



(12) 发明专利

(10) 授权公告号 CN 107229636 B

(45) 授权公告日 2021.08.13

(21) 申请号 201610173402.0

(22) 申请日 2016.03.24

(65) 同一申请的已公布的文献号  
申请公布号 CN 107229636 A

(43) 申请公布日 2017.10.03

(73) 专利权人 腾讯科技(深圳)有限公司  
地址 518000 广东省深圳市福田区振兴路  
赛格科技园2栋东403室

(72) 发明人 孙义康 万翼龙 项丹 周大龙  
余洋洋

(74) 专利代理机构 深圳市深佳知识产权代理事  
务所(普通合伙) 44285  
代理人 王仲凯

(51) Int. Cl.  
G06F 16/35 (2019.01)  
G06F 16/36 (2019.01)  
G06F 16/951 (2019.01)  
G06F 16/953 (2019.01)  
G06F 16/9532 (2019.01)

(56) 对比文件

CN 105005589 A, 2015.10.28

CN 104978328 A, 2015.10.14

CN 103729459 A, 2014.04.16

CN 104123302 A, 2014.10.29

TW 201346813 A, 2013.11.16

US 2004054627 A1, 2004.03.18

US 2003055689 A1, 2003.03.20

魏慧玲. 文本情感分析在产品评论中的应用研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2014, (第6期), 第1138-1208页.

Xiao Sun等.Chinese microblog sentiment classification based on convolution neural network with content extension method.《2015 International Conference on Affective Computing and Intelligent Interaction (ACII)》.2015, 第408-414页.

审查员 彭傲雪

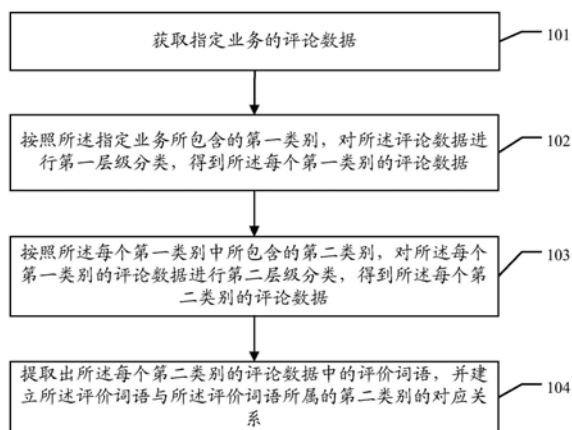
权利要求书2页 说明书9页 附图9页

(54) 发明名称

一种词语分类的方法及装置

(57) 摘要

本发明公开了一种词语分类的方法,包括:获取指定业务的评论数据;按照所述指定业务所包含的第一类别,对所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;按照所述每个第一类别中所包含的第二类别,对所述每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;提取出所述每个第二类别的评论数据中的评价词语,并建立所述评价词语与所述评价词语所属的第二类别的对应关系。本发明实施例提供的词语分类的方法,可以针对指定业务进行词语分类,从而可以按照每种业务统计出与该业务优劣相关的评价词语,提高了监测的有效性。



1. 一种词语分类的方法,其特征在于,包括:

获取指定业务的评论数据;

按照所述指定业务所包含的第一类别,对所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;

按照所述每个第一类别中所包含的第二类别,对所述每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;

提取出所述每个第二类别的评论数据中的评价词语,并建立所述评价词语与所属的第二类别的对应关系,所述第一类别与所述第二类别均为所述指定业务的类别,所述评价词语包括正面评价词语和负面评价词语;

获取所述指定业务更新的评论数据;

根据所述更新的评论数据,确定所述第二类别中更新的评价词语;

当所述第二类别中已包含所述更新的评价词语时,则在与所述更新的评价词语相同的评价词语处做数量累加;

当所述第二类别中未包含所述更新的评价词语时,建立所述更新的评价词语与所述第二类别的对应关系;

统计出与所述指定业务类别优劣势相关的所述评价词语;

当所述指定业务为发布会时,所述发布会所包含的第一类别包括内容、人物、服务、宣传和整体,其中,所述内容包含的第二类别包括主题、产品、设施、音乐、场地、环境和时间中的至少一项,所述人物包含的第二类别包括嘉宾、团队、主持人、表演人员和观众中的至少一项,所述服务包含的第二类别包括售票、邀请函和现场服务中的至少一项,所述宣传包含的第二类别为媒体阵容和宣传素材中的至少一项,所述整体包含的第二类别也为整体。

2. 根据权利要求1所述的方法,其特征在于,所述提取出所述每个第二类别的评论数据中的评价词语之后,所述方法还包括:

确定所述评价词语的属性;

所述建立所述评价词语与所属的第二类别的对应关系,包括:

按照所述评价词语的属性,建立所述评价词语与所属的第二类别的对应关系。

3. 根据权利要求1或2所述的方法,其特征在于,所述获取指定业务的评论数据之前,所述方法还包括:

根据接收到的评论数据中的关键词,确定接收到的所述评论数据所属的业务。

4. 一种词语分类的装置,其特征在于,包括:

获取单元,用于获取指定业务的评论数据;

第一分类单元,用于按照所述指定业务所包含的第一类别,对所述获取单元获取的所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;

第二分类单元,用于按照所述每个第一类别中所包含的第二类别,对所述第一分类单元分类得到后的每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;

提取单元,用于提取出所述第二分类单元分类后的所述每个第二类别的评论数据中的评价词语;

关系建立单元,用于建立所述提取单元提取的评价词语与所述评价词语所属的第二类别的对应关系,所述第一类别与所述第二类别均为所述指定业务的类别,所述评价词语包括正面评价词语和负面评价词语;

所述装置用于获取所述指定业务更新的评论数据;根据所述更新的评论数据,确定所述第二类别中更新的评价词语;当所述第二类别中已包含所述更新的评价词语时,则在与所述更新的评价词语相同的评价词语处做数量累加;当所述第二类别中未包含所述更新的评价词语时,建立所述更新的评价词语与所述第二类别的对应关系;

所述装置用于统计出与所述指定业务类别优劣势相关的所述评价词语;

所述装置用于当所述指定业务为发布会时,所述发布会所包含的第一类别包括内容、人物、服务、宣传和整体,其中,所述内容包含的第二类别包括主题、产品、设施、音乐、场地、环境和时间中的至少一项,所述人物包含的第二类别包括嘉宾、团队、主持人、表演人员和观众中的至少一项,所述服务包含的第二类别包括售票、邀请函和现场服务中的至少一项,所述宣传包含的第二类别为媒体阵容和宣传素材中的至少一项,所述整体包含的第二类别也为整体。

5. 根据权利要求4所述的装置,其特征在于,所述装置还包括第一确定单元,

所述第一确定单元,用于确定所述提取单元提取的所述评价词语的属性;

所述关系建立单元,用于按照所述第一确定单元确定的所述评价词语的属性,建立所述评价词语与所述评价词语所属的第二类别的对应关系。

6. 根据权利要求4或5所述的装置,其特征在于,所述装置还包括第三确定单元,

所述第三确定单元,用于在所述获取单元获取指定业务的评论数据之前,根据接收到的评论数据中的关键词,确定接收到的所述评论数据所属的业务。

7. 一种词语分类的装置,其特征在于,包括:存储器与处理器;

所述存储器,用于存储操作指令;

所述处理器,用于调用所述操作指令,以执行如权利要求1~3任一项所述的词语分类的方法的步骤。

8. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有程序,所述程序用于实现如权利要求1~3任一项所述的词语分类的方法的步骤。

## 一种词语分类的方法及装置

### 技术领域

[0001] 本发明涉及数据处理技术领域,具体涉及一种词语分类的方法及装置。

### 背景技术

[0002] 无论对于一个真实产品还是一个虚拟产品,口碑对该产品的影响都越来越重要,而口碑的好坏来源于用户的评价,若能及时掌握用户的评价,则可以对该产品做出针对性的修改,以便更好的完善该产品。

