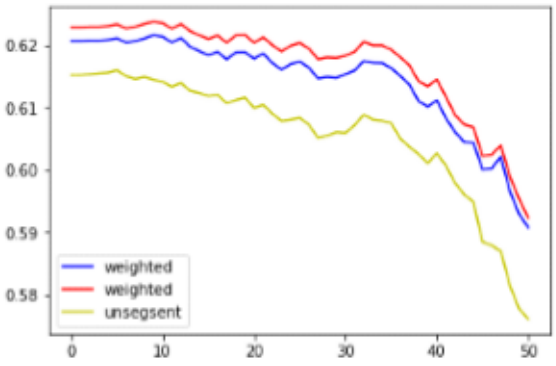
表五 模型在三种情况下的预测结果

- **删除短字数文本**

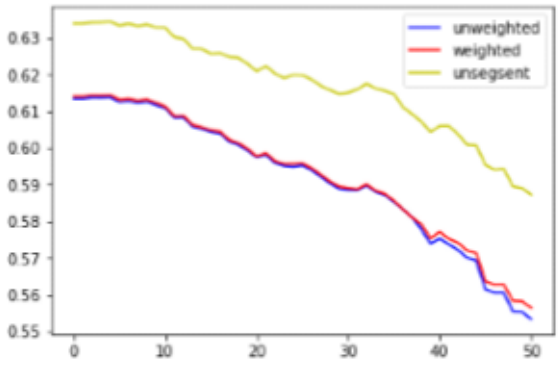
上述模型结果并不理想，我们觉得有可能是数据的不均衡导致了模型间的差异。具体来说就是，很多微博都是短句子，导致预测可信度并不高，这些微博的预测结果反而会变差，因此考虑如果遇见一条短句子，就将这条句子所在微博给删除，这样就保证了所有微博都是只含有长句子的文本，进行分句结果可能更好。各种模型组合结果如下所示：



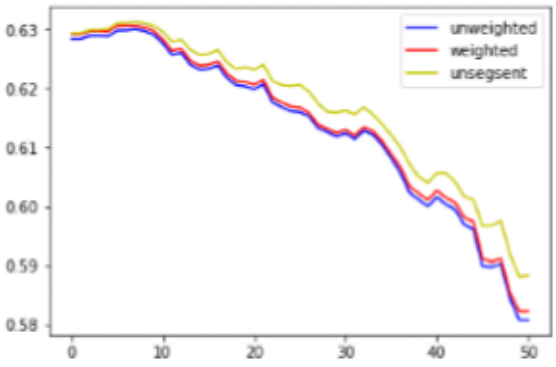
图五 RF+SVM/RF



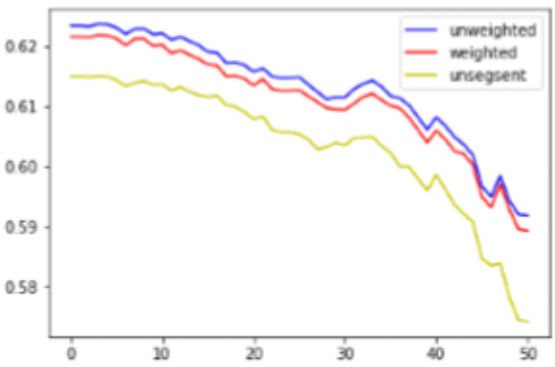
图六 SVM/RF



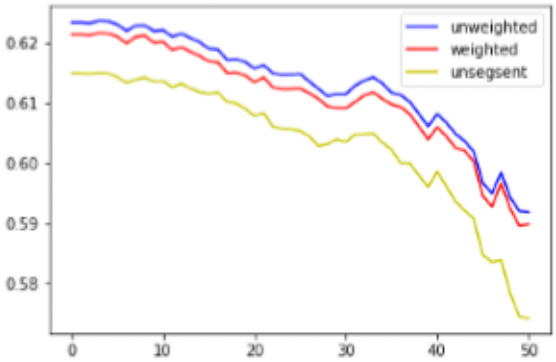
图七 RF/RF+SVM



图八 NB/RF



图九 SVM/logistic



图十 SVM/GBTree

上述六幅图分别对应表五中的六种模型组合的结果，可以发现删除含有短句子的微博之后效果并没有提升很多，表现好的模型组合依旧效果好，表现不好的模型仍然效果不好。

5.2.2. 二分类任务

➢ **二分类任务的句序模型**

如表六所示，通过输出对于情感模型预测后生成的混淆矩阵（0：利空，1：利好，2：中性），我们发现中性文本的模糊性较大，错误大多出现在将中性文本预测错误或将情感倾向不明显的利好/利空文本预测错误为中性文本，这使得情感模型预测结果较差。由于我们的主要目的是想观察句序模型的表现情况，而句序模型中加权投票进行预测时大大依赖于情感模型对句子预测的情感标签，如果置信度大的句子情感模型预测结果错误，在加权时会被赋予较大权重，那么会影响整个句子的预测结果。因此，为了提高情感模型的预测结果，我们删除文本中的中性文本，剩余训练集为 10201 条，剩余测试集为 3501 条，其他设置不变，再次进行实验。

