

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2019 年 11 月 14 日 (14.11.2019)



(10) 国际公布号
WO 2019/214236 A1

- (51) 国际专利分类号:
G06F 17/27 (2006.01)
- (21) 国际申请号: PCT/CN2018/121321
- (22) 国际申请日: 2018 年 12 月 14 日 (14.12.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810447372.7 2018年5月11日 (11.05.2018) CN
- (71) 申请人: 北京三快在线科技有限公司 (**BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD**) [CN/CN]; 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。
- (72) 发明人: 苏婧 (**SU, Jing**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。于志安 (**YU, Zhian**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。王强 (**WANG, Qiang**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。吴尚 (**WU, Shang**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。侯培旭 (**HOU, Peixu**); 中国北京市海淀区北四环西路

9 号 2106-030, Beijing 100080 (CN)。李春阳 (**LI, Chunyang**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。王燕华 (**WANG, Yanhua**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。陈文石 (**CHEN, Wenshi**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。

(74) 代理人: 北京曼威知识产权代理有限公司 (**BEIJING MAVAIP INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 411 室, Beijing 100085 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** USER-GENERATED CONTENT SUMMARY DETERMINING AND USER-GENERATED CONTENT RECOMMENDING

(54) 发明名称: 原创内容摘要确定和原创内容推荐

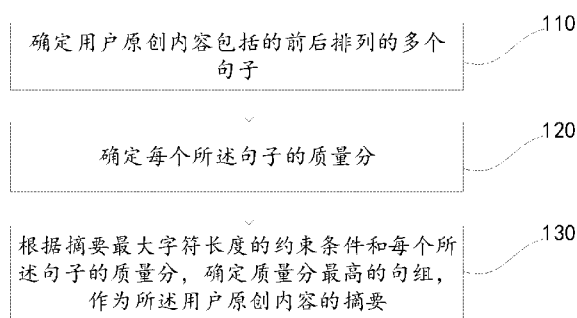


图 1

- 110 Determine a plurality of sentences, arranged in order, included in the user-generated content
- 120 Determine a quality score of each of the sentences
- 130 According to the constraint condition of the maximum character length of the summary and the quality score of each of the sentences, determine a sentence group with the highest quality score as a summary of the user-generated content

(57) **Abstract:** A user-generated content summary determining method, the method comprising: determining a plurality of sentences, arranged in order, included in the user-generated content (110); then determining a quality score of each of the sentences (120); and finally, according to the constraint condition of the maximum character length of the summary and the quality score of each of the sentences, determining a sentence group with the highest quality score as a summary of the user-generated content (130), wherein the sentences included in the sentence group are continuous.

(57) **摘要:** 一种原创内容摘要确定方法, 所述方法包括: 确定用户原创内容包括的前后排列的多个句子 (110); 然后, 确定每个所述句子的质量分 (120); 最后, 根据摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要 (130), 其中, 所述句组中包括的句子是连续的。

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

原创内容摘要确定和原创内容推荐

相关申请的交叉引用

[01] 本专利申请要求于 2018 年 5 月 11 日提交的、申请号为 201810447372.7、发明名称为“原创内容摘要确定方法及装置，原创内容推荐方法及装置”的中国专利申请的优先权，该申请的全文以引用的方式并入本文中。

技术领域

[02] 本申请涉及计算机技术领域的原创内容摘要确定方法及装置，原创内容推荐方法及装置。

背景技术

10 [03] 摘要是一篇文章或一段文字的简要描述，通常表达了文章或文字的核心含义。可以把文章自动生成摘要的方法看作是一个信息压缩过程，将输入的文章或文字压缩为一篇简短的摘要，该过程不可避免有信息损失。

发明内容

[04] 本申请提供一种原创内容摘要确定方法和装置以及用户原创内容推荐方法和装置。

15 [05] 第一方面，本申请实施例提供了一种原创内容摘要确定方法包括：确定用户原创内容包括的前后排列的多个句子；确定每个所述句子的质量分；根据摘要最大字符长度的约束条件和每个所述句子的质量分，确定质量分最高的句组，作为所述用户原创内容的摘要，其中，所述句组中包括的句子是连续的。

[06] 第二方面，本申请实施例提供了一种原创内容摘要确定装置，包括：句子确定模块，
20 用于确定用户原创内容包括的前后排列的多个句子；句子质量分确定模块，用于确定每个所述句子的质量分；摘要确定模块，用于根据摘要最大字符长度的约束条件和每个所述句子的质量分，确定质量分最高的句组，作为所述用户原创内容的摘要，其中，所述句组中包括的句子是连续的。

[07] 第三方面，本申请实施例还公开了一种用户原创内容推荐方法，包括：确定用户的目标商户；根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容；确

25

定所述候选用户原创内容中与所述用户匹配的目标用户原创内容；根据本申请实施例所述用户原创内容摘要确定方法，确定与所述目标用户原创内容的摘要；向所述用户推荐所述目标用户原创内容的摘要。

- 5 [08] 第四方面，本申请实施例还公开了一种用户原创内容推荐装置，包括：目标商户确定模块，用于确定用户的目标商户；候选用户原创内容确定模块，用于根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容；匹配候选用户原创内容确定模块，用于确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容；原创内容摘要确定模块，用于根据本申请实施例所述用户原创内容摘要确定方法，确定所述目标用户原创内容的摘要；推荐模块，用于向所述用户推荐所述目标用户原创内容的摘要。
- 10 [09] 第五方面，本申请实施例还公开了一种电子设备，包括存储器、处理器及存储在所述存储器上并可在处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现本申请实施例所述的用户原创内容摘要确定方法和用户原创内容推荐方法。

- [10] 第六方面，本申请实施例提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时本申请实施例公开的原创内容摘要确定方法和用户原创内容推荐方法的步骤。
- 15

[11] 本申请实施例公开的用户原创内容摘要确定方法，通过确定用户原创内容包括的前后排列的多个句子；然后，确定每个所述句子的质量分；最后，根据摘要最大字符长度的约束条件，确定所述质量分最高的句组，作为所述用户原创内容的摘要，其中，所述句组中包括的句子是连续的。该方法能够高效并准确地提取用户原创内容摘要。

20 附图说明

[12] 为了更清楚地说明本申请实施例的技术方案，下面将对实施例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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

- 25 [13] 图 1 是本申请实施例一的用户原创内容摘要确定方法流程图；

[14] 图 2 是本申请实施例二的用户原创内容摘要确定方法流程图；

[15] 图 3 是本申请实施例三的用户原创内容推荐方法流程图；

[16] 图 4 是本申请实施例四的用户原创内容推荐方法流程图；

[17] 图 5 是本申请实施例五的用户原创内容摘要确定装置的结构示意图之一；

[18] 图 6 是本申请实施例六的用户原创内容推荐装置的结构示意图之一；

[19] 图 7 是本申请实施例六的用户原创内容推荐装置的结构示意图之二。

具体实施方式

5 [20] 下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[21] 在确定摘要的过程中，为了保留尽可能多的重要信息，常用的方法包括信息抽取、
10 文章分类和词法分析等，然后根据获取的信息生成摘要。与传统文章相比，用户原创内容 UGC(User created Content)具有篇幅更短、段落不明显、句子结构不规范、用词相对随意的特点，传统的提取文章或文字摘要的方法可能无法准确提取出用户原创内容的摘要。

实施例一

15 [22] 本实施例公开一种原创内容摘要确定方法，如图 1 所示，该方法包括：步骤 110 至步骤 130。

[23] 步骤 110，确定用户原创内容包括的前后排列的多个句子。

[24] 在一实施例中，首先对用户原创内容进行数据处理，提取出所述用户原创内容中的句子，并将提取出的句子按照各句子在所述用户原创内容中出现的先后顺序进行排列。

20 [25] 由于用户原创内容，如用户点评，没有固定格式要求，所以内容和格式多样。在一实施例中，以预设标点符号作为句子之间的分隔标记，将所述用户原创内容划分为多个句子。其中，所述预设标点符号包括但不限于以下任意一种或多种：句号、感叹号、问号、逗号、空格、分号、顿号、省略号、表情符号、波浪符号。标准标点符号至少包括
25 句号、感叹号、问号、逗号、分号、顿号、冒号、省略号。在一实施例中，先用标准标点符号对用户原创内容进行分句，如果分句后句子还是过长采用其他符号再次分句。按照各句子在所述用户原创内容中出现位置的前后顺序进行排列，得到所述用户原创内容包括的前后排列的 M 个句子。其中，M 为大于等于 1 的自然数。

[26] 步骤 120, 确定每个所述句子的质量分。

[27] 在一实施例中, 可以从句子包括的文本、观点和实体等信息维度的特征确定所述句子的质量分。其中, 文本进一步可以包括: 位置、长度、关键词情感属性、关键词对商户特征的描述等维度的信息。观点维度的信息可以为观点中包括的评价对象、评价词等

