



一种基于支持向量机和主题模型的 评论分析方法RASL

Review Analysis method RASL based on SVM and LDA

陈琪¹，张莉^{1,2}，蒋竞²，黄新越²

1 北京航空航天大学软件学院

2 北京航空航天大学计算机学院

目录

- 研究动机
- 研究方法
- 实验验证
- 论文总结

目录

- 研究动机
- 研究方法
- 实验验证
- 论文总结

研究背景



- 用户抱怨评论 -> 软件需求
- 移动应用程序的评论

好评



360用户 好图

2016-08-26 12:58:24

差评



张夷希 广告越来越多了，现在看视频，翻一页就能看到视频，都是些垃圾游戏视频。

2016-08-26 12:39:05

中评



沉默默杀 闪退啊

2016-08-26 08:38:10

差评



大雨治水 用了几个月了，越用新闻越假！还有一大堆垃圾新闻。。。无爱了！

2016-08-26 08:08:02



- 用户评论的数量庞大，并且是无结构的
- 手动检查耗时且低效

开发者希望用户抱怨信息的**核心**内容**直观**的展现

如何对用户抱怨信息进行信息挖掘？



方法一：

- 对用户评论进行分类

- 优势：

- 从用户评论中发现评论的种类，帮助开发者了解用户关注的要点

- 劣势：

- 分类方法能够了解用户遇到的问题种类，但很难得到评论中针对的软件特征

- 比如：Maalej等人提出了14种类型的问题

- Zhang等人将用户抱怨评论分为17个类型

相关工作



方法一：

- 对用户评论进行分类

- 优势：

(1) “**页面**中的**按钮**没反应。”

从用
户的
注的

(2) “应该在**页面**中添加一个**按钮**。”

了解用户关

.....

- 劣势

分类方法能够了解用户遇到的问题种类，但很难得到评论中针对的

问题类型1：软件错误

问题类型2：请求增加功能

- 比如：Maal

.....

题

Zhang等人将用户抱怨评论分为17个类型

方法二：

- 采用主题提取的方法分析评论
 - 优势：
得到包含用户意图的信息
 - 劣势：
能够得到特征信息，但很难区别用户的意图
- 比如：Blei等人提出LDA模型
Galvis等人提出ASUM模型

相关工作



方法二：

- 采用主题提取的方法分析评论

- 优势
 - (1) “**页面**中的**按钮**没反应。”
 - (2) “应该在**页面**中添加一个**按钮**。”
- 劣势
 -

能够得到特征信息，但很难区分用户的意图

- 比如：Blei等 页面、按钮
- Galvis等

研究目标（1）



对用户评论进行分类

主题提取的方法分析评论

结合

本文：

- (1) 对用户评论进行分类
- (2) 对每个类别的评论进行主题挖掘
产生主题的代表词以及代表性句子

研究目标（2）



举例：

- (1) “**页面**中的**按钮**没反应。”
- (2) “应该在**页面**中添加一个**按钮**。”

.....



问题类型1：软件错误

- 主题1：页面、按钮

.....

问题类型2：请求增加功能

- 主题1：页面、按钮

.....

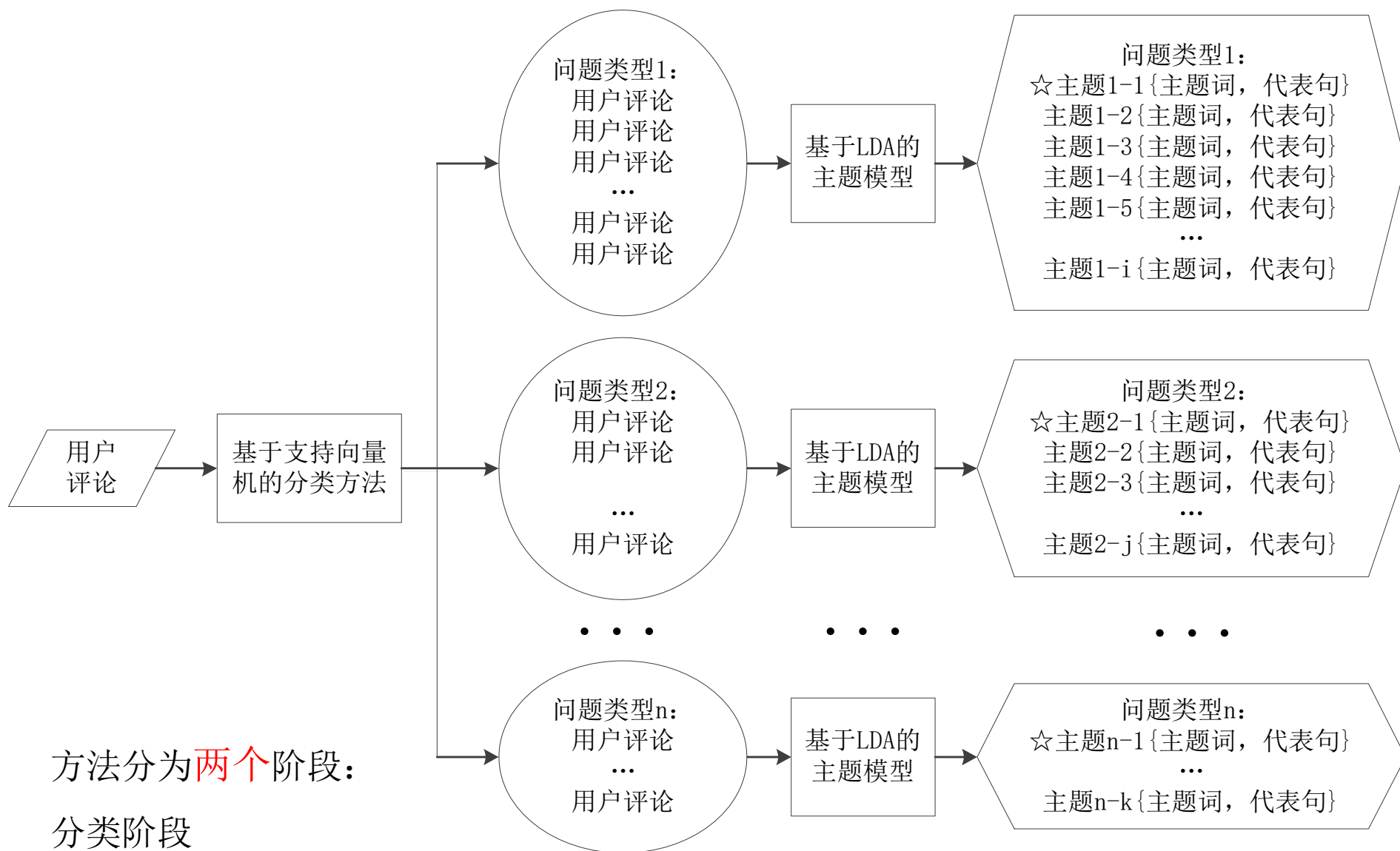
.....