Confusion Matrix			
真实值	预测值		
	0	1	2
0	726	248	249
1	132	1453	253
2	255	463	675

表六 情感模型预测结果的混淆矩阵

对于一个带标签的投资微博文本，我们依旧对其进行分句、添加句序向量以及情感贡献标签等操作，将数据集划分成训练集（25491 条）和测试集（7828 条），利用此训练集得出词序模型。后续流程如上，最终结果见表七所示。

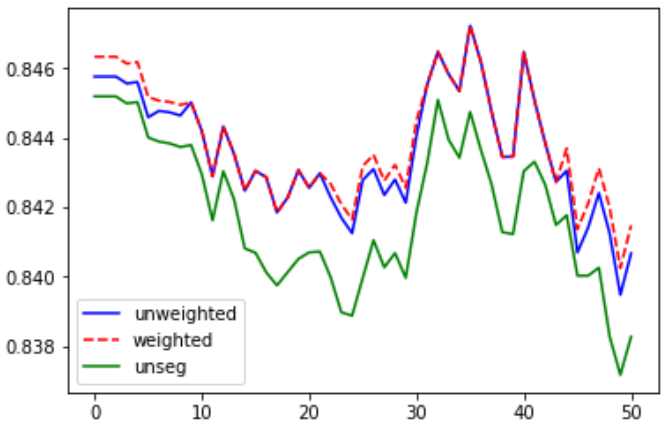
二分类总结（情感/句序）	SVM/RF	SVM/(RF+SVM	(RF+SVM)/RF	SVM/SVM	RF/SVM
weighted	0.838617538	0.83776064	0.84632962	0.838331905	0.839474436
unweighted	0.838617538	0.838617538	0.84575835	0.838617538	0.839474436
unseg	0.834904313	0.834904313	0.84518709	0.834904313	0.84204513

表七 模型在三种情况下的准确率

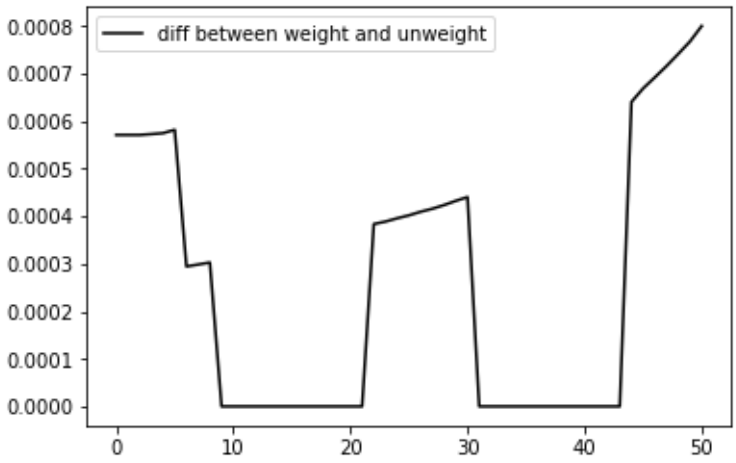
通过结果分析我们发现：

1. 对于二分类，加权或者不加权的结果都没有太大差别，这可能是由于对于二分类问题结果已经比较好，难以有大的提升。
2. 以 SVM 做情感模型分类的，情感模型结果较为不好，但权重模型相比于原来的结果就有大的提高。（0.003~0.004）
3. 对于情感模型表现较好的，加上权重模型之后反而没有提高，甚至有所下降。这可能是由于，权重模型的准确率并不够高，导致情感模型高准确率的结果被影响，反而预测错误。