5 信息。实体维度的信息可以为实体词的出现频次、实体词的类型等维度的信息。

[28] 句子的质量分用于表示该句子对所述用户原创内容的核心思想的贡献或表现能力。

[29] 步骤 130, 根据摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要, 其中, 所述句组中包括的句子是连续的。

[30] 在确定了用户原创内容中包括的前后排列的多个句子之后, 选择信息含量最高的句

10 组作为所述用户原创内容的摘要。在一实施例中, 通过滑动窗口, 找到包含的字符长度满足预设字符长度条件的多组句组。然后, 根据每组句组中各句子的质量分, 确定所述句组的评分。最后, 选择评分最高的句组, 作为所述用户原创内容的摘要。

[31] 本申请实施例公开的用户原创内容摘要确定方法, 通过确定用户原创内容包括的前后排列的至少一个句子, 确定每个所述句子的质量分; 根据摘要最大字符长度的约束条

15 件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要, 能够高效并准确地提取用户原创内容的摘要。

实施例二

[32] 本实施例公开的一种原创内容摘要确定方法, 如图 2 所示, 该方法包括: 步骤 210 至步骤 240。

20 [33] 步骤 210, 构建评价对象库、评价词库和实体词库。

[34] 在一实施例中, 为了确定用户原创内容中包括的句子的质量分, 首先构建评价对象库、评价词库和实体词库, 再基于评价对象库、评价词库和实体词库, 确定句子中包括的实体、评价对象, 以及句子中包括的情感类的关键词等。

[35] 在一实施例中, 根据海量用户在平台上生成的数亿条 UGC 评论和每日千万级别的

25 查询关键词, 使用词法分析器得到名词和形容词等关键词, 结合预设 POI 知识库的内容, 使用 N-Gram 技术得到 UGC 评论中的所述关键词和所述查询关键词的词性类别(例如: 景点、电影院、商区、商场等)。然后, 通过评价对象挖掘, 可以建成一个覆盖率比较高的评价对象库, 为后续评论挖掘提供支持。

[36] 实体是评价对象中的一个子集，选自于商户、用户等的结构化数据中的关键词，例如：商家名称、菜品类别、菜品名称等。

[37] 关键词是指 UGC 文本经过分词后的有意义的词。评价词是指形容词、副词和成语等关键词。在一实施例中，获取 UGC 评论中高频的评价词，统计这些评价词在 5 星评论和 1 星评论中的分布情况，得到评价词的极性（正面、负面和中性）。比如“很好”这个评价词出现在好评的评论中的数量要远大于在差评中的数量，则“很好”这个评价词的极性为正面。通过评价词挖掘，可以建成一个评价词库，为后续评论挖掘提供支持。通过评价词可以确定句子的情感信息。

[38] 步骤 220，确定用户原创内容包括的前后排列的多个句子。

10 [39] 在一实施例中，首先对用户原创内容进行数据处理，提取出所述用户原创内容中的句子，并将提取出的句子按照各句子在所述用户原创内容中出现的先后顺序进行排列。

[40] 由于用户原创内容，如用户点评，没有固定格式要求，所以内容和格式多样。在一实施例中，按照预设标点符号作为句子之间的分隔标记，将所述用户原创内容划分为多个句子。其中，所述预设标点符号包括但不限于以下任意一种或多种：句号、感叹号、15 问号、逗号、空格、分号、顿号、冒号、省略号、表情符号、波浪符号。标准标点符号至少包括句号、感叹号、问号、逗号、分号、顿号、冒号、省略号。在一实施例中，先用标准标点符号对用户原创内容进行分句，如果分句后句子还是过长采用其他符号再次分句。按照各句子在所述用户原创内容中出现位置的前后顺序进行排列，得到所述用户原创内容包括的前后排列的 M 个句子。其中，M 为大于等于 1 的自然数。

20 [41] 在一实施例中，确定用户原创内容包括的前后排列的至少一个句子包括：基于标准标点符号对用户原创内容进行分句，得到所述用户原创内容包括的第一句子；基于扩展标点符号对所述第一句子中字符长度大于预设句子字符长度阈值的第一句子进行再次分句，得到所述第一句子对应的第二句子；将所述第一句子中字符长度未进行再次分句的所述第一句子和所述第二句子，按照在所述用户原创内容中出现位置的前后顺序进行25 排列，得到所述用户原创内容包括的前后排列的 M 个句子。其中，M 为大于等于 1 的自然数。其中，标准标点符号至少包括：句号、逗号、问号、感叹号、省略号、冒号、顿号和分号，扩展标点符号包括：空格、表情符号、破浪符号等。

[42] 以一条用户原创内容为“地道巴蜀陈年酸菜，三年发酵而成，配合来自越南的无污染的龙利鱼^_^味道鲜嫩无比！”且预设句子字符长度阈值为 10 为例，说明如何确定用

户原创内容包括的前后排列的多个句子。首先，基于标准标点符号对用户原创内容进行分句，可以得到“地道巴蜀陈年酸菜”、“三年发酵而成”和“配合来自越南的无污染的龙利鱼^_^味道鲜嫩无比”共 3 个第一句子。对于第一句子“配合来自越南的无污染的龙利鱼^_^味道鲜嫩无比”，其字符长度为 21，大于预设句子字符长度阈值，因此需要基于扩展标点符号进一步对其进行句子划分。由于该句子中包括一个表情符号“^_^”，因此，该句子基于该扩展标点符号进行划分后，得到 2 个第二句子，分别为：“配合来自越南的无污染的龙利鱼”和“味道鲜嫩无比”。最后，确定该用户原创内容中包括的 4 个句子为：第一句子“地道巴蜀陈年酸菜”、“三年发酵而成”，以及第二句子“配合来自越南的无污染的龙利鱼”和“味道鲜嫩无比”。之后，按照上述 4 个句子在所述用户原创内容中出现位置的前后顺序进行排列，得到所述用户原创内容包括的前后排列的 4 个句子，分别为：地道巴蜀陈年酸菜”、“三年发酵而成”、“配合来自越南的无污染的龙利鱼”和“味道鲜嫩无比”。

[43] 步骤 230，确定每个所述句子的质量分。

[44] 句子的质量分用于表示该句子对所述用户原创内容的核心思想的贡献或表现能力。在一实施例中，确定每个所述句子的质量分，包括：根据所述句子的预设维度的信息，确定所述句子的质量分，其中，所述预设维度包括以下维度中的一个或多个：文本、实体和观点。根据所述句子的预设维度的信息，确定所述句子的质量分，包括：对所述句子的实体维度评分和观点维度评分进行加权求和得到初始质量分；通过所述句子的文本维度评分对所述初始质量分进行调整；并将调整后的初始质量分确定为所述句子的质量分。

[45] 在一实施例中，对所述句子的实体维度评分和观点维度评分进行加权求和得到初始质量分，并通过所述句子的文本维度评分对所述初始质量分进行调整，并将调整后的初始质量分确定为所述句子的质量分，包括根据以下公式确定所述句子的质量分：

$$score(sentence_i) = w' \times (\alpha \times score_sentence_i(word \in \text{实体}) + \beta \times score_sentence_i(word \in \text{评价对象}))$$

[46] 其中， $score(sentence_i)$ 表示句子 i 的质量分， $score_sentence_i(word \in \text{实体})$ 表示句子 i 的实体维度评分， $score_sentence_i(word \in \text{评价对象})$ 表示句子 i 的观点维度评分， w' 表示句子 i 的文本维度评分。

[47] 其中，评价对象为句子 i 中包括的观点包括的评价对象， α 表示实体维度评分对应的第一权重调节因子， β 表示观点维度评分对应的第二权重调节因子。即，首先，通过

以下公式计算初始值质量分:

$$\alpha \times score_sentence_i(word \in \text{实体}) + \beta \times score_sentence_i(word \in \text{评价对象})。$$

[48] 然后, 通过文本维度评分 w' 对初始值质量分进行调整, 得到句子 i 的质量分。

[49] 在一实施例中, 根据句子在所述用户原创内容中的前后位置、句子的负面情感信息、

5 商户特色信息确定句子的文本维度评分包括: 提升靠近用户原创内容首部的句子的质量分、降低含有负面情感信息的句子的质量分、提升包括商户特色信息的句子的质量分。

例如, 对于出现在用户原创内容中的前三个句子, 则提高这前三个句子的质量分, 如加 10 分, 以此提升用户原创内容首部位置句子出现在摘要中的概率。例如, 如果句子中包