分类+主题分析

目录

- 研究动机
 - 研究方法
- 实验验证
- 论文总结

评论分析方法RASL

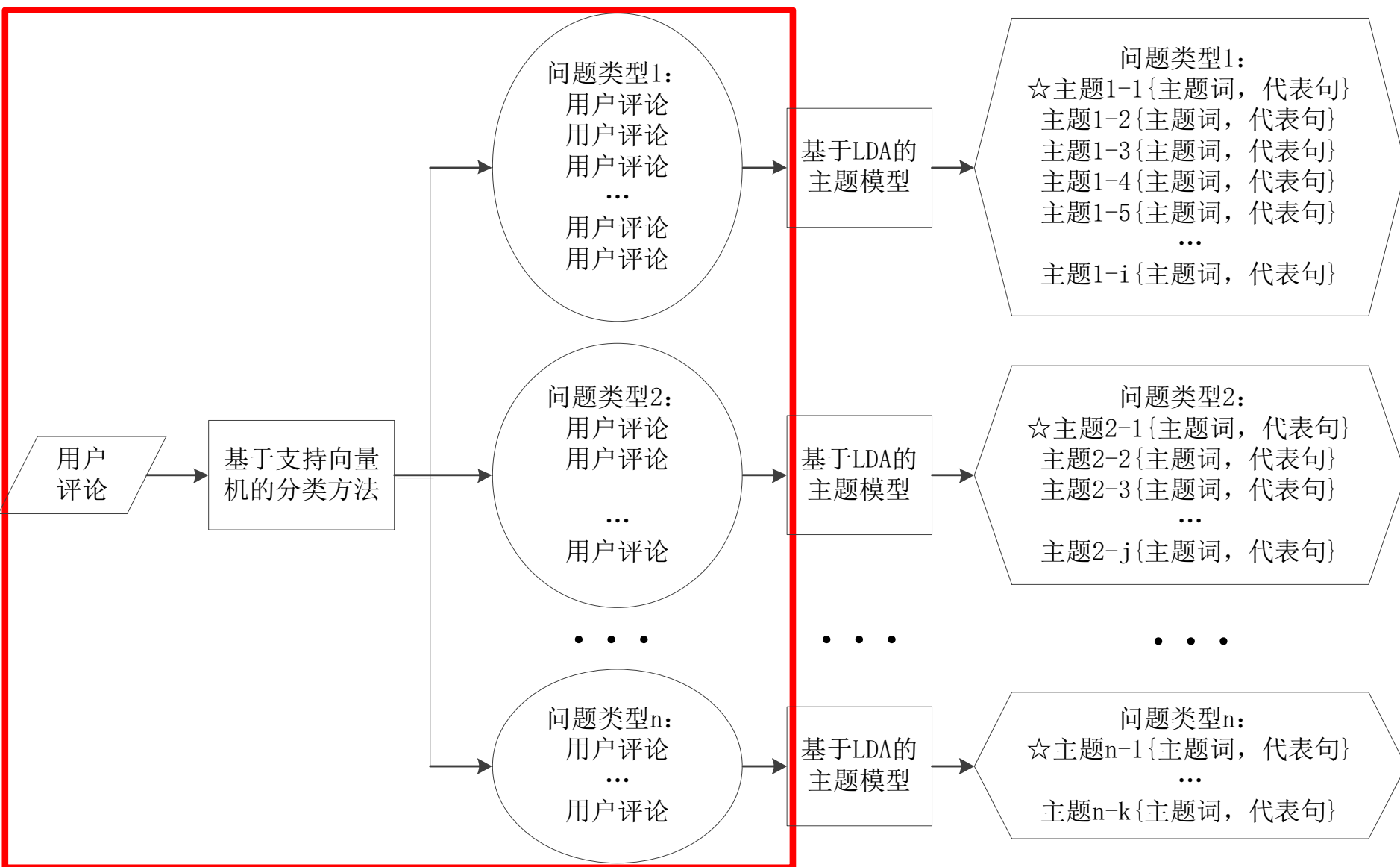


方法分为两个阶段:

分类阶段

主题分析阶段

研究方法——分类阶段



方法分为两个阶段：分类阶段；主题分析阶段



用户抱怨评论的分类类型

- 三人手工并行完成对问题类型的标注;
- 采用组内相关系数 (ICC) 对标注结果的可靠性进行度量

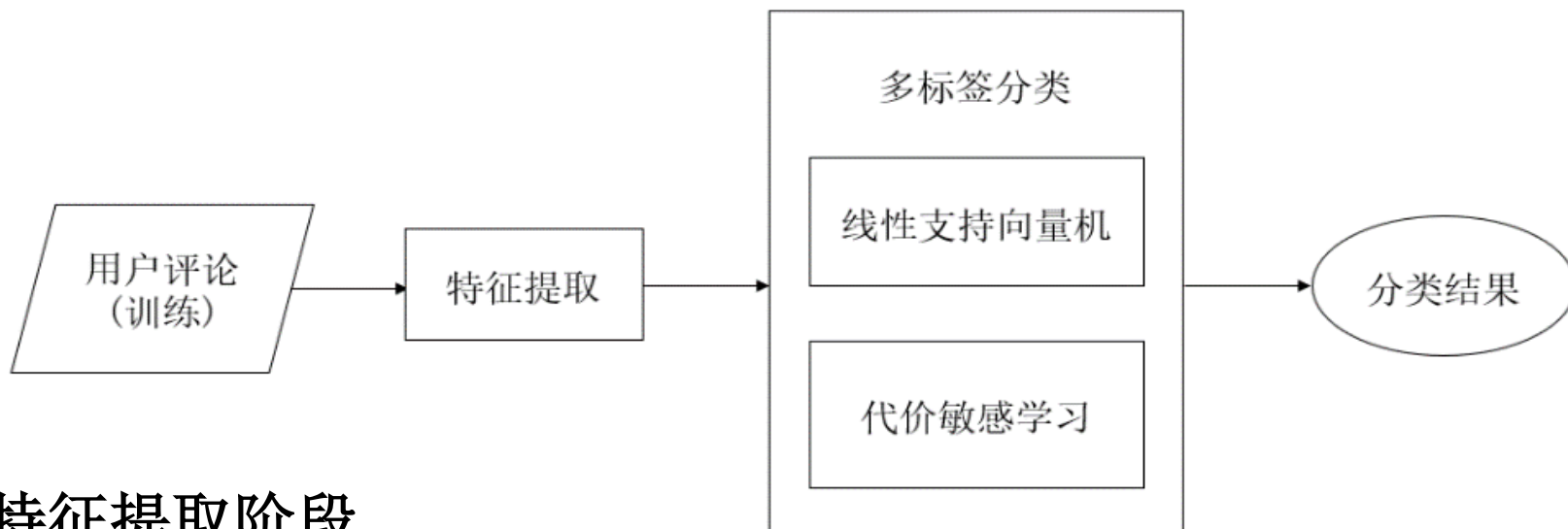
研究方法——分类阶段(1)



评论类型

问题类型	描述
额外开销	抱怨需要额外的花费来享受完整的体验
兼容问题	应用在某个设备或操作系统上有问题
内容抱怨	内容没有吸引力或缺乏某些内容
崩溃	应用出现崩溃
移除特性	应用的某些特性非常糟糕
增加特性	希望应用增加某些特性
功能问题	应用的功能出现异常或失败
内容问题	在安装应用时失败
网络连接	应用出现网络连接方面的问题，如延迟
隐私与道德	应用侵犯用户隐私或者不道德
财产安全	应用威胁到用户的财产安全
资源占用	应用占用了太多电量、内存等
响应时间	应用响应缓慢
流量浪费	应用使用了超出用户预期的流量
更新问题	用户抱怨更新带来了新的问题
界面交互	对界面设计、交互、视觉方面的抱怨
其他	无用的，没有指出问题的评论