[0003] 现有技术中的舆情监测系统,是指通过相关的专业舆情软件从互联网上繁杂的信息中将关键信息抓取出来,然后添加到舆情监测系统的公共词库中。这样,可以从公共词库中了解公众的对舆论事件的反应。

[0004] 现有技术中的舆情监测系统应用到互联网上,可以抽取出互联网上的用户评价的一些关键词,但互联网上用户的评价所涉及到的业务或产品种类繁多,业务运营商无法根据公共词库中的关键词对每种业务或产品的优劣做出正确的判断。

### 发明内容

[0005] 本发明实施例提供一种词语分类的方法,可以按照每种业务统计出与该业务优劣相关的评价词语,从而提高监测的有效性。本发明实施例还提供了相应的装置。

[0006] 本发明第一方面提供一种词语分类的方法,包括:

[0007] 获取指定业务的评论数据;

[0008] 按照所述指定业务所包含的第一类别,对所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;

[0009] 按照所述每个第一类别中所包含的第二类别,对所述每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;

[0010] 提取出所述每个第二类别的评论数据中的评价词语,并建立所述评价词语与所述评价词语所属的第二类别的对应关系。

[0011] 本发明第二方面提供一种词语分类的装置,包括:

[0012] 获取单元,用于获取指定业务的评论数据;

[0013] 第一分类单元,用于按照所述指定业务所包含的第一类别,对所述获取单元获取的所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;

[0014] 第二分类单元,用于按照所述每个第一类别中所包含的第二类别,对所述第一分类单元分类得到后的每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;

[0015] 提取单元,用于提取出所述第二分类单元分类后的所述每个第二类别的评论数据中的评价词语;

[0016] 关系建立单元,用于建立所述提取单元提取的评价词语与所述评价词语所属的第二类别的对应关系。

[0017] 与现有技术中无法根据舆情监测系统的公共词库中的关键词对每种业务或产品的优劣做出正确的判断相比,本发明实施例提供的词语分类的方法,可以针对指定业务进行词语分类,从而可以按照每种业务统计出与该业务优劣相关的评价词语,提高了监测的有效性。

### 附图说明

[0018] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0019] 图1是本发明实施例中舆情监测系统的一实施例示意图;

[0020] 图2是本发明实施例中词语分类的方法的一实施例示意图;

[0021] 图3是本发明实施例中发布会场景的分类架构示意图;

[0022] 图4是本发明实施例中影片场景的分类架构示意图;

[0023] 图5是本发明实施例中漫画场景的分类架构示意图;

[0024] 图6是本发明实施例中游戏场景的分类架构示意图;

[0025] 图7是本发明实施例中文学场景的分类架构示意图;

[0026] 图8是本发明实施例中词语分类的装置的一实施例示意图;

[0027] 图9是本发明实施例中词语分类的装置的另一实施例示意图;

[0028] 图10是本发明实施例中词语分类的装置的另一实施例示意图;

[0029] 图11是本发明实施例中词语分类的装置的另一实施例示意图;

[0030] 图12是本发明实施例中词语分类的装置的另一实施例示意图。

### 具体实施方式

[0031] 本发明实施例提供一种词语分类的方法,可以按照每种业务统计出与该业务优劣相关的评价词语,从而提高监测的有效性。本发明实施例还提供了相应的装置。以下分别进行详细说明。

[0032] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0033] 本发明实施例所提供的词语分类的方法,应用于舆情监测系统,可以提供针对泛娱乐业务事件、活动和产品等进行正负面舆情分析的评价体系词库。也就是说,可以针对一款游戏、一部动漫、一部文学作品、一部电影和一场发布会等提供专门的评价词库。需要说明的是,评价词库不限于上述几个领域,未来针对有大影响力的活动事件、传播素材等涉及传播的内容,都能够进行监测,评价词库所涉及领域会根据需求动态变化。这样,内容提供商就可以结合评价词库进行针对性的改进,而且,本申请中的词库包括正面评价词语和负面评价词语,有利于内容提供商及时掌握所提供作品的优势和劣势,进而制作出更好的内容。而且,本申请的评价词语结合到具体产品、作品或者业务,这样也可以提供机器学习的

效率。

[0034] 舆情监测系统通常包括如图1所示的用户平台、多个分类节点和存储阵列,用户可以通过该用户平台进行业务评价,在用户平台上输入针对某项业务的评论数据。用户平台上接收到的评论数据,可以分配给分类节点进行分类,分类节点分类的依据可以是按照业务进行分类,确定评论数据所属的业务后,将该评论数据输入到存储阵列中存储,存储设备中可以划分成多个存储区域,例如:可以划分成电影存储区域、动漫存储区域、文学存储区域和发布会存储区域等,其中每个存储区域还可以再细分成多个小的存储区域,小的存储区域用于存储具体业务的评论数据。

[0035] 当需要获知用户对指定业务的相关评价时,词语分类的装置可以从存储设备中存储该指定业务的存储区域获取该指定业务的评论数据,再进行分类,提取出评价词语。

[0036] 执行本发明实施例中词语分类的装置可以是独立的用于词语分类的服务器,也可以是包含多个分类节点的服务集群。

[0037] 图2为本发明实施例中词语分类的方法的一实施例示意图。

[0038] 如图2所述,本发明实施例中词语分类的方法的一实施例包括:

[0039] 101、获取指定业务的评论数据。

[0040] 指定业务可以是内容开发商所指定的业务,例如:某部电影、某部文学作品和某本漫画等。

[0041] 102、按照所述指定业务所包含的第一类别,对所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据。

[0042] 指定业务可以包括多个第一类别,例如:电影可以包括电影特性、人物、电影内在和整体等第一类别。

[0043] 可以把针对该电影的评论数据按照电影特性、人物、电影内在和整体等几个第一类别进行分类。

[0044] 103、按照所述每个第一类别中所包含的第二类别,对所述每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据。

[0045] 第一类别中还可以包括多个第二类别,例如:电影特性中可以包括台词对白、情节/剧情、悬念、音响效果、色彩、摄影/画面、音乐、细节、特技/特效,以及场景等多个第二类别;

[0046] 将电影特性的评论数据按照台词对白、情节/剧情、悬念、音响效果、色彩、摄影/画面、音乐、细节、特技/特效,以及场景进行分类,就可以得到每个第二类别的评论数据。

[0047] 当然,第二类别下还可以再划分第三类别,第三类别下也可以再划分第四类别,无限划分是不能穷举的,因此,本申请中只是以划分到第二类别为例进行说明。

[0048] 104、提取出所述每个第二类别的评论数据中的评价词语,并建立所述评价词语与所述评价词语所属的第二类别的对应关系。

[0049] 从评论数据中提取出评价词语可以采用关键词的提取方法,例如:第二类别剧情下的评论数据为:《地球上的星星》主题太棒了,题材好新颖,从中可以提取出关键词“新颖”,然后,建立剧情和新颖的对应关系。

[0050] 本发明实施例中,获取指定业务的评论数据;按照所述指定业务所包含的第一类别,对所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;按照所述每个

第一类别中所包含的第二类别,对所述每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;提取出所述每个第二类别的评论数据中的评价词语,并建立所述评价词语与所属的第二类别的对应关系。与现有技术中无法根据舆情监测系统的公共词库中的关键词对每种业务或产品的优劣做出正确的判断相比,本发明实施例提供的词语分类的方法,可以针对指定业务进行词语分类,从而可以按照每种业务统计出与该业务优劣相关的评价词语,提高了监测的有效性。

[0051] 可选地,所述提取出所述每个第二类别的评论数据中的评价词语之后,所述方法还可以包括:

[0052] 确定所述评价词语的属性;

[0053] 所述建立所述评价词语与所属的第二类别的对应关系,包括:

[0054] 按照所述评价词语的属性,建立所述评价词语与所属的第二类别的对应关系。

[0055] 评价词语的属性可分正面和负面,正面评价词语有利于估计内容提供商继续延续好的方面,负面评价词语有利于估计内容提供商及时修改内容,从而可以提供更好的业务给用户。