10 括预设评价词库中的负面词语, 则确定所述句子包含负面情感, 通过降低句子的质量分, 如减 20 分, 降低这个句子出现在摘要中的概率。如果句子中包括预设评价词库中的广告词语, 则通过降低句子的质量分, 如减 10 分, 降低该句子出现在摘要中的概率。再例如, 如果句子中含有商户排名前三的推荐菜, 或者含有商户类目下特色的评价对象, 提高该句子的质量分, 如加 10 分, 从而提升该句子出现在摘要中的概率。

15 [50] 实体维度评分反映了实体在用户原创内容中的权重。在一实施例中, 根据句子中包括的实体词的逆向文本词频确定句子的实体维度评分。例如, 实体维度评分为所述句子中包括的实体的逆向文本词频之和, 通过以下公式确定句子的实体维度评分:

$$score_sentence_i(word \in \text{实体}) = \sum_{word \in \text{实体}} idf(word_j)$$

[51] 该公式中, $idf(word_j)$ 为句子包括的实体词 $word_j$ 的逆向文本词频。其中, 所述实体的逆向文本词频可通过如下公式确定:

20

$$idf(word_j) = \log \frac{|shop_num|}{1 + |\{k : word(j) \in shop_k\}|}$$

[52] 该公式中, $|shop_num|$ 为所有用户原创内容覆盖的商户总数, $\{k : word(j) \in shop_k\}$ 表示出现关键词 $word(j)$ 的商户总数。

[53] 在一实施例中, 根据句子中包括的观点涉及的评价对象的逆向文本词频确定句子的观点维度评分。

25 [54] 观点维度评分反映了观点中的评价对象在用户原创内容中的权重。在一实施例中, 根据句子中包括的评价对象词的逆向文本词频确定句子的观点维度评分。例如, 所述观点维度的信息为所述句子中包括的观点所涉及的评价对象的逆向文本词频之和, 通过如

下公式确定句子的观点维度评分:

$$score_sentence_i(word \in \text{评价对象}) = \sum_{word \in \text{评价对象}} idf(word_i)$$

[55] 该公式中 $idf(word_i)$ 为句子包括的评价对象 $word_i$ 的逆向文本词频。其中, 所述评价对象的逆向文本词频通过如下公式确定:

$$idf(word_i) = \log \frac{|shop_num|}{1 + |\{k : word(l) \in shop_k\}|}$$

[56] 该公式中, $|shop_num|$ 为所有用户原创内容覆盖的商户总数, $\{k : word(l) \in shop_k\}$ 表示出现关键词 $word(l)$ 的商户总数。

[57] 在一实施例中, 根据句子中包括的观点涉及的评价对象的逆向文本词频确定句子的观点维度评分。例如, 通过公式确定句子的观点维度评分,

$$score_sentence_i(word \in \text{评价对象}) = \sum_{word \in \text{评价对象}} idf(word_i)$$

[58] 该公式中, $idf(word_i)$ 为句子包括的评价对象 $word_i$ 的逆向文本词频。

[59] 通过上述公式可以看出, 如果实体或评价对象出现在用户原创内容 (如商户评论) 中的频率低, 则相应的实体维度评分或观点维度评分权重高。进一步的, 通过对实体维度评分和观点维度评分进行加权求和, 得到句子的质量分。在一实施例中, 实体维度评分和观点维度评分的权值通过经验统计设置。

[60] 步骤 240, 根据摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要, 其中, 所述句组中包括的句子是连续的。

[61] 在确定了用户原创内容中包括的前后排列的多个句子之后, 选择信息含量最高的句组作为所述用户原创内容的摘要。

[62] 在一实施例中, 通过如下公式确定 $begin$ 和 end 之间的句组, 作为所述用户原创内容的摘要:

$$\begin{cases} \arg \max(begin, end) = w \times \sum_{i=begin}^{end} score(sentence_i) \\ \text{s.t. } 0 \leq begin < N, \sum_{begin}^{end} length(sentence_i) < max_length \end{cases}$$

[63] 其中, $begin$ 和 end 是所述用户原创内容中句子的顺序号, max_length 为预设摘要

最大字符长度, $length(sentence_i)$ 为句子 i 中的字符长度, w 是总分调节因子, w 根据句子 $sentence_i, begin \leq i \leq end$ 是否含有实体和观点、以及 $\sum_{begin}^{end} length(sentence_i)$ 确定。

[64] 根据摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要, 包括: 通过滑窗技术确定满足摘要最大字符长度的约束条件的至少一组句组; 针对各个所述句组, 将所述句组中包括的各个句子的质量分的加权和确定为所述句组的质量分; ; 确定质量分最高的句组作为所述用户原创内容的摘要。在一实施例中, 所述句组的质量分中各个句子的质量分的权重根据所述句组中各个句子是否含有实体和观点、所述句组的字符长度、所述句组中是否包含所述用户原创内容的首个句子或末尾句子中的任意一项或多项因素确定。

10 [65] 在一实施例中, 假设预设摘要最大字符长度为 35, 以某条用户原创内容中包括前后排列的 9 个句子, 每个句子的质量分和字符长度如下表所示为例, 说明确定摘要的方法。其中, 句子编号 1 至 9 为句子的前后排列序号, 且各个句子的质量分的权重相同, 例如, 均为 1。

	句子 1	句子 2	句子 3	句子 4	句子 5	句子 6	句子 7	句子 8	句子 9
字符长度	10	9	6	8	16	7	8	9	10
质量分	0.5	0.2	1	2	-10	2	3	3	2

15 [66] 在一实施例中, 首先, 从句子 1 开始, 通过调整窗口的长度, 找到长度不超过 35 个字符的句组, 如{句子 1},{句子 1, 句子 2},{句子 1, 句子 2, 句子 3},{句子 1, 句子 2, 句子 3, 句子 4}。然后, 分别确定每组句组的质量分, 并保留质量分最高的一组句组, 如{句子 1, 句子 2, 句子 3, 句子 4}组成的句组作为候选摘要, 所述候选摘要的质量分为 3.7 分。

20 [67] 接下来, 滑动窗口, 从句子 2 开始, 通过调整窗口的长度, 找到长度不超过 35 个字符的句组, 如{句子 2},{句子 2, 句子 2},{句子 2, 句子 3, 句子 4}。然后, 分别确定每组句组的质量分, 并保留质量分最高的一组句组, 如{句子 2, 句子 3, 句子 4}组成的句组, 质量分为 3.2 分。

25 [68] {句子 1, 句子 2, 句子 3, 句子 4}组成的候选摘要的质量分大于{句子 2, 句子 3, 句子 4}组成的句组的质量分(3.2 分), 因此, 暂时保留{句子 1, 句子 2, 句子 3, 句子 4}句组构成的候选摘要。

[69] 以此类推，通过滑窗技术，分别确定以每个句子开始的长度不超过 35 个字符的多组句组，并确定每组句组的质量分，以通过质量分更高的句组更新暂时保留的候选摘要，直至最后找到最高得分的句组，作为所述用户原创内容的摘要。以上表中的句子为例，最终将确定质量分为 10 分的句组{句子 6, 句子 7, 句子 8, 句子 9}作为所述用户原创内容的摘要。

[70] 在一实施例中，根据摘要最大字符长度的约束条件和每个所述句子的质量分，确定质量分最高的句组，作为所述用户原创内容的摘要，包括：通过滑窗技术确定满足摘要最大字符长度的约束条件的一个或多个句组；针对各个所述句组，将所述句组中包括的各个句子的质量分的加权和确定为所述句组的质量分；确定所述质量分最高的句组，作为所述用户原创内容的摘要。

[71] 在确定所述句组的质量分时，所述句组中各个句子的质量分可具有相同的权重，也可以具有不同的权重。

[72] 在一实施例中，假设所述句组中各个句子的质量分具有相同的权重，该权重与句组字符长度和预设摘要最大字符长度的比值为 T ， T 为大于 1 的数，如 $T=1.5$ ，这样可以避免确定的摘要的字符长度过短。在一实施例中，假设所述句组中各个句子的质量分具有不同的权重，如果句子的实体维度评分为零，例如该句子中不包括实体，则降低所述句子的质量分的权重；如果句子的观点维度评分为零，例如句子中不包括评价对象，则降低所述句子的质量分的权重；如果句子是所述用户原创内容的第一个句子或最后一个句子，则提升所述句子的质量分的权重。根据所述句子是否是所述用户原创内容的首个句子或末尾句子确定该句子的质量分的权重，可以提升确定的摘要中句子的完整性。

[73] 本申请实施例公开的用户原创内容摘要确定方法，通过确定用户原创内容包括的前后排列的多个句子；然后，确定每个所述句子的质量分；最后，根据摘要最大字符长度的约束条件和每个所述句子的质量分，确定质量分最高的句组，作为所述用户原创内容的摘要，能够高效并准确地提取用户原创内容的摘要题。本申请实施例中，通过所述用户原创内容的文本、实体和观点三个维度加权计算得到句子的质量分，通过这种方法，能找到用户原创内容中信息价值密度最高的句组。并且，本申请实施例公开的用户原创内容摘要确定方法，支持标点符号使用不规范，甚至语句不通顺的用户原创内容摘要的抽取，鲁棒性更强；可以根据对摘要长度的不同要求，自适应抽取商户特色的用户原创内容摘要。

