## 修改内容总结：

比较版本，系统提交版和当前版本。

1. 英文题目翻译
2. 摘要段间距
3. 摘要“解决上述问题”
4. 摘要第二、三段的组织
5. 摘要第三段“首先”
6. 去掉“大数据可视化关键词”
7. 英文摘要对应上面的问题
8. 绪论帽子
9. 研究背景精简（删除太具体工作）
10. 非理性繁荣参考文献
11. 预测泡沫的准确表述
12. 研究背景由京东换为美联航的例子
13. 实时性与具体性改为“预测的实时性与预测对象的具体性”
14. 研究背景总结点题
15. 研究目标再次点题
16. 研究内容删除无用内容，仅保留自己工作的梗概，与原研究目标重合部分全部删除，研究目标不再作为分节
17. “传统情感分析算法难以胜任该数量级数据的分析工作。有可靠性保证的高并发流处理平台才能满足本课题的需求。”算法改为算法平台
18. “根据计算，如果对某些股票的预测准确率达到60%，投资这些股票的产生风险便会下降20%[12]。所以相对比于研究传统情感分析算法时较高准确率的要求，本论文对算法准确率要求并不高” 表述不清意义不大，删除
19. 主流的研究方向添加“分别为随机游走模型以及非理性模型”
20. 1.2.1的海量改为实时
21. 将第五章非自己的内容全部移到第二章相关原理技术部分
22. 针对所有需要加参考文献的相关原理技术全部加了参考文献
23. 相关技术与系统所需要的特性相互结合叙述，体现技术的必要性
24. 2.2推特数据改为文本数据
25. 系统的需求分析按照功能-非功能需求重新整合
26. 删除了数据层需求中啰嗦的部分
27. 在Netty报文头添加关键字信息，补充说明了来源
28. 修改了表格格式问题，包括边框和后面的断续列表

“根据本课题研究目的对系统组成进行拆分，系统功能可划分为以下6个子功能：”改为“根据业务层功能的具体需求，预测系统需要满足获取、存储、处理、分析、展示等多个系统层的功能，具体如下：

1. 解释QPS
2. 相对于提交部分重新进行了设计，分层主要依照第三章的分层结构
3. “如何保证流数据能够持续地背压输出（即数据消耗能力满足数据生产的需求），这是系统是否满足可用性的关键。”补充说明
4. “如何解决数据库IO的压力” 确实存在该问题并且读写分离解决

“数据获取层完成获取股价数据、推特数据获取。

数据处理层完成数据的清洗、筛选（预处理）。

数据传输层完成数据的跨主机传输。

数据计算层完成情感数据的计算以及股价预测计算。

数据存储层完成数据的存储。

数据展示层完成计算结果的展示。

六个层次的结构大致如图4-3所示：六段变一段

1. ” 技术架构示意图重新画了紧凑的
2. “流数据形式的社交网络文本数据，”改为“流形式的...”
3. 如19所述第五章整体只留下自己的算法设计部分
4. 修改了公式编号没有对齐的问题
5. “前人word2vec算法的工作以及已经成熟的投入生产的工业化方案使得该算法已经有比较好的表现。”修改表述方式“在第二章提到的word2vec算法的先前工作以及已经成熟的投入生产的工业化方案已经有比较好的表现。”
6. 第五章严格叙述了哪些是他人的工作，哪些是自己的工作
7. 第五章仔细修改了小节开头原表述“利用动态淘汰策略的在线 SVM 算法，可以自适应的调整模型的更改速度， 而不用重新计算所有数据，但是存在由于参数设置问题导致模型更新过快，旧的 数据尚未产生其足够的舆论影响即被淘汰的情况，或者历史数据积累的效应无法 在时间上对模型产生完整影响。在此基础上，本文提出使用一种新的在线 SVM 算法，称为在线被动攻击 SVM 算法，通过对目标函数进行变形，从而能达到满 足一定条件下最小化每一条数据对模型产生的影响，使得每一条数据都能产生其 对模型的影响而不使得历史数据的影响被快速覆盖。 将”

新表述“利用动态淘汰策略的在线SVM算法，可以自适应地随着新到来数据调整模型，而不用重新计算所有数据。但是该算法存在由于参数设置问题导致模型更新过快，旧的数据尚未产生其足够的舆论影响即被淘汰的情况。该算法也会出现历史数据积累的效应无法在时间上对模型产生完整影响的情况。在此基础上，本文提出使用一种新的在线被动攻击SVM算法，通过对目标函数进行变形，从而能满足在一定条件下最小化每一条数据对模型产生的影响，且历史数据的影响不被快速覆盖。”

还包括其他小节。

1. 图5-1修改参数的位置
2. 基于时间序列的股价预测算法改为基于情感分析的时间序列股价预测算法
3. 严格区分了“舆论”与“舆论情感”这不一样的概念
4. 参变量*a, b*的组合有多种可以选择
5. 数据集的字段大致如下 删除文中所有大致
6. 第六章存在部分表述语病情况，已修改。如“对于基于词袋的PCA+TF-IDF降维算法，并采取LR分类的进行评估，并验证了其准确率，AUC，以及F1分数”
7. 总结其算法误差效果可以得到以下表格
8. 修改了图名
9. 规范了“可以看到”口语表述
10. 除了American公司的预测结果中两个算法误差均比较大之外
11. “其逻辑框架如图7-4所示”更改了所有类似表述
12. “数据计算模块的部署按照算法与主题进行物理区分，计算不同的主题的实例分布在不同的机器上。”更容易理解的表述
13. “Web模块主要由Spring MVC搭建，按照Web项目的架构分为视图（View）、控制器（Controller）、服务（Service）三个部分：”增加中文表述
14. 第七章示意图都进行了紧凑处理
15. 系统测试开头重新组织了语言“本章主要对系统功能进行详细的测试，测试需要优秀完备的测试环境进行支持，测试主要包括功能性测试以及性能测试两个方面。功能测试测试系统主要模块的功能，包括单元功能测试和整体功能测试。性能测试主要包括系统的压力测试。”
16. 8.3.2.3 整个小节表述优化
17. 修改图8-1的表述
18. 8.3.2 删除大
19. 第九章重写，叙述完成的论文工作，捎带一提系统设计工作。删除工作不足汇总进入未来工作。
20. 参量引入股价预测，并得到预测结果
21. 新加了更多的参考文献，参考文献按照格式给出的参考文献格式进行了统一修改
22. 修改参考文献的标号
23. //todo 补充参考文献12