[0056] 本发明实施例中,针对发布会、影片、动漫、游戏和文学作品都有进行第一类别和第二类别的关系构建,有的第二类别下还构建了第三类别,下面分别结合图3至图7说明本发明实施例中针对不同的业务进行词语分类的过程。

[0057] 图3为本发明实施例中发布会场景的分类架构示意图。

[0058] 如图3所示,发布会业务的第一类别可以包括:内容、人物、服务、宣传和整体,其中,内容下的第二类别可以包括:主题、产品、设施、音乐、场地、环境和时间。人物下的第二类别可以包括嘉宾、团队、主持人、表演人员和观众。服务下的第二类别可以包括售票、邀请函和现场服务。宣传下的第二类别可以包括媒体阵容和宣传素材。

[0059] 对发布会业务的评论数据可以包括正面评论数据和反面评论数据,对评论数据可以分别从正面和反面进行分类,正面的评论例如可以参阅表1进行理解,负面的评论例如可以参阅表2进行理解。

[0060] 表1:发布会业务的正面评价词语与类别的对应关系

| 第一类别 | 第二类别 | 评价词语 |
|------|------|------|
| 内容   | 音乐   | 应景   |
| 内容   | 场地   | 气派   |
| 人物   | 主持人  | 独特   |
| 人物   | 观众   | 守秩序  |
| 服务   | 邀请函  | 精美   |
| 宣传   | 媒体素材 | 新鲜   |
| 整体   | 整体   | 华丽   |

[0062] 表2:发布会业务的负面评价词语与类别的对应关系

|        | 第一类别 | 第二类别 | 评价词语 |
|--------|------|------|------|
|        | 内容   | 环境   | 拥挤   |
|        | 内容   | 场地   | 偏远   |
| [0063] | 人物   | 嘉宾   | 二流   |
|        | 人物   | 团队   | 经验不足 |
|        | 服务   | 售票   | 贵    |
|        | 宣传   | 媒体阵容 | 凑数   |

|        |    |    |    |
|--------|----|----|----|
| [0064] | 整体 | 整体 | 无趣 |
|--------|----|----|----|

[0065] 表1和表2只是发布会业务的正负面评价词语的几个举例,实际上,舆情监测系统中会收录很多个评价词语与第二类别的对应关系。这样,发布会的举办方就可以通过这些评价了解自己的优点和不足,优点可以在下次发布会中继续延续,不足可以在下次发布会中尽量弥补。

[0066] 图4是本发明实施例中影片场景的分类架构示意图。

[0067] 如图4所示,影片业务的第一类别可以包括:电影特性、人物、电影内在和整体。电影特性中的第二类别可以包括台词对白、剧情、悬念、音响效果、摄影、色彩、音乐、细节、特技和场景等。人物中的第二类别可以包括导演和演员。电影内在中的第二类别可以包括价值观、情怀、故事和风格。整体就是从整体上去评判,可以认为整体的第二类别也为整体。

[0068] 对影片业务的评论数据可以包括正面评论数据和反面评论数据,对评论数据可以分别从正面和反面进行分类,正面的评论例如可以参阅表3进行理解,负面的评论例如可以参阅表4进行理解。

[0069] 表3:影片业务的正面评价词语与类别的对应关系

|        | 第一类别 | 第二类别 | 评价词语 |
|--------|------|------|------|
| [0070] | 电影特性 | 剧情   | 温暖   |
|        | 电影特性 | 特效   | 冲击力  |
|        | 人物   | 演员   | 入戏   |
|        | 人物   | 导演   | 导的好  |
|        | 电影内在 | 故事   | 生动   |
|        | 电影内在 | 价值观  | 正能量  |
|        | 整体   | 整体   | 经典   |

[0071] 表4:影片业务的负面评价词语与类别的对应关系

|        | 第一类别 | 第二类别 | 评价词语 |
|--------|------|------|------|
| [0072] | 电影特性 | 剧情   | 温暖   |
|        | 电影特性 | 特效   | 特效太多 |



|        |      |    |      |
|--------|------|----|------|
|        | 人物   | 演员 | 丑    |
|        | 人物   | 导演 | 选景不好 |
| [0073] | 电影内在 | 情怀 | 不动情  |
|        | 电影内在 | 风格 | 俗气   |
|        | 整体   | 整体 | 平庸   |

[0074] 表3和表4只是影片业务的正负面评价词语的几个举例,实际上,舆情监测系统中会收录很多个评价词语与第二类别的对应关系。这样,影片的制作方就可以通过这些评价了解自己的优点和不足,优点可以在下次发布会中继续延续,不足可以在下次发布会中尽量弥补。

[0075] 动漫、游戏和文学作品的场景的分类架构示意图可分别参阅图5至图7进行理解。

[0076] 关于动漫、游戏和文学作品的第一类别和第二类别可以参阅图5至图7并结合上述图3的发布会业务和图4的影片业务进行理解,本申请中不再一一列举。

[0077] 本发明实施例中,所述建立所述评价词语与所述评价词语所属的第二类别的对应关系之后,所述方法还可以包括:

[0078] 获取所述指定业务更新的评论数据;

[0079] 根据所述更新的评论数据,确定所述第二类别中更新的评价词语;

[0080] 根据所述更新的评价词语,更新所述对应关系。

[0081] 内容提供商在获知一些评论词语后可能会尽量完善指定业务,例如:可以给游戏打补丁,改善游戏中的缺点,这样,可能会有一些新的评论数据,针对新的评论数据,可以再提出去更新的评价词语,更新第二类别与评价词语的对应关系。从而使内容提供上可以实时了解到新的评价词语,进而做出更好的改善。

[0082] 可选地,所述根据所述更新的评价词语,更新所述对应关系,可以包括:

[0083] 当所述第二类别中已包含所述更新的评价词语时,则在与所述更新的评价词语相同的评价词语处做数量累加;

[0084] 当所述第二类别中未包含所述更新的评价词语时,建立所述更新的评价词语与所述第二类别的对应关系。

[0085] 本发明实施例中,当影片的评论语中已经包括剧情--温暖的评价词语时,可以累加剧情--温暖的次数,而不需要重复列出剧情--温暖的对应关系。当影片的评论词语中未包括剧情--一环相扣的评价词语时,则可以增加剧情--一环相扣的对应关系。

[0086] 可选地,所述获取指定业务的评论数据之前,所述方法还可以包括:

[0087] 根据接收到的评论数据中的关键词,确定接收到的所述评论数据所属的业务。

[0088] 本发明实施例中,确定接收到的所述评论数据所属的业务可以参阅图1部分的描述进行理解,本处不做过多赘述。

[0089] 参阅图8,本发明实施例提供的词语分类的装置20的一实施例包括:

[0090] 获取单元201,用于获取指定业务的评论数据;

[0091] 第一分类单元202,用于按照所述指定业务所包含的第一类别,对所述获取单元201获取的所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;

[0092] 第二分类单元203,用于按照所述每个第一类别中所包含的第二类别,对所述第一分类单元202分类得到后的每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;

[0093] 提取单元204,用于提取出所述第二分类单元203分类后的所述每个第二类别的评论数据中的评价词语;

[0094] 关系建立单元205,用于建立所述提取单元204提取的评价词语与所述评价词语所属的第二类别的对应关系。

[0095] 本发明实施例中,获取单元201获取指定业务的评论数据;第一分类单元202按照所述指定业务所包含的第一类别,对所述获取单元201获取的所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;第二分类单元203按照所述每个第一类别中所包含的第二类别,对所述第一分类单元202分类得到后的每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;提取单元204提取出所述第二分类单元203分类后的所述每个第二类别的评论数据中的评价词语;关系建立单元205建立所述提取单元204提取的评价词语与所述评价词语所属的第二类别的对应关系。与现有技术中无法根据舆情监测系统的公共词库中的关键词对每种业务或产品的优劣做出正确的判断相比,本发明实施例提供的词语分类的装置,可以针对指定业务进行词语分类,从而可以按照每种业务统计出与该业务优劣相关的评价词语,提高了监测的有效性。