实施例三

[74] 本实施例公开的一种原创内容推荐方法，如图 3 所示，该方法包括：步骤 310 至步骤 350。

[75] 步骤 310，确定用户的目标商户。

5 [76] 在一实施例中，首先根据用户的历史行为数据，确定用户发生过预设历史行为的商户，作为第一目标商户；然后，确定与所述第一目标商户相似的商户，作为第二目标商户；最后，将所述第一目标商户和所述第二目标商户，作为所述用户的目标商户。

[77] 步骤 320，根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容。

10 [78] 获取所述目标商户的用户原创内容，并进一步确定每条用户原创内容的评价得分。在一实施例中，可以根据用户原创内容的文本信息、实体信息以及观点信息等确定用户原创内容的评价得分。在一实施例中，评价得分越高表示所述用户原创内容的质量越高，即所述用户原创内容展示给用户的信息更有价值。然后，按照用户原创内容的评价得分由高到低的顺序，对每个所述目标用户的用户原创内容分别进行排序。之后，对于每一个目标用户，分别选择评价得分最高的预设数量的用户原创内容，作为候选用户原创内容。

15 [79] 步骤 330，确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容。

[80] 在一实施例中，可以分别提取用户的特征向量，以及每一条候选用户原创内容的特征向量，然后，通过计算用户的特征向量与每一条候选用户原创内容的特征向量之间的相似度，确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容。在一实施例中，可以通过计算用户的特征向量与候选用户原创内容的特征向量之间的相似度距离，
20 确定用户与某一条候选用户原创内容的之间的匹配度；或者，通过预先训练的机器学习排序模型，根据输入的用户的特征向量和某一条所述候选用户原创内容的特征向量，计算用户与所述某一条候选用户原创内容之间的匹配度。

[81] 然后，选择与所述用户匹配度最高的一个或预设数量的所述候选用户原创内容，作为所述目标用户原创内容。

25 [82] 步骤 340，确定所述目标用户原创内容的摘要。

[83] 根据实施例一和实施例二所述用户原创内容摘要确定方法，确定所述目标用户原创内容的摘要。

[84] 步骤 350，向所述用户推荐所述目标用户原创内容的摘要。

[85] 在确定了与所述用户匹配的目标用户原创内容后，向所述用户推荐所述目标用户原创内容的摘要。

[86] 本申请实施例公开的用户原创内容推荐方法，通过确定用户的目标商户；根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容；确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容；最后，向所述用户推荐所述目标用户原创内容的摘要，其中，所述目标用户原创内容的摘要根据实施例一或实施例二所述的用户原创内容摘要确定方法确定，这样，与根据用户原创内容的热度为用户推荐用户原创内容的方案相比，实现了根据用户需求推荐更准确的用户原创内容。本申请实施例公开的用户原创内容推荐方法，通过把与用户匹配的用户原创内容推荐给用户，实现了有针对性的进行信息推荐，有效提升了用户原创内容推荐的准确性。同时，通过在为

5 用户推荐原创内容时，仅展示原创内容的摘要，简洁清晰的为用户展示推荐的关键信息，便于用户准确快速的做出决策，进一步提升了用户体验。

10

实施例四

[87] 本实施例公开的一种用户原创内容推荐方法，如图 4 所示，该方法包括：步骤 410 至步骤 470。

15

[88] 步骤 410，构建评价对象库、评价词库和实体词库。

[89] 构建评价对象库、评价词库和实体词库的具体实施方式参见实施例二，本实施例不再赘述。

[90] 步骤 420，确定用户的目标商户。

[91] 在一实施例中，确定用户的目标商户，包括：确定所述用户产生过预设行为的商户，作为第一目标商户；基于商户向量的相似度，确定与所述第一目标商户相似的

20

第二目标商户；将所述第一目标商户和所述第二目标商户，作为所述用户的目标商户。

[92] 在一实施例中，首先根据用户的历史行为数据，确定用户发生过预设历史行为的商户，作为第一目标商户。其中，用户产生过预设行为的商户包括但不限于：用户点击过的商户、用户浏览过的商户、用户收藏过的商户、用户购买过商品的商户。

25

[93] 然后，进一步确定与所述第一目标商户相似的商户，作为第二目标商户。

[94] 在一实施例中，基于商户向量的相似度，确定与所述第一目标商户相似的

第二目标商户之前，还包括：将用户点击的商户序列作为词向量模型输入，训练商户向量模型；

通过所述商户向量模型确定商户的商户向量。

[95] 在一实施例中，把用户在商户上的行为转变为时间序列事件，然后，把时间序列事件作为输入，采用深度学习算法训练商户向量模型，即把商户特征从高维的离散空间映射到低维的连续空间。例如，当用户先后点击了商户 1、商户 2 和商户 3，那么，可以
5 把商户 1、商户 2 和商户 3 的商户标识序列作为输入样本，用于训练商户向量模型。然后，通过预先训练的商户向量模型，可以获得某个商户标识对应的商户向量。

[96] 在确定了每个商户的商户向量之后，通过计算各个商户向量与所述第一目标商户之间的相似度，可以确定与所述第一目标商户相似的第二目标商户。

[97] 最后，将所述第一目标商户和所述第二目标商户，作为所述用户的目标商户。例如，
10 根据用户的历史行为，确定用户曾经点击过商户 1，则将商户 1 作为用户的第一目标商户。然后，通过计算商户向量的相似度，确定与商户 1 相似的商户 2，则将商户 2 作为所述用户的第二目标商户。最后，将商户 1 和商户 2 作为所述用户的目标商户。

[98] 步骤 430，根据所述目标商户的用户原创内容的文本、实体和观点三个维度的信息，确定所述用户原创内容的评价得分。

15 [99] 根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容之前，还包括：根据所述目标商户的用户原创内容的文本、实体和观点三个维度的信息，确定所述用户原创内容的评价得分。例如，根据所述目标商户的用户原创内容的文本、实体和观点三个维度的信息，确定所述用户原创内容的评价得分，可以包括：通过对用户原创内容的文本得分、实体得分和观点得分进行加权求和，所述用户原创内容的评价得分。

20 [100] 在一实施例中，首先，对于平台的用户原创内容，如用户点评，选取最近预设时间（如半年内）的用户原创内容。然后，根据用户原创内容中的文本、实体和观点三个维度的信息，确定所述用户原创内容的评价得分。因为优质商户或者高星级用户下也存在低质的用户原创内容，所以在对用户原创内容进行评分时，不考虑商户和用户的特征，只从用户原创内容的内容质量本身来分析，通过文本、实体和观点三个维度计算得到用
25 户原创内容的评价得分。

[101] 在一实施例中，文本得分与用户原创内容中包含的不同文字的数量成正比。即，用户原创内容中包含的不同文字越多，文本得分越高。根据用户原创内容中包含的不同文字的数量确定文本得分，可以有效过滤掉用户重复使用同一个标点符号或者文字来充当字数的用户原创内容。

[102] 在一实施例中，实体得分可以通过用户原创内容中包含的实体的逆向文本词频表示；观点得分可以通过用户原创内容中包含的观点所涉及的评价对象的逆向文本词频表示。

[103] 在确定实体得分和观点得分之前，首先，将用户原创内容划分为多个句子。将用户原创内容划分为多个句子的具体方法可以参考实施例二中确定用户原创内容中的句子的方法，本实施例不再赘述。

[104] 然后，通过预设的实体词库，确定由用户原创内容中划分得到的每个句子中包括的实体和观点。

[105] 实体是指用户原创内容中涉及的评论对象，例如，商户名、地址、类目、商场、星级酒店、商场、小区、电影院、行政区和城市等。实体是用户原创内容中的重要信息，例如，一条用户原创内容中提到的推荐菜、地址和类目等信息，可以作为该条用户原创内容的重要特征。O2O（online to offline, O2O）场景下的信息抽取有别于传统的人名、地名和公司名识别，需要挖掘不同维度下不同关键词的权重信息，例如在美食品类下的商家评论中，“龙之梦”的评论数很少，其逆向文本词频要高于“粤菜”。在一实施例中，可以通过如下公式确定一条用户原创内容的实体得分：

$$score_ugc = \sum_{word \in \text{实体}} idf(word_p)$$

[106] 该公式中， $idf(word_p)$ 为该条用户原创内容包括的实体词 $word_p$ 的逆向文本词频。其中，所述实体词的逆向文本词频通过如下公式确定。

$$idf(word_p) = \log \frac{|shop_num|}{1 + |\{k : word(p) \in shop_k\}|}$$

[107] 该公式中， $|shop_num|$ 为所有用户原创内容覆盖的商户总数， $\{k : word(p) \in shop_k\}$ 表示出现关键词 $word_p$ 的商户总数。