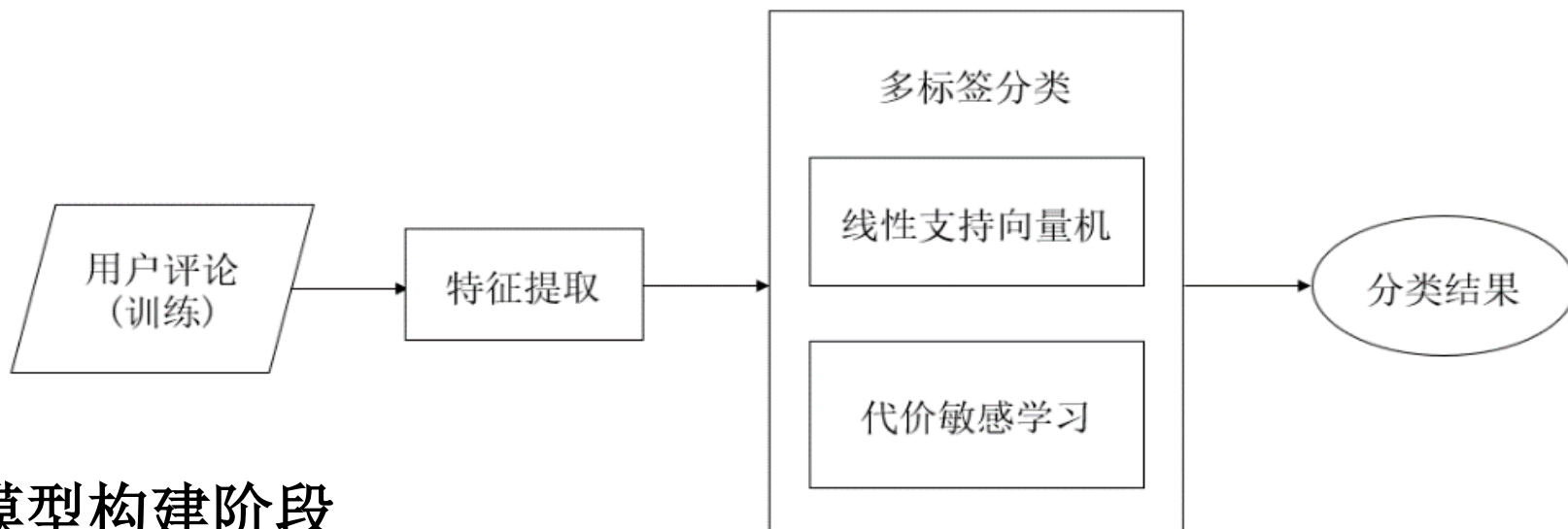
研究方法——分类阶段(2)



特征提取阶段

- **目标：**提取评论文本的特征
- **输入：**用户评论
- **输出：**特征向量
- 运用向量空间模型（VSM）将文本转化为特征向量组成的向量空间

研究方法——分类阶段(3)



模型构建阶段

- **目标：** 构建一个多标签的分类模型
- **输入：** 特征向量
- **输出：** 评论所属的问题类型
- 本文选择使用线性支持向量机进行分类，将多分类问题转化为多个二分类问题

研究方法——分类阶段(4)



代价敏感矩阵

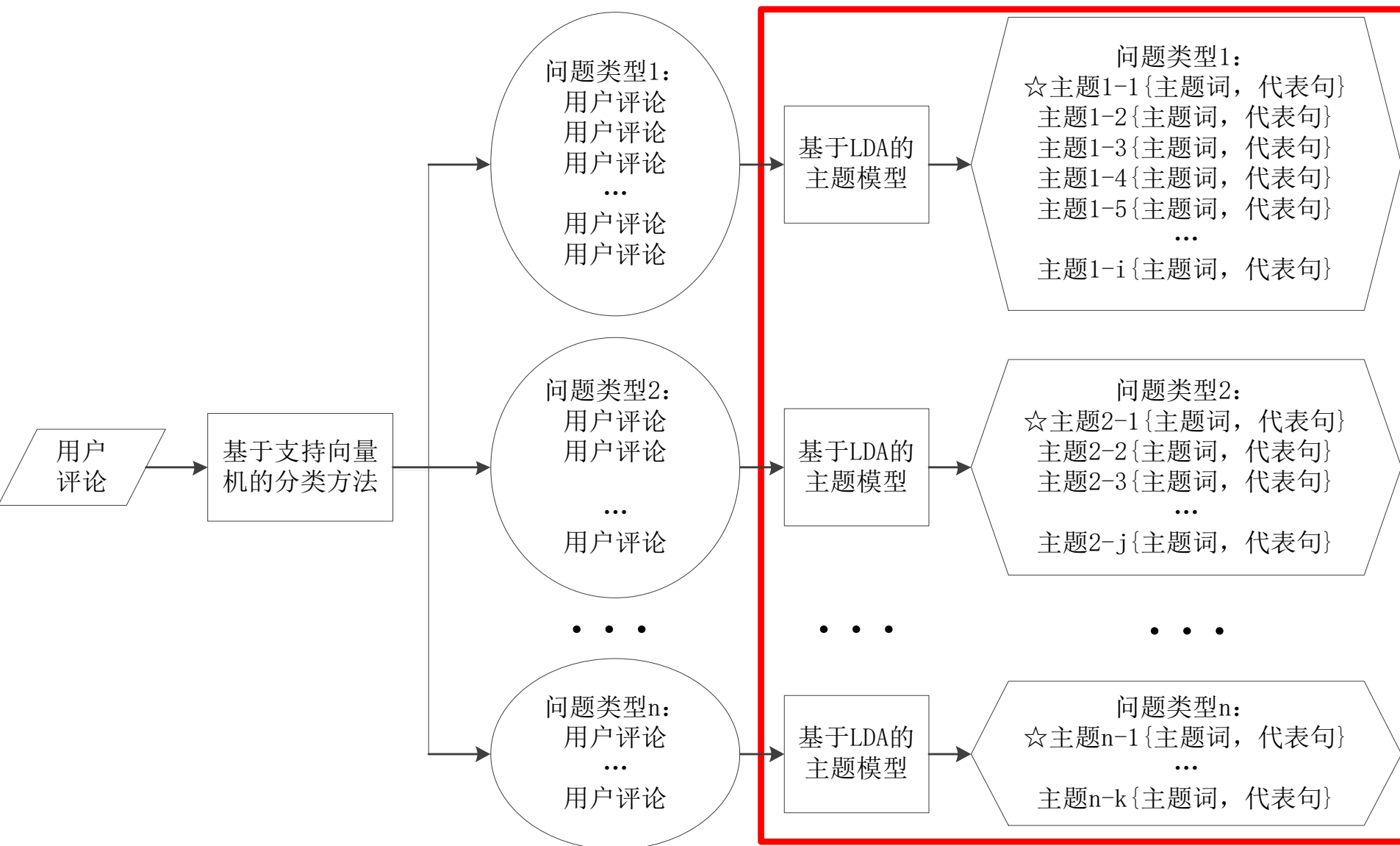
- 对于一些问题类型，负样本的数量远大于正样本
- 这些不平衡的数据可能导致分类器更倾向于将新样本预测为负样本，为了减轻不平衡数据的影响采用代价敏感学习的方法