[0096] 可选地,在上述图8对应的实施例的基础上,参阅图9,本发明实施例提供的词语分类的装置的另一个可选实施例中,所述装置20还包括第一确定单元206,

[0097] 所述第一确定单元206,用于确定所述提取单元204提取的所述评价词语的属性;

[0098] 所述关系建立单元205,用于按照所述第一确定单元206确定的所述评价词语的属性,建立所述评价词语与所述评价词语所属的第二类别的对应关系。

[0099] 可选地,在上述图8对应的实施例的基础上,参阅图10,本发明实施例提供的词语分类的装置20的另一可选实施例中,所述装置还包括第二确定单元207和更新单元208,

[0100] 所述获取单元201,还用于获取所述指定业务更新的评论数据;

[0101] 所述第二确定单元207,用于根据所述获取单元201获取的所述更新的评论数据,确定所述第二类别中更新的评价词语;

[0102] 所述更新单元208,用于根据所述第二确定单元207确定的所述更新的评价词语,更新所述对应关系。

[0103] 可选地,在上述图10对应的实施例的基础上,本发明实施例提供的词语分类的装置20的另一可选实施例中,

[0104] 所述更新单元208用于:

[0105] 当所述第二类别中已包含所述更新的评价词语时,则在与所述更新的评价词语相同的评价词语处做数量累加;

[0106] 当所述第二类别中未包含所述更新的评价词语时,建立所述更新的评价词语与所述第二类别的对应关系。

[0107] 可选地,在上述图8对应的实施例的基础上,参阅图11,本发明实施例提供的词语分类的装置20的另一可选实施例中,所述装置还包括第三确定单元209,

[0108] 所述第三确定单元209,用于在所述获取单元201获取指定业务的评论数据之前,

根据接收到的评论数据中的关键词,确定接收到的所述评论数据所属的业务。

[0109] 图8至图11所描述的词语分类的装置可以参阅图1至图7部分的描述进行理解,本处不再重复赘述。

[0110] 图12是本发明实施例提供的词语分类的装置20的结构示意图。所述词语分类的装置20应用于舆情监测系统,所述词语分类的装置20包括处理器210、存储器250和收发器230,存储器250可以包括只读存储器和随机存取存储器,并向处理器210提供操作指令和数据。存储器250的一部分还可以包括非易失性随机存取存储器(NVRAM)。

[0111] 在一些实施方式中,存储器250存储了如下的元素,可执行模块或者数据结构,或者他们的子集,或者他们的扩展集:

[0112] 在本发明实施例中,通过调用存储器250存储的操作指令(该操作指令可存储在操作系统中),

[0113] 获取指定业务的评论数据;

[0114] 按照所述指定业务所包含的第一类别,对所述评论数据进行第一层级分类,得到所述每个第一类别的评论数据;

[0115] 按照所述每个第一类别中所包含的第二类别,对所述每个第一类别的评论数据进行第二层级分类,得到所述每个第二类别的评论数据;

[0116] 提取出所述每个第二类别的评论数据中的评价词语,并建立所述评价词语与所述评价词语所属的第二类别的对应关系。

[0117] 与现有技术中无法根据舆情监测系统的公共词库中的关键词对每种业务或产品的优劣做出正确的判断相比,本发明实施例提供的词语分类的装置,可以针对指定业务进行词语分类,从而可以按照每种业务统计出与该业务优劣相关的评价词语,提高了监测的有效性。

[0118] 处理器210控制词语分类的装置20的操作,处理器210还可以称为CPU(Central Processing Unit,中央处理单元)。存储器250可以包括只读存储器和随机存取存储器,并向处理器210提供指令和数据。存储器250的一部分还可以包括非易失性随机存取存储器(NVRAM)。具体的应用中词语分类的装置20的各个组件通过总线系统220耦合在一起,其中总线系统220除包括数据总线之外,还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见,在图中将各种总线都标为总线系统220。

[0119] 上述本发明实施例揭示的方法可以应用于处理器210中,或者由处理器210实现。处理器210可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器210中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器210可以是通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现成可编程门阵列(FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器250,处理器210读取存储器250中的信息,结合其硬件完成上述方法的步骤。

[0120] 可选地,处理器210还用于确定所述评价词语的属性;按照所述评价词语的属性,建立所述评价词语与所述评价词语所属的第二类别的对应关系。

[0121] 可选地,处理器210还用于获取所述指定业务更新的评论数据;根据所述更新的评论数据,确定所述第二类别中更新的评价词语;根据所述更新的评价词语,更新所述对应关系。

[0122] 可选地,处理器210用于当所述第二类别中已包含所述更新的评价词语时,则在与所述更新的评价词语相同的评价词语处做数量累加;当所述第二类别中未包含所述更新的评价词语时,建立所述更新的评价词语与所述第二类别的对应关系。

[0123] 可选地,处理器210还用于根据接收到的评论数据中的关键词,确定接收到的所述评论数据所属的业务。

[0124] 图12所提供的词语分类的装置,可以参阅图1至图11部分的描述进行理解,本处不做过多赘述。

[0125] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成,该程序可以存储于一计算机可读存储介质中,存储介质可以包括:ROM、RAM、磁盘或光盘等。

[0126] 以上对本发明实施例所提供的词语分类的方法以及装置进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