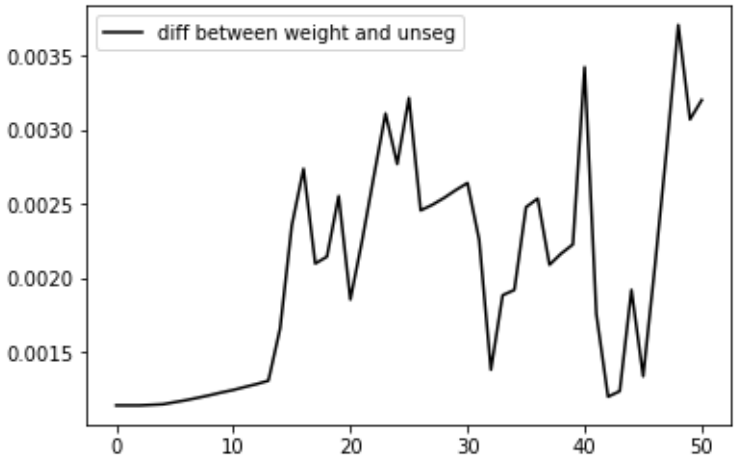
相比较来看，情感模型用 RF+SVM， 句序权重模型用 RF 的结果是比较理想的。最后得到的对比图如下所示



图十一 RF+SVM/RF



图十二 加权与不加权的差别图



图十三 加权与不分句的差别图

- **回归任务的句序模型**

我们在句序模型阶段采取了分类模型，软投票时采用了句子为真（即标签为 1）的置信度作为权重加权投票。

此种句序模型进行训练时使用的标签为 0/1，模型采用的是分类任务。但由于我们进行软加权时考虑的是连续型的权重。因此我们考虑建立一个回归任务。

我们在预测完句子的情感标签后，会得到三列情感标签置信度，若微博全文情感标签为 1，我们则选择三列数据中为 1 的置信度作为该句的情感贡献权重。然后，我们将句序向量作为输入，此时连续型的情感贡献权重作为输出，进行回归模型拟合。

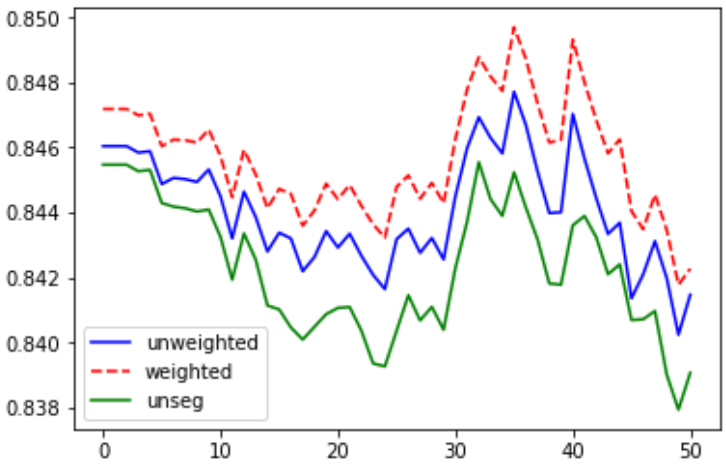
在预测阶段，我们利用新的句序模型通过句序向量预测出句子的情感贡献权重，然后进行加权投票得到最终的预测结果。我们在二分类数据集上进行了实验，结果如表八所示。

回归二分类总结（情感/句序）	SVM/RF	(RF+SVM)/RF	RF/SVM
weighted	0.83947	0.847186518	0.83862
unweighted	0.83833	0.846043987	0.83947
unseg	0.8349	0.845	0.84205