预测 \ 真实	0	1
	0	1
0	C_{00}	C_{01}
1	C_{10}	C_{11}

详细过程见我们前期工作：

an approach for labelling mobile app reviews. 计算机科学技术学报(英文版)2017 19

研究方法——主题分析阶段



方法分为两个阶段：分类阶段；主题分析阶段

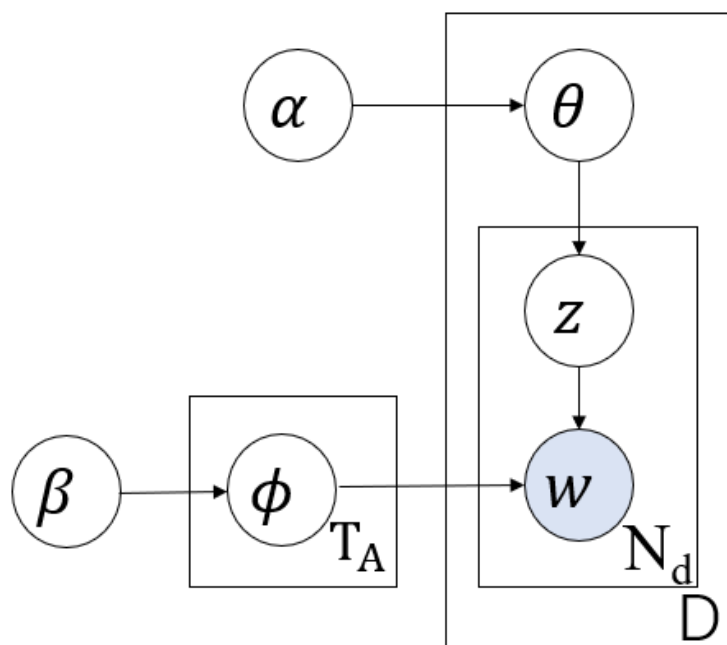


主题数的设定

- 问题类型根据分类的结果中各自所占比例计算出每个问题类型对应的主题数目，后续再对各问题类型下的评论进行主题分析
- 主题总数为**T**，分类A的评论数量占总数量的比重为**Ratio**，则分类A的主题数为 **$T * \text{Ratio}$**



主题词与代表句的生成



θ 是文档的主题分布
 α 是主题分布 θ 的Dirichlet分布的超参数
 z 是文档中词语的主题
 N_d 是主题 z 下的全部词语数
 D 是文档数目
 ϕ 是每个主题对应的词语分布
 β 是每个主题的词分布 ϕ 的Dirichlet分布的超参数
 T_A 是主题数目
 w 是生成的词语

- 本文采用了LDA模型来进行主题词和代表词的生成
- 这个模型是一种典型的词袋模型，即一个评论是由一组词构成，不去考虑词语的顺序，因而简化了语义关联问题的复杂性



主题词与代表句的生成



主题词与代表句的生成分为三个阶段：

主题生成阶段

根据阈值挑选关键词阶段

代表句生成阶段



T_A 是主题数目
 w 是生成的词语

- 本文采用了LDA模型来进行主题词和代表词的生成
- 这个模型是一种典型的词袋模型，即一个评论是由一组词构成，不去考虑词语的顺序，因而简化了语义关联问题的复杂性



主题生成阶段

- **输入：** 分类的评论 **输出：** 指定数目的主题
- **基本思想：** LDA通过词库自动分析分类所有评论得到对应主题，选择最相关的主题作为该分类下的主题

根据阈值挑选关键词阶段

- **输入：** 主题生成阶段得到的分类A下的 T_A 个主题
- **输出：** 每个主题下指定数目的主题词
- **基本思想：** 在主题词生成时，选择对应主题下概率值最高的词汇作为主题词



代表句生成阶段

- 输入：根据阈值挑选关键词阶段得到的主题词
- 输出：每个主题一下的代表句
- 基本思想：
 - d 向量为 $[d_1, d_2, \dots, d_n]$ (主题数为 n)
 - 每个 d_i 代表一条用户评论被分配到每个主题的概率

假设：第一条评论的向量 $d_1 = [0.75, 0.05, 0.05, 0.15]$

第二条评论的向量 $d_2 = [0.2, 0.75, 0.001, 0.049]$

第三条评论的向量 $d_3 = [0.05, 0.2, 0.15, 0.6]$

则对于目标主题1，评论 d_1 作为目标主题的代表句

研究方法——展示结果



- 按照每个类别下评论的数量从多到少排序

- 按照每个主题下评论的数量从多到少排序

方便阅读者对提出较多的问题进行重点关注

分类类型1

☆主题1-1:

主题词

代表句

主题1-2:

主题词

代表句

...