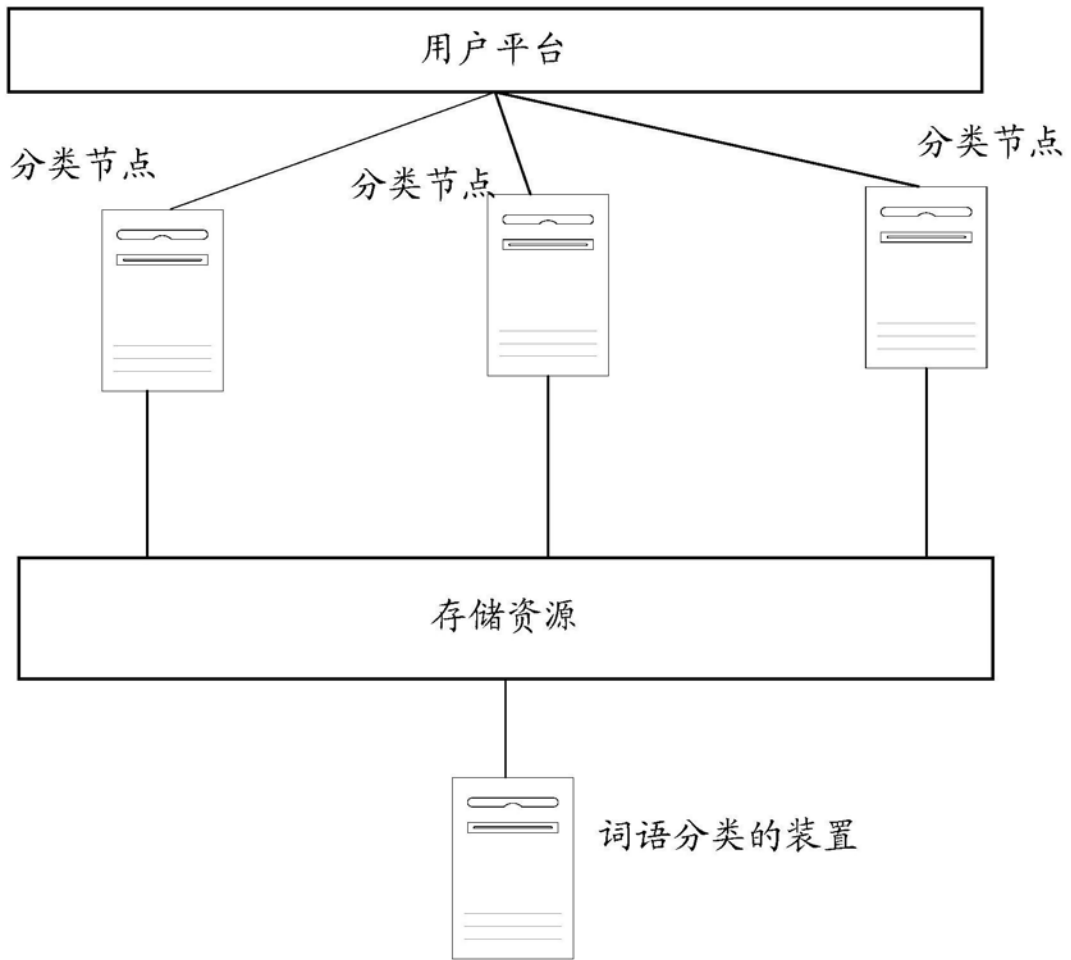


图1

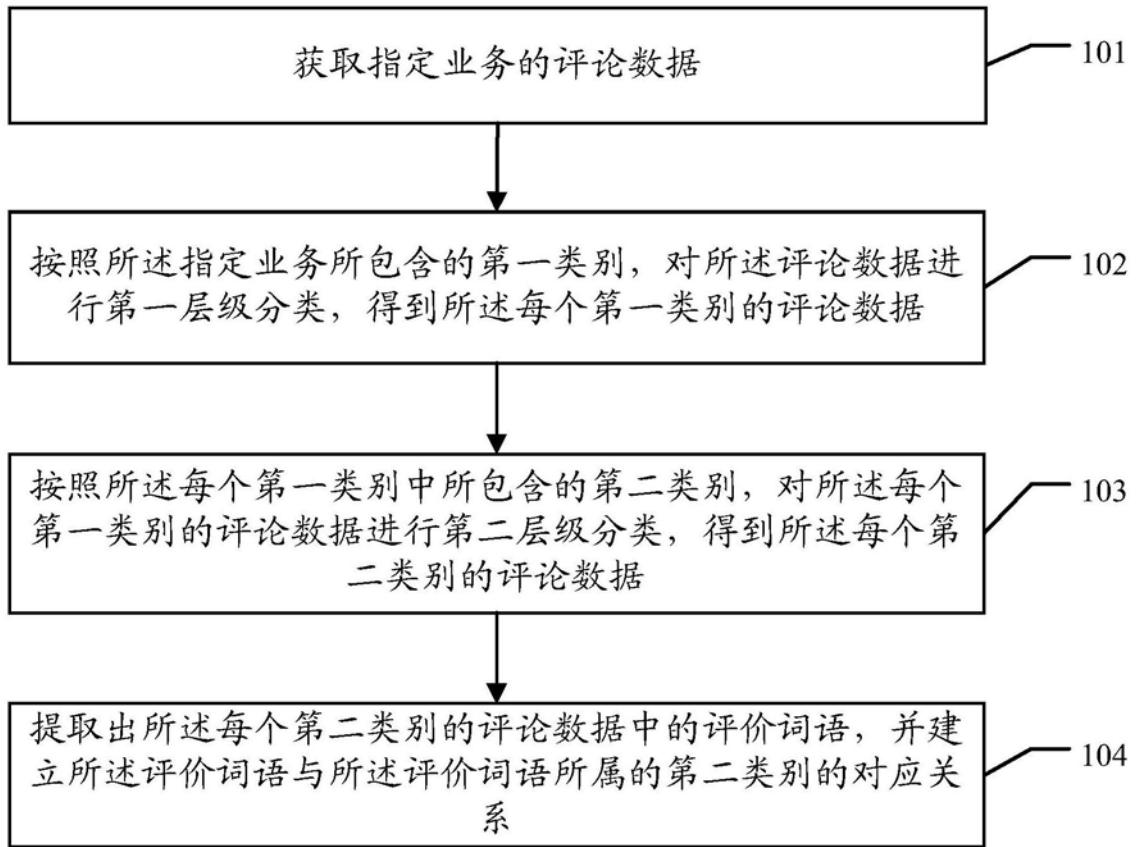


图2

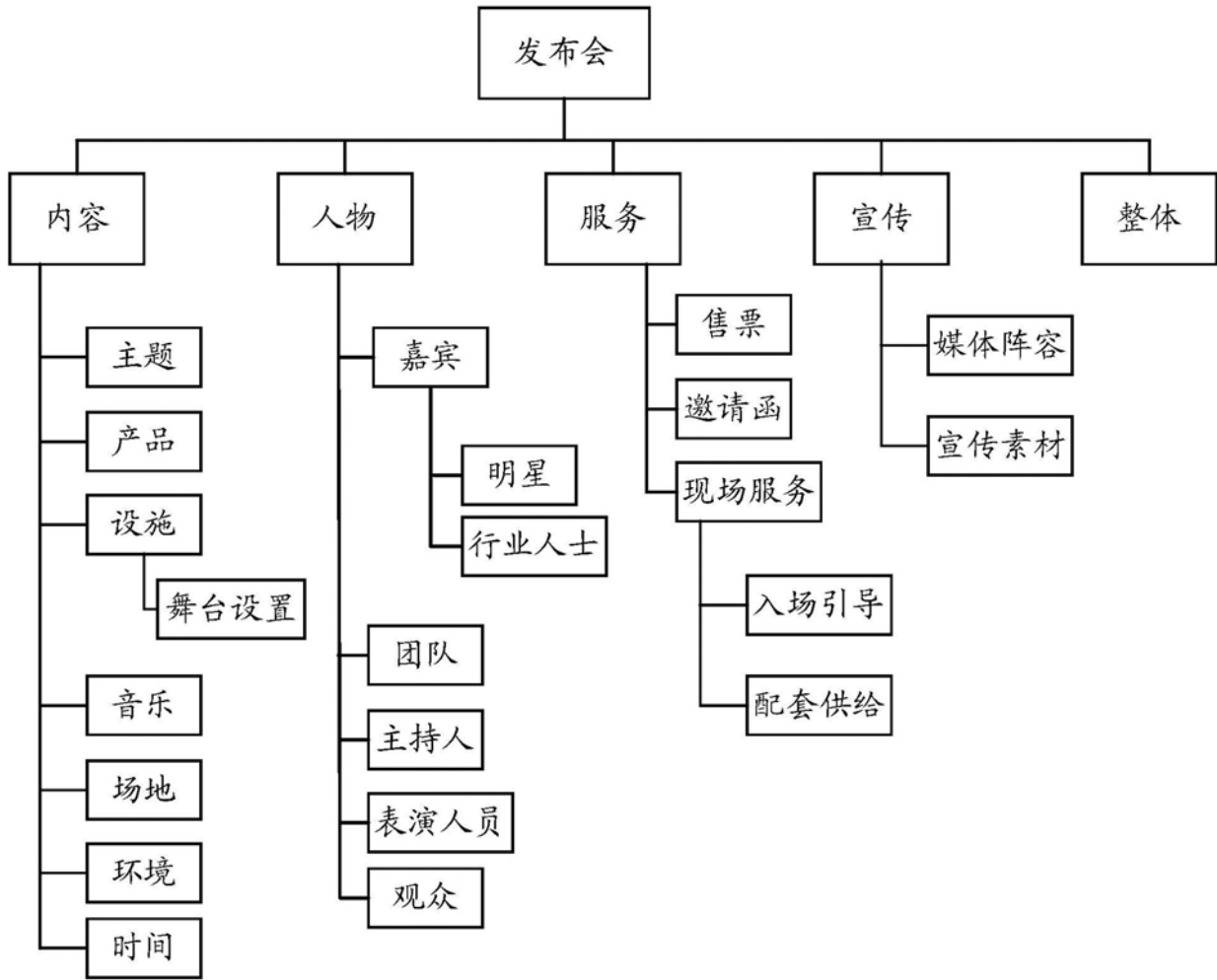


图3