[108] 观点表示对具体的评价对象的主客观判断信息，本申请中，主要从句子中抽取观点。例如，对于一条用户原创内容中的一个句子“浓缩咖啡豆是皮爷家的经典”，从该句子中抽取观点的具体方法如下：根据预先构建的评价对象库可以确定该句子中包括的评价对象是：咖啡豆；根据预先构建的评价词库可以确定该句子中包括的评价词是：“浓缩”、“经典”；将该句子中包括的评价对象和评价词组合，得到该句子中包括的观点，即：“咖啡豆-经典”和“咖啡豆-浓缩”。再后，根据上述两个观点在所述用户原创内容中出现的比率得到每个观点的置信度，在一实施例中，观点出现越频繁则置信度越高。最

后得到一条所述用户原创内容中所有的观点，以及每个观点的置信度。

[109] 对于一条用户原创内容中得到的每个观点，通过对所述观点包括的评价对象和评价词的词向量进行求和，得到所述观点的向量表示。通过向量对观点进行表示之后，就可以采用余弦定理计算向量之间的距离，来判断观点之间的相似关系。在一实施例中，

5 通过对句子进行分析，可以得到如下观点数据结构表：

字段名称	字段说明	示例
Opinion	观点	咖啡豆-经典
SemanticVector	词向量	[0,1,0.32,0.16,0.07...]
Aspect	评价对象	咖啡豆
Evaluate	评价词	经典
Confidence	置信度	0.87
Updatetime	更新时间	2018-03-12 09:00:00

[110] 在一实施例中，基于用户产生的全部用户原创内容，通过分词处理后得到训练样本，使用本领域技术人员熟知的词向量技术，得到训练样本中每个关键词的词向量。在一实施例中，关键词包括实体词、评价词以及各种有意义的通用词汇。词向量是关键词的向量表示。在一实施例中，关键词的词向量为固定长度的浮点型一维向量，例如，采用 skip-gram 模型的负采样方法训练词向量模型。采用词向量技术后，所有关键词都可以用一个固定长度的向量表示，将原来稀疏的巨大维度压缩到一个更小维度空间，例如“披萨”和“pizza”这两个词在文本上没有相似性，但是通过词向量表示后，其语义距离比较接近。

[111] 最后，通过对一条用户原创内容中包括的实体的实体得分、观点的观点得分，以及文本得分进行加权求和，将得到的总得分和作为该条用户原创内容的评价得分。在一实施例中，对实体得分、观点得分以及文本得分进行加权，各项得分的加权值根据具体需求设置，通常，观点得分的权值最高，文本得分的权值最低。

[112] 步骤 440, 根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容。

[113] 如前所述，假设将商户 1 和商户 2 作为用户的目标商户，则进一步根据用户原创内容的评价得分，在所述商户 1 和商户 2 的用户原创内容中，分别选择评价得分满足预设条件的多条用户原创内容作为用户的候选用户原创内容。例如，按照评价得分由高到低的顺序，分别对商户 1 和商户 2 的用户原创内容排序，然后选择商户 1 评价得分最高

的 M 条用户原创内容和商户 2 评价得分最高的 M 条用户原创内容，作为候选用户原创内容。

[114] 步骤 450，确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容。

5 [115] 在一实施例中，确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容，包括：根据每条所述候选用户原创内容的排序特征和所述用户的用户特征，分别确定每条所述候选用户原创内容与所述用户的匹配度；确定所述匹配度满足预设条件的所述候选用户原创内容，作为与所述用户匹配的目标用户原创内容。

10 [116] 在一实施例中，可以首先基于用户原创内容的排序特征和用户的用户特征通过机器学习训练匹配度识别模型。例如，将用户原创内容的排序特征和发布该原创内容的用户的用户特征组合为正样本，将用户原创内容的排序特征和踩了该原创内容的用户的用户特征组合为负样本，训练匹配度识别模型。然后，通过该匹配度识别模型基于输入的用户原创内容的排序特征和用户的用户特征识别所述用户原创内容和所述用户的匹配度。其中，所述排序特征包括：点赞数、评论数、分享数、文本质量分、图片质量分、实体词、用户原创内容发布者等级、发布者与所述用户的关系中的任意一项或多项；所
15 述用户特征包括：用户历史行为特征、商区偏好特征、类目偏好特征、相似用户特征中的任意一项或多项，所述用户历史行为特征包括：搜索、浏览、购买、到店行为中的任意一项或多项的特征。

[117] 在一实施例中，可以确定所述匹配度得分最高的预设数量的所述候选用户原创内容，作为与所述用户匹配的目标用户原创内容；或者，确定每个商户对应的所述候选用户原创内容中，与所述用户的所述匹配度得分最高的一条所述候选用户原创内容，作为
20 与所述用户匹配的目标用户原创内容。由于进行匹配度识别时，结合了用户偏好、用户社交关系等特征，因此，所确定的目标用户原创内容，是用户偏好的用户原创内容。

[118] 步骤 460，确定所述目标用户原创内容的摘要。

25 [119] 在一实施例中，通过是实施例一和实施例二所述的用户原创内容摘要确定方法确定所述目标用户原创内容的摘要，本实施例中，对摘要的具体确定方法不再赘述。

[120] 步骤 470，向所述用户推荐所述目标用户原创内容的摘要。

[121] 在确定了与所述用户匹配的目标用户原创内容后，向所述用户推荐所述目标用户原创内容的摘要。

[122] 本申请实施例公开的用户原创内容推荐方法，通过确定用户的目标商户；然后，

确定所述目标商户的用户原创内容的评价得分，并根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容；确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容及其摘要；最后，向所述用户推荐所述目标用户原创内容的摘要。这样，与根据用户原创内容的热度为用户推荐用户原创内容的方案相比，能够根据用户需求推荐更加准确的用户原创内容。本申请实施例公开的用户原创内容推荐方法，通过把与用户匹配的用户原创内容推荐给用户，实现了有针对性的进行信息推荐，有效提升了用户原创内容推荐的准确性。同时，通过在为用户推荐用户原创内容时，仅展示用户原创内容的摘要，简洁清晰的为用户展示推荐的关键信息，便于用户准确快速的做出决策，进一步提升了用户体验。

- 5 [123] 通过用户原创内容的文本、实体和观点的信息，确定用户原创内容的评价得分，能够提升用户原创内容质量评价的准确性，进一步提升用户原创内容推荐的准确性。

实施例五

[124] 本实施例公开的一种原创内容摘要确定装置，如图 5 所示，所述装置包括：

[125] 句子确定模块 510，用于确定用户原创内容包括的前后排列的至少一个句子。

- 15 [126] 句子质量分确定模块 520，用于确定每个所述句子的质量分。

[127] 摘要确定模块 530，用于根据摘要最大字符长度的约束条件和每个所述句子的质量分，确定质量分最高的句组，作为所述用户原创内容的摘要，其中，所述句组中包括的句子是连续的。

[128] 可选的，所述句子质量分确定模块 520 进一步用于：

- 20 [129] 根据每个所述句子的预设维度的信息，确定每个所述句子的质量分，其中，所述预设维度包括以下维度中的一个或多个：文本、实体和观点。

[130] 可选的，所述根据每个所述句子的预设维度的信息，确定每个所述句子的质量分，包括：对每个所述句子的实体维度评分和观点维度评分进行加权求和得到初始质量分，并通过所述句子的文本维度评分对所述初始质量分进行调整；并将调整后的初始质量分
25 确定为所述句子的质量分。在本申请的一个实施例中，所述对每个所述句子的实体维度评分和观点维度评分进行加权求和得到初始质量分，并通过所述句子的文本维度评分对所述初始质量分进行调整；并将调整后的初始质量分确定为所述句子的质量分，进一步包括：

[131] 根据如下公式确定每个所述句子的质量分:

$$score(sentence_i) = w' \times (\alpha \times score_sentence_i(word \in \text{实体}) + \beta \times score_sentence_i(word \in \text{评价对象}))$$

[132] 其中, $score(sentence_i)$ 表示句子 i 的质量分, $score_sentence_i(word \in \text{实体})$ 表示句子 i 的实体维度评分, $score_sentence_i(word \in \text{评价对象})$ 表示句子 i 的观点维度评分, w' 表示句子 i 的文本维度评分。其中, 评价对象为句子中包括的观点针对的评价对象, α 表示实体维度评分对应的第一权重调节因子, β 表示观点维度评分对应的第二权重调节因子。