主题1-i:

主题词

代表句

⋮

分类类型16

☆主题16-1:

主题词

代表句

...

目录

- 研究动机
- 研究方法
 - 实验验证
- 论文总结

实验验证——实验对象



- 比较对象：经典方法ASUM
- ASUM方法是一种结合情感分析的主题模型
 - 将句子看作文档，句子中每个词都是隐含主题的分布，然后进行主题挖掘。
 - 在此基础上，融合主题特征和情感信息来分析用户对这些主题的偏好，并以<主题词，代表句>序对作为输出
- 实验对象：

360手机助手中随机抽取一个评分高的应用(评分9以上)和一个评分低的应用(评分6以下)



RQ1: ASUM方法和本文方法RASL的困惑度如何？

RQ2: 和原始评论相比，ASUM方法和本文方法RASL是否包含完整的信息？

RQ3: 和原始评论相比，ASUM方法和本文方法RASL是否包含冗余的信息？

RQ4: ASUM方法和本文方法RASL的阅读理解性如何？

RQ5: ASUM方法和本文方法RASL的评论阅读时间如何？



困惑度分析

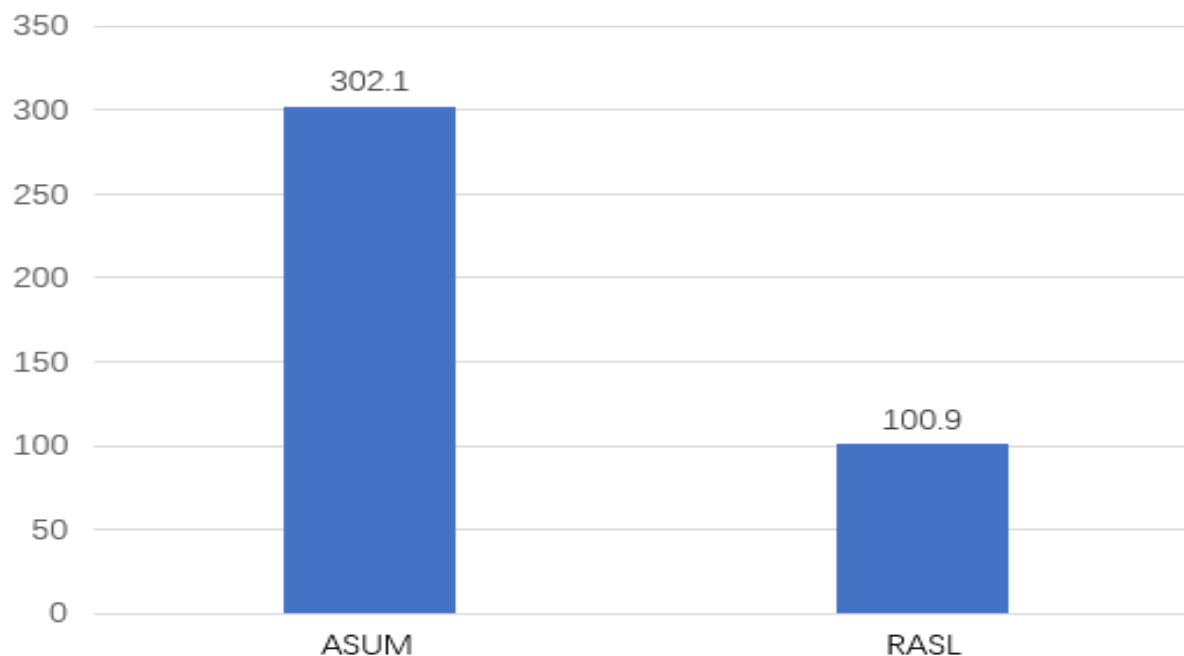
- 在这里是指：对于一个文档 d ，所训练出来的模型对于文档 d 属于哪个主题的确定程度
- 针对主题模型的优劣进行判断
- 困惑度越小，说明模型效果越好
- 困惑度的计算公式为：

$$\text{Perplexity}(D) = \exp \left\{ - \frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right\}$$

实验结果——RQ1



RQ1: ASUM方法和本文方法RASL的困惑度如何？



- 本文方法得到的困惑度小于ASUM方法所得到的困惑度
- 在困惑度方面：本文提出的方法优于ASUM方法



RQ1: ASUM方法和本文方法RASL的困惑度如何？

350

202.1

分析：

- 由于我们的方法在主题分析前加入了分类，使得问题的类型更加明确；
- 在同一类型中进行主题分析，一条评论属于某个主题的确定性就越大。

ASUM

RASL

- 本文方法得到的困惑度小于ASUM方法所得到的困惑度
- 在困惑度方面：本文提出的方法优于ASUM方法



问卷调查

问卷调查对象：6 位北京航空航天大学软件工程专业研究生
提供给受试者三份文件：

- 1) 原始的中差评集合
- 2) ASUM方法提供的分析结果
- 3) RASL方法提供的分析结果

实验验证——实验方法2(1)



问卷
问卷
提供

- 1 原始评论
- 2 原始评论
- 3 原始评论
- 4 原始评论
- 5 原始评论
- ...
- n 原始评论

Original review example

原始评论示例

主题1:
 主题词
 代表句
主题2:
 主题词
 代表句
 ...
主题m:
 主题词
 代表句

ASUM Method results example

ASUM方法结果示例

分类类型1

☆主题1-1:

 主题词

 代表句

主题1-2:

 主题词

 代表句

 ...