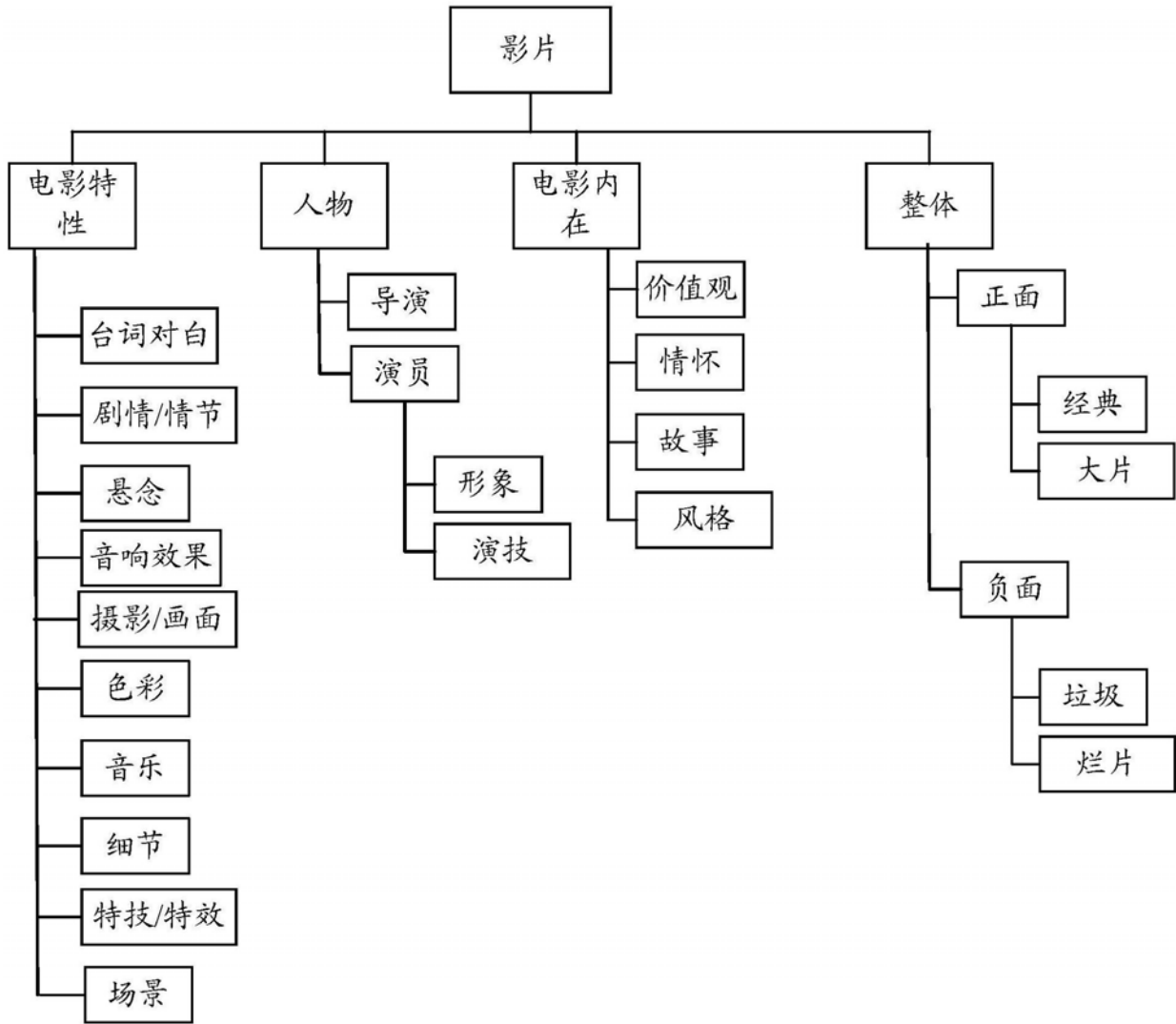


图4



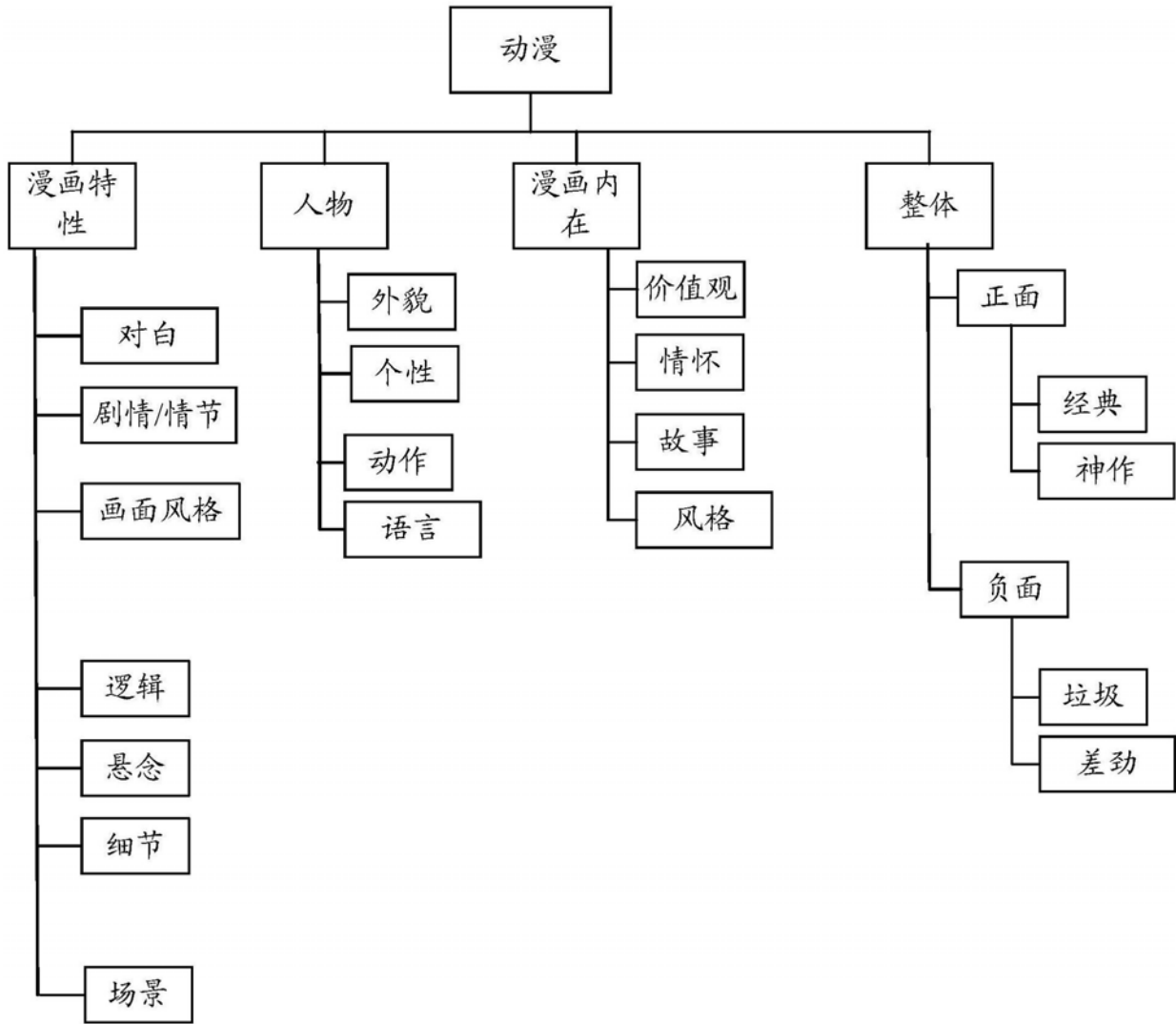


图5

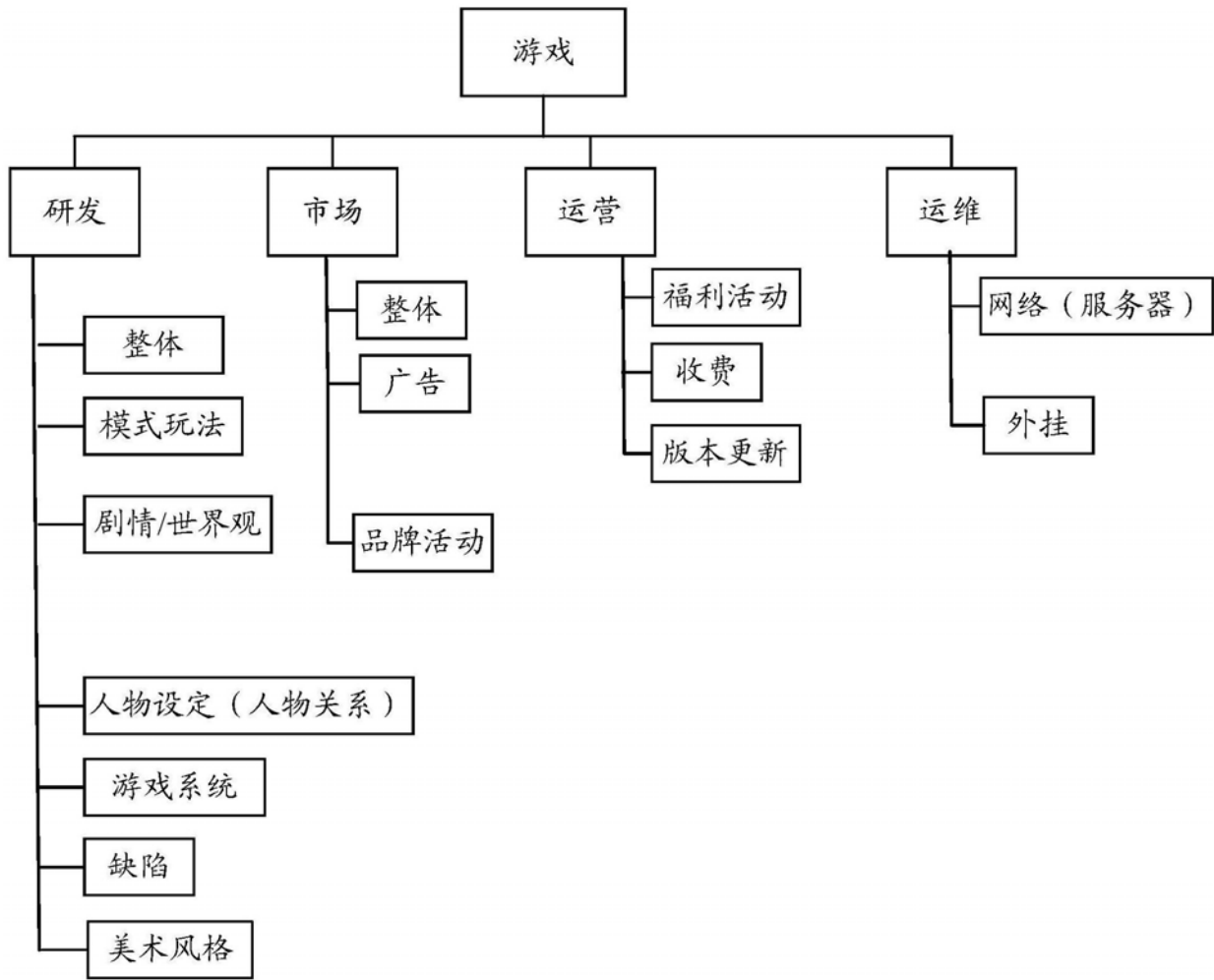


图6