[133] 可选的, 所述摘要确定模块 530 进一步用于:

[134] 通过滑窗技术确定满足摘要最大字符长度的约束条件的一个或多个句组;

10 [135] 针对各个所述句组, 将所述句组中包括的各个句子的质量分的加权和确定为所述句组的质量分;

[136] 确定所述质量分最高的句组, 作为所述用户原创内容的摘要。

[137] 可选的, 所述句组的质量分中各个句子的质量分的权重根据所述句组中各个句子是否含有实体和观点、所述句组的字符长度、所述句组中是否包含所述用户原创内容的首个句子或末尾句子中的任意一项或多项因素确定。

15 [138] 本实施例是与实施例一和实施例二对应的装置实施例, 本实施例中各模块的具体实现方式参见实施例一和实施例二中的相关步骤的描述, 此处不再赘述。

[139] 通过确定用户原创内容包括的前后排列的多个句子并确定每个所述句子的质量分, 然后, 根据摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要, 其中, 所述句组中包括的句子是连续的。本公开实例中的用户原创内容推荐装置解决了无法准确提取用于原创内容摘要的问题。经过大量用户原创内容的测试, 本申请公开的用户原创内容摘要确定装置, 可以高效、准确的确定用户原创内容的摘要。通过用户原创内容的文本、实体和观点三个维度加权计算得到句子质量的方法, 本公开实例能找到用户原创内容中信息价值密度最高的句组。并且, 25 本申请实施例公开的远传内容摘要确定方法, 支持标点符号使用不规范, 甚至语句不通顺的用户原创内容摘要的抽取, 鲁棒性更强; 可以根据对摘要长度的不同要求, 自适应抽取商户特色的用户原创内容摘要。

实施例六

[140] 本实施例公开的一种原创内容推荐装置，如图 6 所示，所述装置包括：

[141] 目标商户确定模块 610，用于确定用户的目标商户。

[142] 候选用户原创内容确定模块 620，用于根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容。

5 [143] 匹配候选用户原创内容确定模块 630，用于确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容。

[144] 原创内容摘要确定模块 640，用于根据本申请实施例所述用户原创内容摘要确定方法，确定所述目标用户原创内容的摘要。

10 [145] 推荐模块 650，用于向所述用户推荐所述目标用户原创内容的摘要，其中，所述目标用户原创内容的摘要根据实施例一和实施例二所述的用户原创内容摘要确定方法确定。

[146] 可选的，如图 7 所示，所述装置还包括：

[147] 用户原创内容评价得分确定模块 660，用于根据所述用户原创内容的文本、实体和观点三个维度的信息，确定所述用户原创内容的评价得分。

15 [148] 可选的，所述目标商户确定模块 610 进一步用于：

[149] 确定所述用户产生过预设行为的商户，作为第一目标商户；基于商户向量的相似度，确定与所述第一目标商户相似的第二目标商户；将所述第一目标商户和所述第二目标商户，作为所述用户的目标商户。

[150] 可选的，所述目标商户确定模块 610 还用于：

20 [151] 将用户点击的商户序列作为词向量模型的输入，训练商户向量模型；通过所述商户向量模型确定所述第一目标商户的所述商户向量。

[152] 可选的，所述匹配候选用户原创内容确定模块 630 进一步用于：

25 [153] 根据每条所述候选用户原创内容的排序特征和所述用户的用户特征，分别确定每条所述候选用户原创内容与所述用户的匹配度；确定所述匹配度满足预设条件的所述候选用户原创内容，作为与所述用户匹配的目标用户原创内容。

[154] 其中，所述排序特征包括：点赞数、评论数、分享数、文本质量分、图片质量分、实体词、用户原创内容发布者等级、发布者与所述用户的关系中的任意一项或多项；所

述用户特征包括：用户历史行为特征、商区偏好特征、类目偏好特征、相似用户特征中的任意一项或多项，所述用户历史行为特征包括：搜索、浏览、购买、到店行为中的任意一项或多项的特征。

5 [155] 本实施例是与实施例三和实施例四对应的装置实施例，本实施例中各模块的具体实现方式参见实施例三和实施例四中的相关步骤的描述，此处不再赘述。

[156] 通过确定用户的目标商户；然后，确定所述目标商户的用户原创内容的评价得分，并根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容；确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容及其摘要；最后，向所述用户推荐所述目标用户原创内容的摘要。本公开实例中的用户原创内容推荐装置解决了根据用
10 户原创内容的热度为用户推荐用户原创内容时，推荐的用户原创内容不准确，无法满足用户需求的问题。通过把与用户匹配的用户原创内容推荐给用户，实现了有针对性的进行信息推荐，本公开实例中的用户原创内容推荐装置有效提升了用户原创内容推荐的准确性。同时，通过在为用户推荐原创内容时，仅展示原创内容的摘要，简洁清晰的为用户展示推荐的关键信息，便于用户准确快速的做出决策，进一步提升了用户体验。

15 [157] 通过用户原创内容的文本、实体和观点的信息，确定用户原创内容的评价得分，能够提升用户原创内容质量评价的准确性，进一步提升用户原创内容推荐的准确性。

[158] 相应的，本申请还公开了一种电子设备，包括存储器、处理器及存储在所述存储器上并可在处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如本申请实施例一和实施例二所述的原创内容摘要确定方法、实施例三和实施例四所述的用于
20 原创内容推荐方法。所述电子设备可以为 PC 机、移动终端、个人数字助理、平板电脑等。

[159] 本申请还公开了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本申请实施例一和实施例二所述的原创内容摘要确定方法或实施例三和实施例四所述的用户原创内容推荐方法。

25 [160] 本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[161] 以上对本申请提供的一种用户原创内容摘要确定方法及装置，用户原创内容推荐

方法及装置进行了详细介绍，本文中应用了具体个例对本申请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申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，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申请的限制。

- 5 [162] 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件实现。基于这样的理解，上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如 ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络
- 10 络设备等）执行各个实施例或者实施例的某些部分所述的方法。

权利要求书

1. 一种用户原创内容的摘要确定方法, 包括:
确定用户原创内容包括的前后排列的多个句子;
确定每个所述句子的质量分;
- 5 根据摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要, 其中, 所述句组中包括的句子是连续的。
2. 根据权利要求 1 所述的方法, 确定所述句子的质量分, 包括:
根据所述句子的预设维度的信息, 确定所述句子的质量分,
其中, 所述预设维度包括以下维度中的一个或多个: 文本、实体和观点。
- 10 3. 根据权利要求 2 所述的方法, 根据所述句子的预设维度的信息, 确定所述句子的质量分, 包括:
对所述句子的实体维度评分和观点维度评分进行加权求和, 得到初始质量分;
通过所述句子的文本维度评分对所述初始质量分进行调整;
将调整后的所述初始质量分确定为所述句子的质量分。
- 15 4. 根据权利要求 1 所述的方法, 根据所述摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的所述句组, 作为所述用户原创内容的摘要, 包括:
通过滑窗技术确定满足所述摘要最大字符长度的约束条件的一个或多个句组;
针对各个所述句组, 将所述句组中包括的各个句子的质量分的加权和确定为所述句组的质量分;
- 20 确定所述质量分最高的句组, 作为所述用户原创内容的摘要。
5. 根据权利要求 4 所述的方法, 所述句组中包括的各个句子的质量分的权重由以下任意一项或多项因素确定:
针对所述句组中包括的各个句子, 所述句子是否含有实体和观点;
所述句组的字符长度; 和
- 25 所述句组中是否包含所述用户原创内容的首个句子或末尾句子。
6. 一种用户原创内容推荐方法, 包括:
确定用户的目标商户;
根据所述目标商户的用户原创内容的评价得分, 确定候选用户原创内容;
确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容;
- 30 根据权利要求 1 至 5 任一项所述用户原创内容摘要确定方法, 确定所述目标用户原创内容的摘要;

向所述用户推荐所述目标用户原创内容的摘要。

7. 根据权利要求6所述的方法, 所述方法还包括:

根据所述用户原创内容的文本、实体和观点三个维度的信息, 确定所述用户原创内容的评价得分。

5 8. 根据权利要求6所述的方法, 确定所述用户的目标商户, 包括:

确定所述用户产生过预设行为的商户, 作为第一目标商户;

基于商户向量的相似度, 确定与所述第一目标商户相似的第二目标商户;

将所述第一目标商户和所述第二目标商户, 作为所述用户的目标商户。

9. 根据权利要求8所述的方法, 所述方法还包括:

10 将用户点击的商户序列作为词向量模型的输入, 训练商户向量模型;

通过所述商户向量模型确定所述第一目标商户的所述商户向量。

10. 根据权利要求6所述的方法, 确定所述候选用户原创内容中与所述用户匹配的所述目标用户原创内容, 包括:

根据每条所述候选用户原创内容的排序特征和所述用户的用户特征, 分别确定每条

15 所述候选用户原创内容与所述用户的匹配度;

确定所述匹配度满足预设条件的所述候选用户原创内容, 作为与所述用户匹配的所述目标用户原创内容;

其中, 所述排序特征包括: 点赞数、评论数、分享数、文本质量分、图片质量分、实体词、用户原创内容发布者等级、发布者与所述用户的关系中的任意一项或多项;

20 所述用户特征包括: 用户历史行为特征、商区偏好特征、类目偏好特征、相似用户特征中的任意一项或多项;

所述用户历史行为特征包括: 搜索、浏览、购买、到店行为中的任意一项或多项的特征。

11. 一种原创内容摘要确定装置, 包括:

25 句子确定模块, 用于确定用户原创内容包括的前后排列的多个句子;

句子质量分确定模块, 用于确定每个所述句子的质量分;

摘要确定模块, 用于根据摘要最大字符长度的约束条件和每个所述句子的质量分, 确定质量分最高的句组, 作为所述用户原创内容的摘要, 其中, 所述句组中包括的句子是连续的。

30 12. 根据权利要求11所述的装置, 所述句子质量分确定模块进一步用于:

根据每个所述句子的预设维度的信息, 确定每个所述句子的质量分, 其中, 所述预

设维度包括以下维度中的一个或多个：文本、实体和观点。

13. 根据权利要求 12 所述的装置，根据每个所述句子的预设维度的信息，确定每个所述句子的质量分，包括：

对所述句子的实体维度评分和观点维度评分进行加权求和，得到初始质量分；

5 通过所述句子的文本维度评分对所述初始质量分进一步加权调整；

将调整后的所述初始质量分确定为所述句子的质量分。

14. 根据权利要求 11 所述的装置，所述摘要确定模块进一步用于：

通过滑窗技术确定满足所述摘要最大字符长度的约束条件的一个或多个句组；

10 针对各个所述句组，将所述句组中包括的各个句子的质量分的加权和确定为所述句组的质量分；

确定所述质量分最高的句组。

15. 根据权利要求 14 所述的装置，所述句组中包括的各个句子的质量分的权重由以下任意一项或多项因素确定：

针对所述句组中包括的各个句子，所述句子是否含有实体和观点；

15 所述句组的字符长度；和

所述句组中是否包含所述用户原创内容的首个句子或末尾句子。

16. 一种用户原创内容推荐装置，包括：

目标商户确定模块，用于确定用户的目标商户；

20 候选用户原创内容确定模块，用于根据所述目标商户的用户原创内容的评价得分，确定候选用户原创内容；

匹配候选用户原创内容确定模块，用于确定所述候选用户原创内容中与所述用户匹配的目标用户原创内容；

原创内容摘要确定模块，用于根据权利要求 1 至 5 任一项所述用户原创内容摘要确定方法，确定所述目标用户原创内容的摘要；

25 推荐模块，用于向所述用户推荐所述目标用户原创内容的摘要。

17. 根据权利要求 16 所述的装置，还包括：

用户原创内容评价得分确定模块，用于根据所述用户原创内容的文本、实体和观点三个维度的信息，确定所述用户原创内容的评价得分。

18. 根据权利要求 16 所述的装置，所述目标商户确定模块进一步用于：

30 确定所述用户产生过预设行为的商户，作为第一目标商户；

基于商户向量的相似度，确定与所述第一目标商户相似的第二目标商户；

将所述第一目标商户和所述第二目标商户，作为所述用户的目标商户。

19. 根据权利要求 16 所述的装置，所述目标商户确定模块还用于：

将用户点击的商户序列作为词向量模型的输入，训练商户向量模型；

通过所述商户向量模型确定所述第一目标商户的所述商户向量。

5 20. 根据权利要求 16 所述的装置，所述匹配候选用户原创内容确定模块进一步用于：

根据每条所述候选用户原创内容的排序特征和所述用户的用户特征，分别确定每条所述候选用户原创内容与所述用户的匹配度；

10 确定所述匹配度满足预设条件的所述候选用户原创内容，作为与所述用户匹配的所述目标用户原创内容；

其中，所述排序特征包括：点赞数、评论数、分享数、文本质量分、图片质量分、实体词、用户原创内容发布者等级、发布者与所述用户的关系中的任意一项或多项；

所述用户特征包括：用户历史行为特征、商区偏好特征、类目偏好特征、相似用户特征中的任意一项或多项；

15 所述用户历史行为特征包括：搜索、浏览、购买、到店行为中的任意一项或多项的特征。

20 21. 一种电子设备，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现权利要求 1 至 5 任意一项所述的用户原创内容摘要确定方法或实现权利要求 6 至 10 任意一项所述的用户原创内容推荐方法。

22. 一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现权利要求 1 至 5 任意一项所述用户原创内容摘要确定方法或实现权利要求 6 至 10 任意一项所述用户原创内容推荐方法。