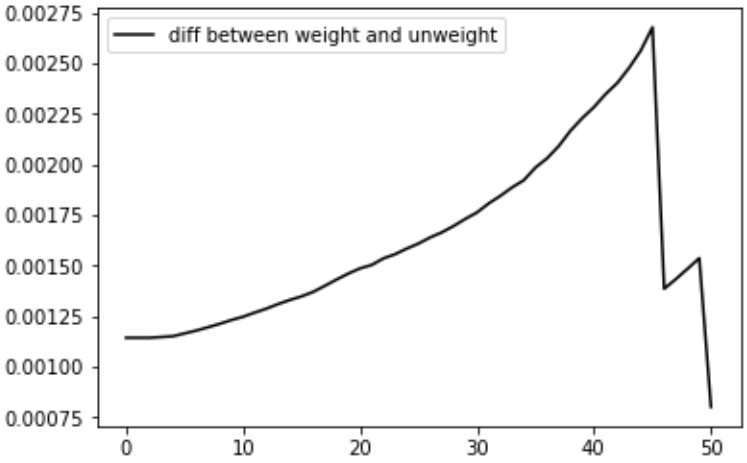
表八：回归任务下的模型结果图

最终通过比较结果，我们发现：

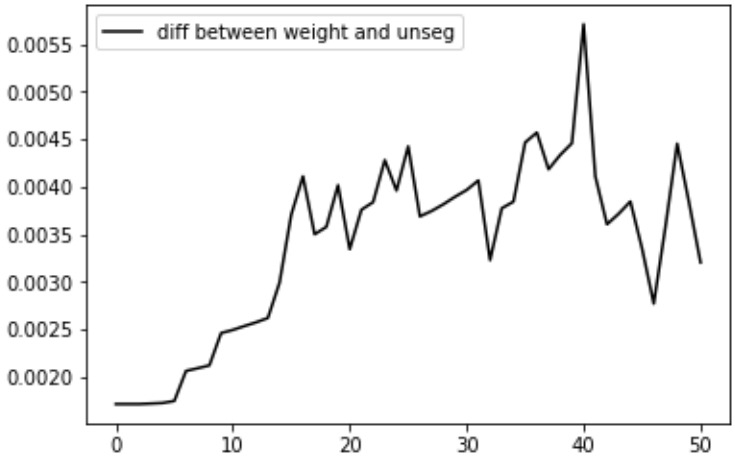
1. 对句序模型换为回归任务之后，在所有情况下，不加权投票和加权投票的区别都被拉开；
2. 对于某些模型，加权的结果从比较差变为了比较好，但是也有模型未能变化；
3. 对于之前最好的情感模型使用(RF+SVM)投票，句序模型使用 RF 回归，整体结果都提高，并且，不加权的结果和加权的结果的区别被大大拉开。结果如下图所示。



图十四 (RF+SVM)/ RF 回归任务



图十五 加权与不加劝的差别



图十六 加权与不分句的差别

6. 总结与展望

在本项目中，我们基于微博文本的特征提出了一种融合句序信息的情感分类模型。我们通过训练得到了每条句子的可信度标签，从而通过加权投票的方法预测出一条投资微博的情感倾向。在句序模型中，我们获得了 10 维向量的每一维的重要程度，如表九所示，不难发现最后一维（显示含义为倒数第一句）的重要程度远高于其他维度。这在某种程度上也可以印证我们对于句序不同时所表达的情感倾向对全文情感倾向预测的贡献程度也会不同的猜想。

clf.feature_importance					
句序向量	1	2	3	4	5
重要性	0.0405	0.1635	0.0932	0.0453	0.0370
句序向量	-5	-4	-3	-2	-1
重要性	0.0347	0.0429	0.0493	0.0774	0.4151

表九 重要性程度

此外，可以看出，在情感模型的分类器表现较差时，句序模型能很好的对模型起到提升作用；但是当情感模型较好时，句序模型的提升作用不是特别明显。通过分析讨论，我们提出了一些未来的改进方向并正在尝试中，比如更换数据集或者更换句序表达方式进行尝试。

参考文献

[1]周胜臣，瞿文婷，石英子，等. 中文微博情感分析研究综述[J] . 计算机应用与软件, 2013, 30 (3) : 161-164.

[2]Bifet, A., & Frank, E. (2010, January). Sentiment knowledge discovery in twitter streaming data. In Discovery Science (pp. 1-15). Springer Berlin Heidelberg.

[3]Hanzhe Li. Sentiment Analysis and Opinion Mining on Twitter with GMO Keyword. North Dakota State University. 2016.

[4]Kouloumpis E, Wilson T, Moore J. Twitter sentiment analysis: the good the bad and the OMG [C]//Proceeding of the Fifth International Conference on Weblogs and Social Media, 2011:17-21.

[5]姜杰,夏睿. 机器学习与语义规则融合的微博情感分类方法[J]. 北京大学学报(自然科学版),2017,(02):247-254.

[6]Turney Peter. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Meeting of t he Association for Computational Linguistics [C] . 2002. 417- 424.

[7]Sanjiv Das and Mike Chen. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA), 2001.

[8] Dhar, Lisa, Schnoes, Melinda G, Wysocki, Theresa L,et al. Erratum:

“Temperature-induced changes in photopolymer volume holograms” [Appl. Phys. Lett. 73, 1337 (1998)][J]. Applied Physics Letters, 1998, 73(15):2220-2220.

[9]Bagnoli, Beneish and Watts (1999) examined the predictive validity of whisper forecasts, and found them to be superior to those of First Call analysts.

[20]周立柱,贺宇凯,王建勇. 情感分析研究综述[J]. 计算机应用,2008,(11):2725-2728.

[10]Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv: 1301.3781, 2013.

[11]Richard Socher, Jeffrey Pennington, Eric Huang, et al. Manning Conference on Empirical Methods in Natural Language Processing(EMNLP 2011, Oral) Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. 2011.