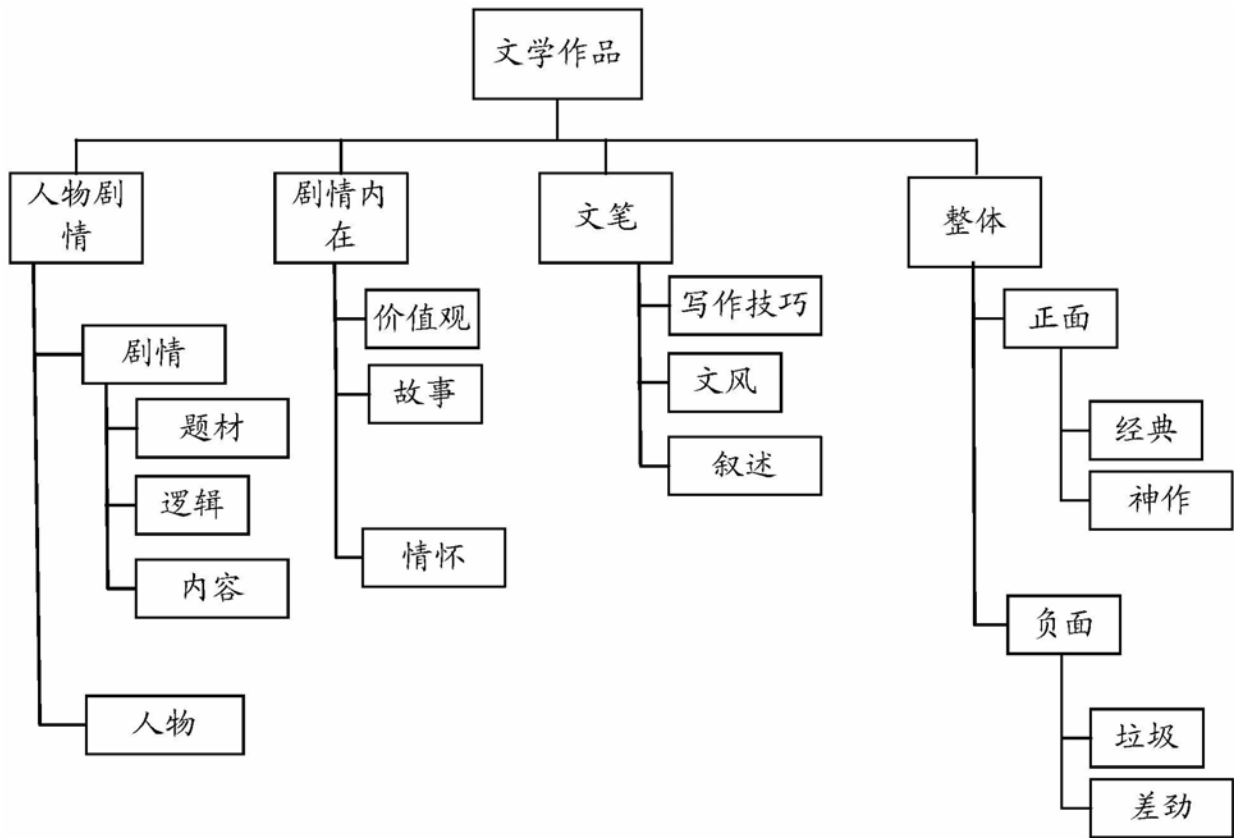


图7

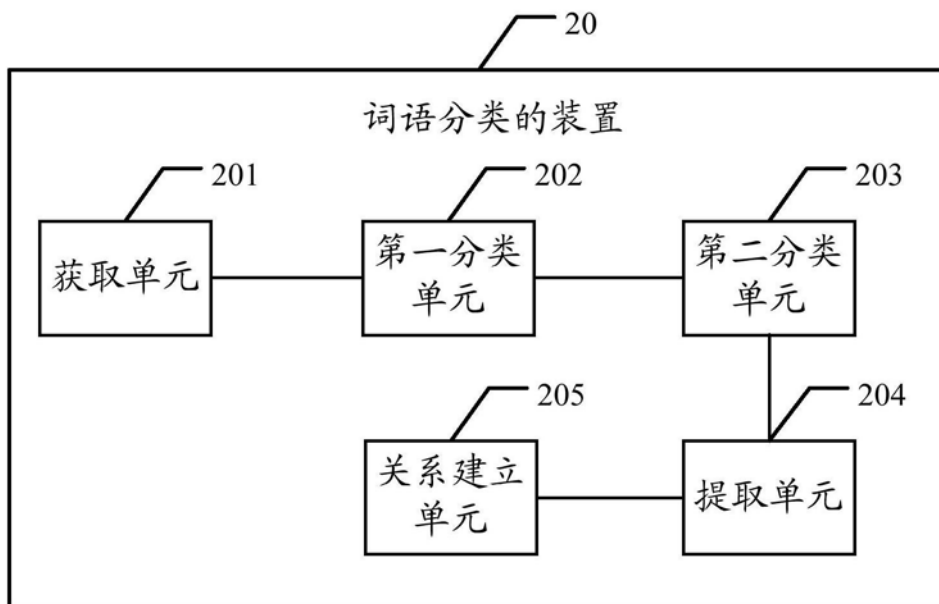


图8

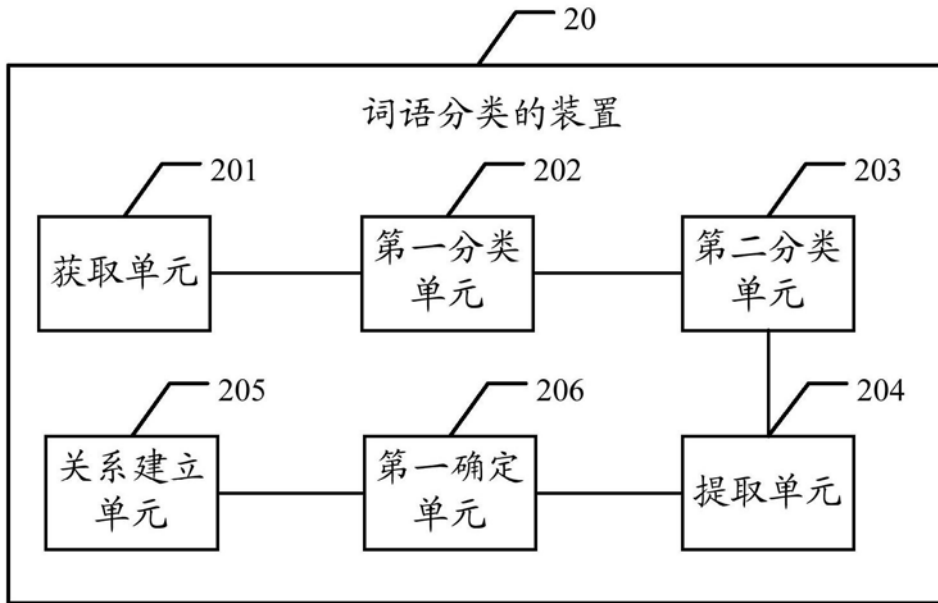


图9