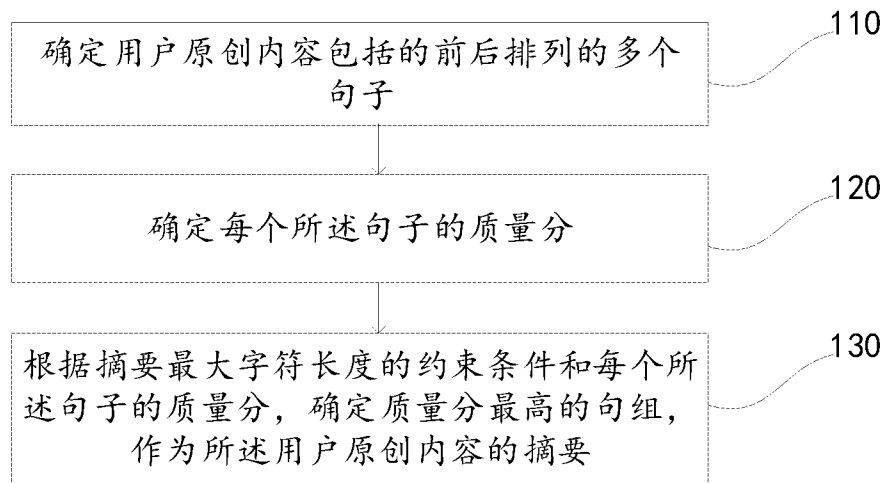


图 1

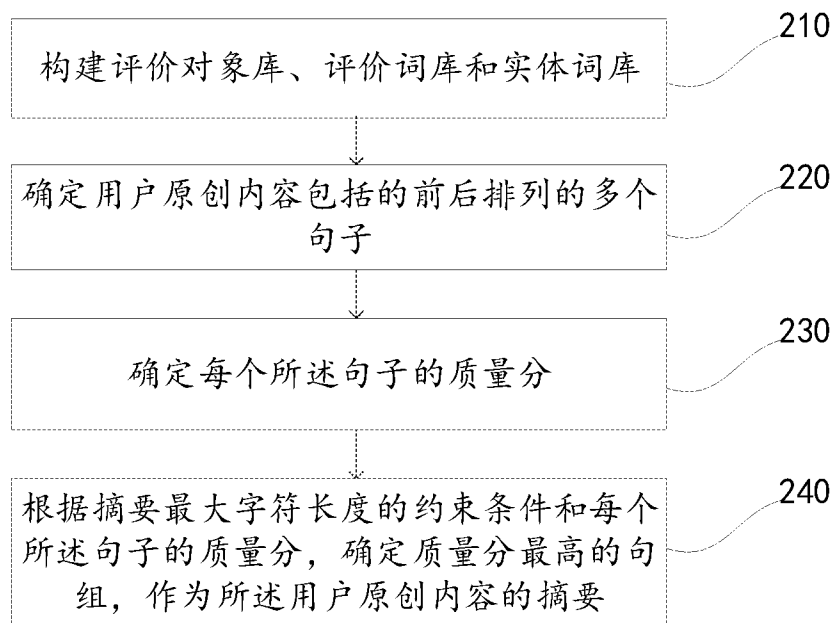


图 2

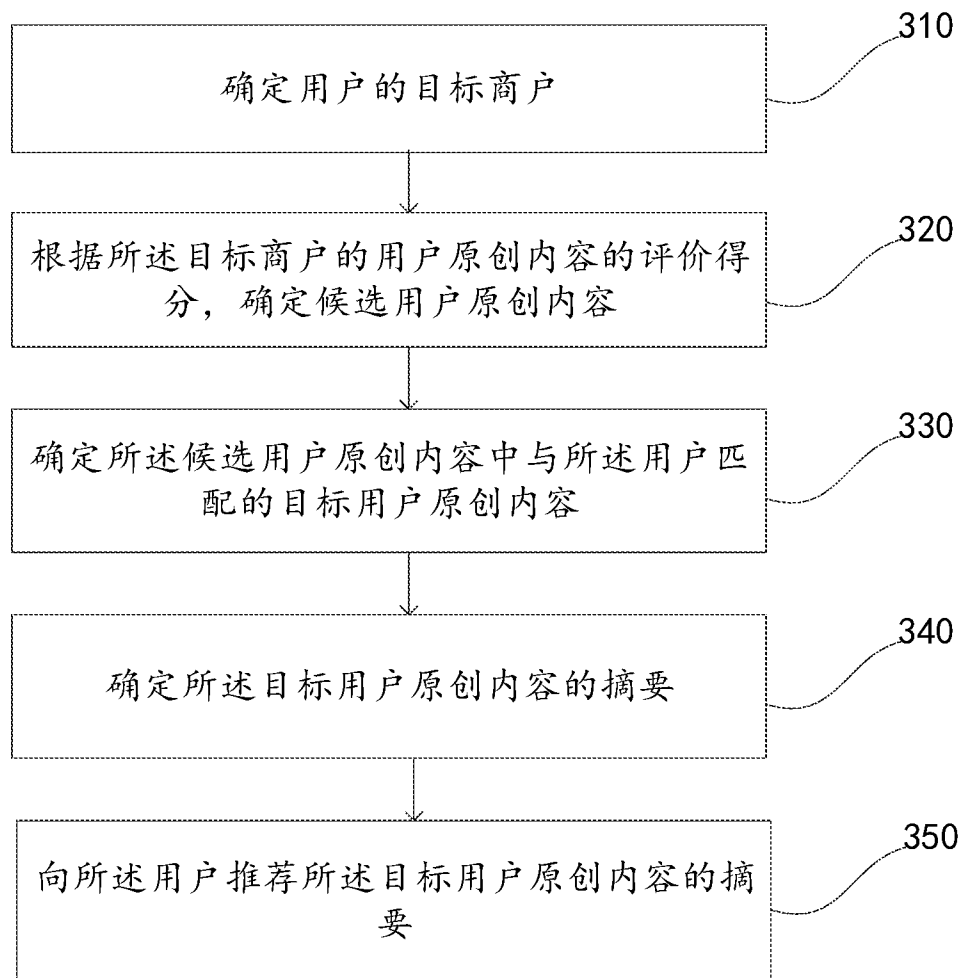


图 3



图 4

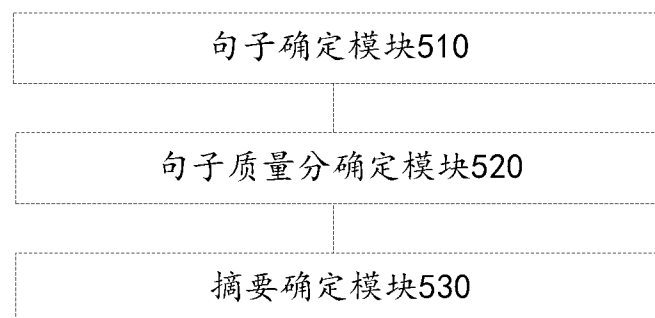


图 5

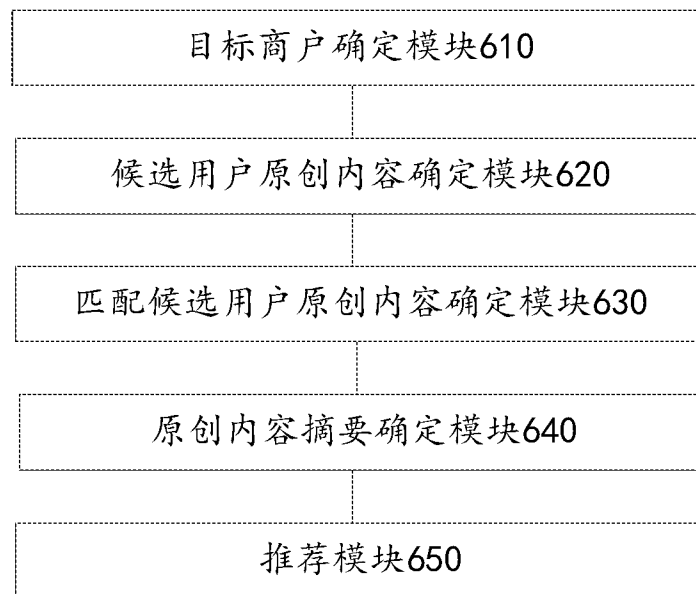


图 6

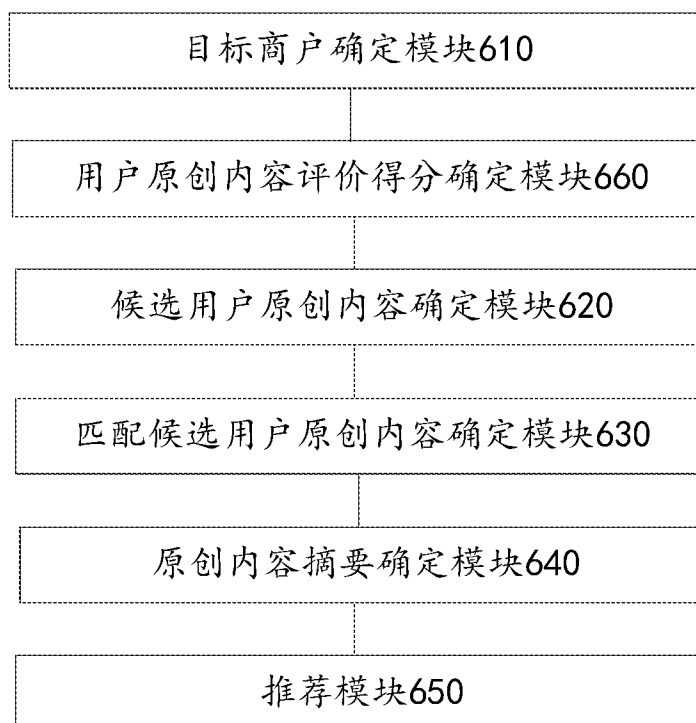


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/121321

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/27(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI, IEEE: 摘要, 提取, 抽取, 句子, 语句, 质量, 分数, 得分, 连续, 句组, 组合, abstract, extract, sentence, quality, score, successive, group

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 108628833 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 09 October 2018 (2018-10-09) claims 1-22	1-22
Y	CN 105868175 A (LETV INFORMATION TECHNOLOGY (BEIJING) CO., LTD.) 17 August 2016 (2016-08-17) description, paragraphs [0057]-[0065]	1-22
Y	CN 107609960 A (KOUBEI (SHANGHAI) INFORMATION TECHNOLOGY CO., LTD.) 19 January 2018 (2018-01-19) description, paragraphs [0088]-[0145] and [0238]-[0289]	1-22
A	CN 101667194 A (PEKING UNIVERSITY) 10 March 2010 (2010-03-10) entire document	1-22



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

27 February 2019

Date of mailing of the international search report

13 March 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/121321

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108628833	A	09 October 2018	None			
CN	105868175	A	17 August 2016	WO	2017092316	A1	08 June 2017
				US	2017161259	A1	08 June 2017
CN	107609960	A	19 January 2018	None			
CN	101667194	A	10 March 2010	None			

国际检索报告

国际申请号

PCT/CN2018/121321

A. 主题的分类

G06F 17/27 (2006.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

WPI, EPDOC, CNPAT, CNKI, IEEE: 摘要, 提取, 抽取, 句子, 语句, 质量, 分数, 得分, 连续, 句组, 组合, abstract, extract, sentence, quality, score, successive, group

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 108628833 A (北京三快在线科技有限公司) 2018年 10月 9日 (2018 - 10 - 09) 权利要求1-22	1-22
Y	CN 105868175 A (乐视网信息技术北京股份有限公司) 2016年 8月 17日 (2016 - 08 - 17) 说明书第[0057]-[0065]段	1-22
Y	CN 107609960 A (口碑上海信息技术有限公司) 2018年 1月 19日 (2018 - 01 - 19) 说明书第[0088]-[0145]、 [0238]-[0289]段	1-22
A	CN 101667194 A (北京大学) 2010年 3月 10日 (2010 - 03 - 10) 全文	1-22

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 2月 27日

国际检索报告邮寄日期

2019年 3月 13日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

王晓敏

电话号码 86-(10)-53961359

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/121321

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	108628833	A	2018年 10月 9日	无			
CN	105868175	A	2016年 8月 17日	WO	2017092316	A1	2017年 6月 8日
				US	2017161259	A1	2017年 6月 8日
CN	107609960	A	2018年 1月 19日	无			
CN	101667194	A	2010年 3月 10日	无			