主题1-i:

 主题词

 代表句

 ...

分类类型16

☆主题16-1:

 主题词

 代表句

 ...

RASL Method results example

RASL方法结果示例



问卷调查

问卷调查研究问题：

RQ2: 和原始评论相比，ASUM方法和本文方法RASL是否包含完整的信息？

RQ3: 和原始评论相比，ASUM方法和本文方法RASL是否包含冗余的信息？

RQ4: ASUM方法和本文方法RASL的阅读理解性如何？

RQ5: ASUM方法和本文方法RASL的评论阅读时间如何？

实验验证——实验方法2(2)



Q1-1	和原始评论相比，方法 A 包含的信息是否完整？请打分(10 分制：1 分为缺失极多信息，10 分为没有缺失任何信息)
Q1-2	和原始评论相比，方法 B 包含的信息是否完整？请打分(10 分制：1 分为缺失极多信息，10 分为没有缺失任何信息)
Q2-1	和原始评论相比，方法 A 是否包含冗余的信息？请打分(10 分制：1 分为有许多冗余的信息，10 分为没有冗余的信息)
Q2-2	和原始评论相比，方法 B 是否包含冗余的信息？请打分(10 分制：1 分为有许多冗余的信息，10 分为没有冗余的信息)
Q3-1	方法 A 是否易于阅读理解？请打分(10 分制：1 分为难以阅读理解，10 分为非常易于阅读理解)
Q3-2	方法 B 是否易于阅读理解？请打分(10 分制：1 分为难以阅读理解，10 分为非常易于阅读理解)
Q4-1	请填写采用方法 A，阅读评论所花费的时间(分钟)
Q4-2	请填写采用方法 B，阅读评论所花费的时间(分钟)

实验结果——RQ2 (1)



RQ2: 和原始评论相比, ASUM方法和本文方法RASL是否包含完整的信息?

1分为缺失极多信息, 10分为没有缺失任何信息

受试者编号	360 云盘		今日头条	
	ASUM 方法	RASL 方法	ASUM 方法	RASL 方法
1	5	8	2	8
2	1	8	2	8
3	7	8	8	9
4	5	8	4	7
5	5	7	3	6
6	6	8	5	7

本文进行Mann-Whitney U检验检测这种差异的显著性

☆最终得到的结果显示, **RASL的完整性在0.05的显著性水平下明显优于ASUM**

实验结果——RQ2 (2)



完整性举例：

- (1) “**页面**中的**按钮**没反应。”
 - (2) “应该在**页面**中添加一个**按钮**。”
-

ASUM方法

主题词：
“按钮” “页面”
代表句：
页面中的按钮没反应。

RASL方法

问题类型1： 软件错误
主题词：“按钮” “页面”
代表句：
页面中的按钮没反应。
问题类型2： 请求增加功能
主题词：“按钮” “页面”
代表句：
应该在页面中添加一个按钮。

实验结果——RQ3



RQ3: 和原始评论相比, ASUM方法和本文方法RASL是否包含冗余的信息?

1分为有许多冗余的信息, 10分为没有冗余的信息

受试者编号	360 云盘		今日头条	
	ASUM 方法	RASL 方法	ASUM 方法	RASL 方法
1	4	9	4	7
2	2	8	1	9
3	9	9	9	9
4	5	9	4	8
5	3	5	4	5
6	7	6	5	6

本文进行Mann-Whitney U检验检测这种差异的显著性

☆最终得到的结果显示, 在0.05的显著性水平下, **RASL**包含的冗余信息明显少于**ASUM**

实验结果——RQ4 (1)



RQ4: ASUM方法和本文方法RASL的的阅读理解性如何?

1分为难以阅读理解, 10分为非常易于阅读理解

受试者编号	360 云盘		今日头条	
	ASUM 方法	RASL 方法	ASUM 方法	RASL 方法
1	6	9	5	9
2	3	8	2	9
3	3	8	2	8
4	6	9	6	9
5	6	8	5	7
6	7	8	6	7