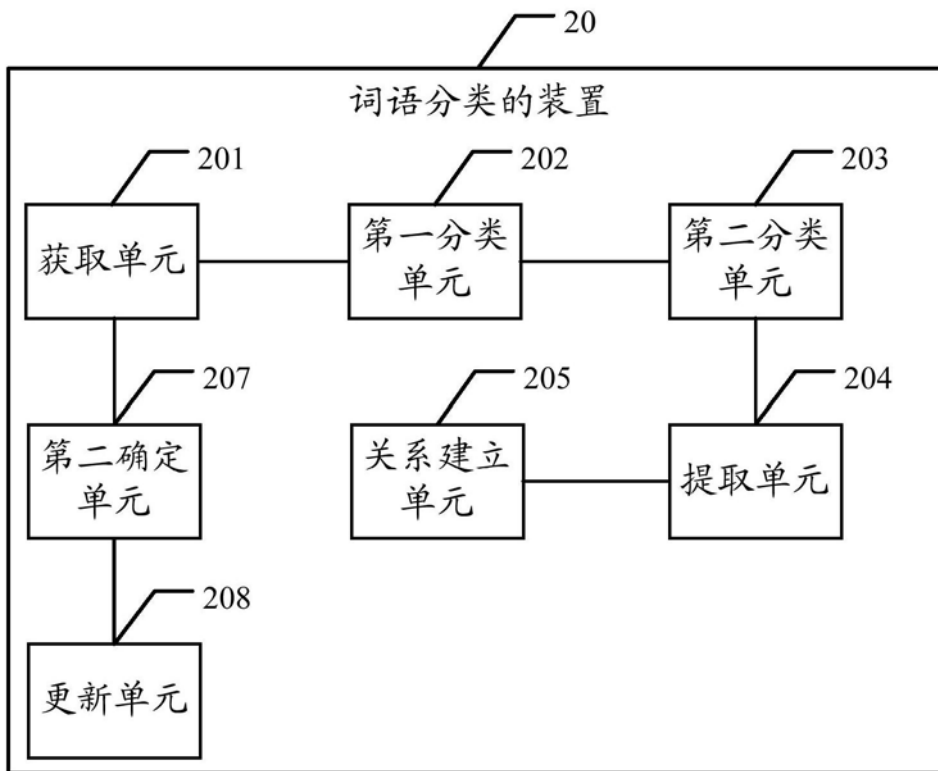


图10

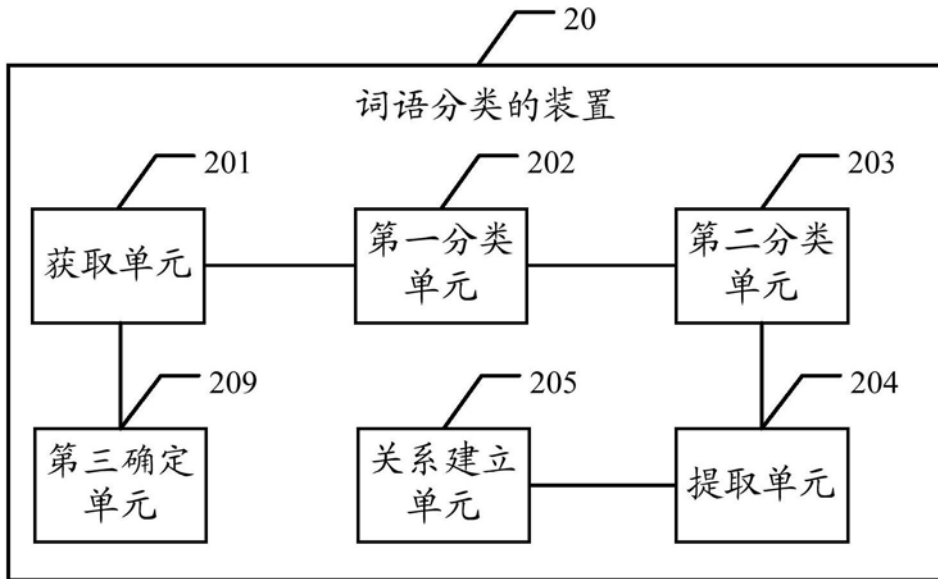


图11

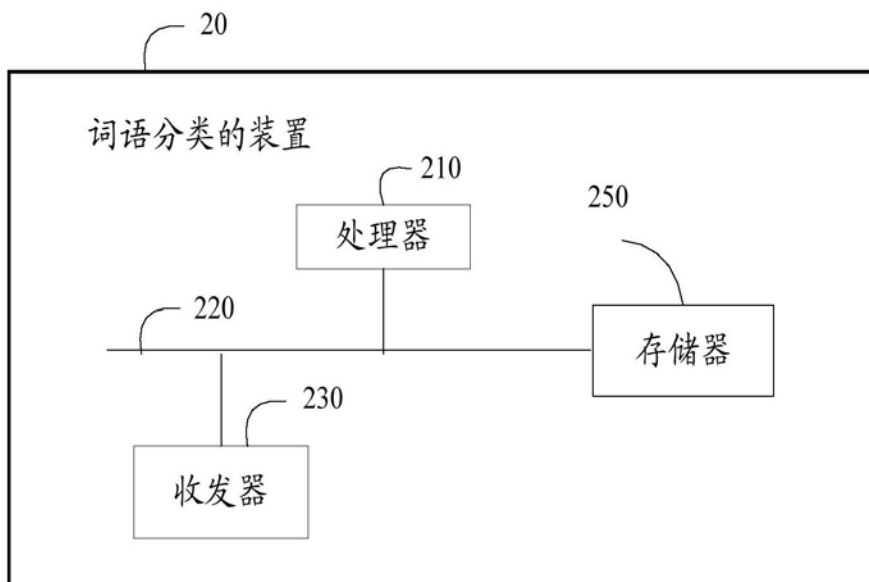


图12