本文进行Mann-Whitney U检验检测这种差异的显著性

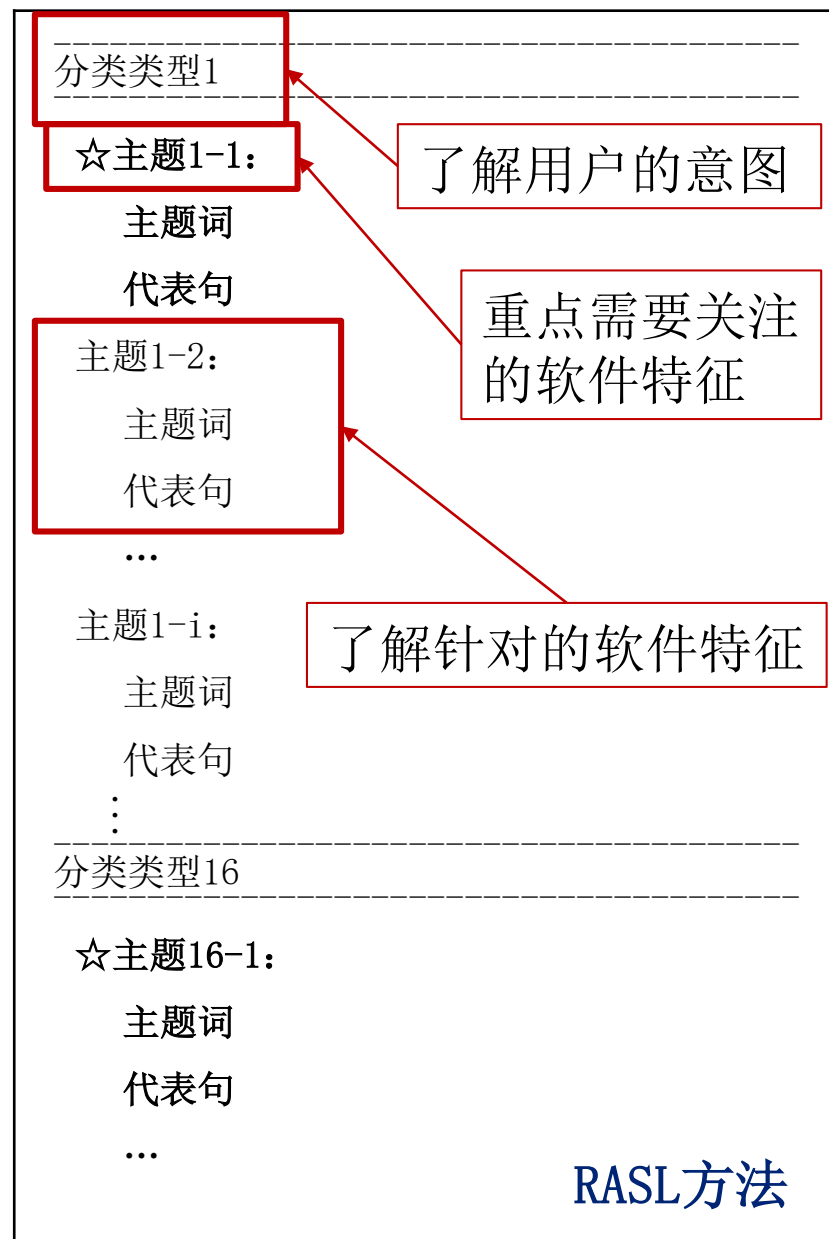
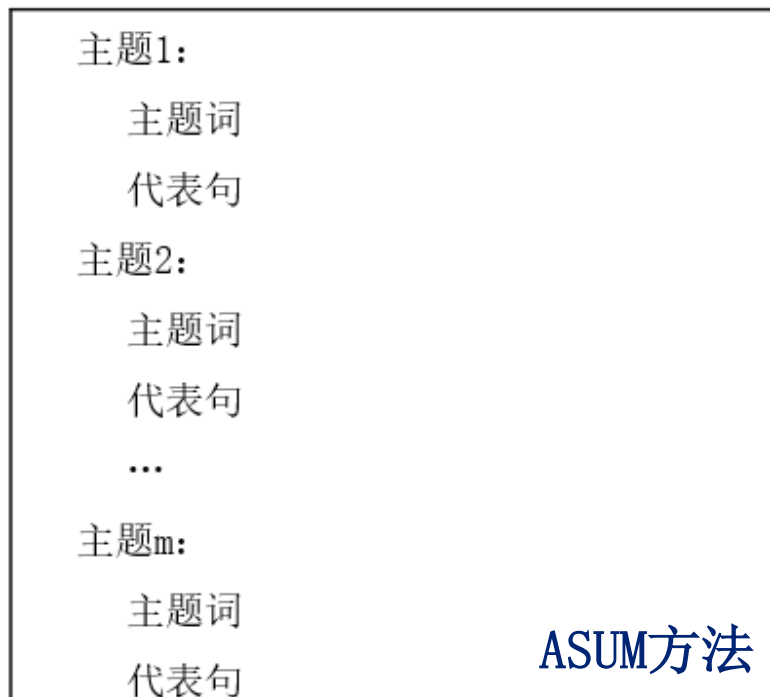
☆最终得到的结果显示, **RASL的可阅读性在0.05的显著性水平下明显优于ASUM**

实验结果——RQ4 (2)



阅读理解性分析:

我们的方法结构性更强



RQ5: ASUM方法和本文方法RASL的评论阅读时间如何?

受试者编号	360 云盘		今日头条	
	ASUM 方法	RASL 方法	ASUM 方法	RASL 方法
1	8	7	6	6
2	8	16	6	9
3	4	2	3.5	2
4	5	5	4	4
5	5	6	5	5
6	16	18	18	16

本文进行Mann-Whitney U检验检测这种差异的显著性

☆最终得到的结果显示，在0.05的显著性水平下，ASUM方法和RASL方法之间的时间差异不具有统计的显著性

目录

- 研究动机
- 研究方法
- 实验验证
- 论文总结

- 提出了一种基于支持向量机和主题模型的评论分析方法RASL

该方法通过收集用户使用软件后产生的反馈信息，挖掘其中的各类需求，使开发者能快速、直观地理解用户反馈

- 设计实验对比ASUM方法对RASL方法进行评估

- 1) RASL方法困惑度明显减少

- 2) 问卷结果证明，和ASUM相比，发现RASL方法可理解性、完整性更佳，包含的冗余信息也更少



谢谢！

报告人：
陈琪(北京航空航天大学)