(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
2020 年 6 月 25 日 (25.06.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/125015 A1

(51) 国际专利分类号:
G06F 16/9535 (2019.01) *G06Q 30/00* (2012.01)

(21) 国际申请号: PCT/CN2019/098210

(22) 国际申请日: 2019 年 7 月 29 日 (29.07.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811545314.4 2018年12月17日 (17.12.2018) 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: 阎晗 (**YAN, Han**); 中国北京市海淀区北四环西路9号2106-030, Beijing 100080 (CN)。 杨涵 (**YANG, Han**); 中国北京市海淀区北四环西路9号2106-030, Beijing 100080 (CN)。

(74) 代理人: 北京润泽恒知识产权代理有限公司 (**BEIJING RUN ZEHENG INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区中关村南大街 31 号神舟大厦 702, Beijing 100081 (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, 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)。

(54) **Title:** CONTEXTUALIZED MERCHANT RECALL

(54) **发明名称**: 场景化商家召回

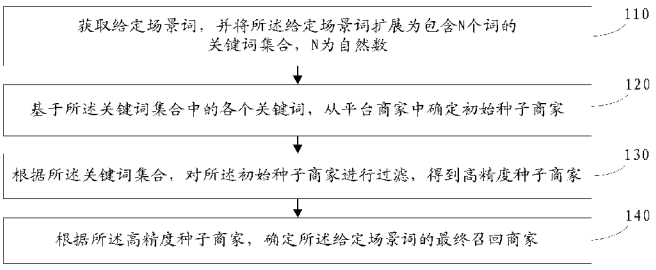


图 1

110 Acquire given scenario words, and expand the given scenario words to keyword sets comprising N words, N being a natural number

120 On the basis of each keyword in the keyword sets, determine initial seed merchants from within a platform merchant

130 Filter the initial seed merchants according to the keyword sets to obtain a highly precise seed merchant

140 According to the highly precise seed merchant, determine a final recall merchant for the given scenario words

(57) **Abstract:** Disclosed in the present application are a contextualized merchant recall method and device, an electronic device and a readable storage medium. The method comprises: acquiring given scenario words, and expanding the given scenario words to keyword sets comprising N words, N being a natural number; on the basis of each keyword in the keyword sets, determining initial seed merchants from within a platform merchant; filtering the initial seed merchants according to the keyword sets to obtain a highly precise seed merchant; and according to the highly precise seed merchant, determining a final recall merchant for the given scenario words.

(57) **摘要**: 本申请公开了一种场景化商家召回方法、装置、电子设备及可读存储介质。所述方法，包括：获取给定场景词，并将所述给定场景词扩展为包含N个词的关键词集合，N为自然数；基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家；根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家；根据所述高精度种子商家，确定所述给定场景词的最终召回商家。

WO 2020/125015 A1

[见续页]

本国际公布:
— 包括国际检索报告(条约第21条(3))。

场景化商家召回

本申请要求在 2018 年 12 月 17 日提交中国专利局、申请号为 201811545314.4、发明名称为“场景化商家召回方法、装置、电子设备及可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及数据挖掘技术领域，具体涉及场景化商家召回。

背景技术

商家召回是 O2O (Online To Offline，在线离线/线上到线下)等互联网模式中的重要问题。其中，场景化商家召回是基于不同场景对互联网平台上入驻的商家进行分类召回，可实现对商家更好的管理。在现有的场景化商家召回相关技术中，一般是采用有监督学习方法将物品与每个场景的相关度获取转化为分类任务，将属于每个场景类别的概率作为商家跟场景的相关度；或者，通过无监督学习方法，如通过直接将物品文本和场景文本表征为向量，进而获取相似度。

但是，采用有监督学习方法需要人工标注大量数据，耗费大量人力并且时间周期长；而采用无监督学习方法生成的结果无法兼顾精度和召回率，召回商家列表从头部到腰部相关度衰减很快，无法产出工业界能够直接使用的结果。而且业务常用的相关度获取方法主要是应用于纯文本召回，比如新闻、文档、文章等召回，而很多互联网平台的召回物品是商家，文本分布在不同的属性上，各属性对物品的重要性不同，并且不同属性文本中包含大量冗余信息和特殊停用词，需要对文本进行大量针对性优化。

发明内容

依据本申请第一方面，提供了一种场景化商家召回方法，包括:

获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数;

基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家;

根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家;

根据所述高精度种子商家，确定所述给定场景词的最终召回商家。

根据本申请的第二方面，提供了一种场景化商家召回装置，包括:

关键词集合扩展模块，用于获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数;

初始种子商家确定模块，用于基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家;

高精度种子商家确定模块，用于根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家;

最终召回商家确认模块，用于根据所述高精度种子商家，确定所述给定场景词的最终召回商家。

根据本申请的第三方面，提供了一种电子设备，包括:

处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的获取机程序，所述处理器执行所述程序时实现前述的场景化商家召回方法。

根据本申请的第四方面，提供了一种可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行前述的场景化商家召回方法。

根据本申请的场景化商家召回方法，可以获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数；基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家；根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家；根据所述高精度种子商家，确定所述给定场景词的最终召回商家，取得了降低时间成本以及人力成本，同时兼顾精度和召回率并提高适应性的有益效果。

上述说明仅是本申请技术方案的概述，为了能够更清楚了解本申请的技术手段，而可依照说明书的内容予以实施，并且为了让本申请的上述和其它目的、特征和优点能够更明显易懂，以下特举本申请的具体实施方式。

附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本申请的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中:

图 1 示出了根据本申请一个实施例的一种场景化商家召回方法的步骤流程图;

图 2 示出了根据本申请一个实施例的一种场景化商家召回方法的步骤流程图;

图 3 示出了根据本申请一个实施例的一种场景化商家召回装置的结构示意图; 以及

图 4 示出了根据本申请一个实施例的电子设备的示意图;

图 5 示出了用于执行根据本申请的方法的另一电子设备的框图; 以及,

图 6 示出了用于保持或者携带实现根据本申请的方法的程序代码的存储单元。

具体实施例

下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

实施例一

详细介绍本申请实施例提供的一种场景化商家召回方法。

参照图 1，示出了本申请实施例中一种场景化商家召回方法的步骤流程图。

步骤 110，获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数。

本申请实施例中所述的给定场景词用于指示当前商家召回场景。在实际应用中，如果直接基于给定场景词进行商家召回，召回率较低。因此在本申请实施例中，为了提高召回率，在获取得到给定场景词之后，可以将给定场景词扩展为包含 N 个词的关键词集合。其中，N 为自然数。一般而言，针对同一场景的给定场景词具体可以包括至少一个词语，而且 N 的具体取值可以根据需求进行预先设置，对此本申请实施例不加以限定。

而且，在本申请实施例中，可以通过任何可用的语义相似度方法将定场景词扩展为包含 N 个词的关键词集合。例如，可以通过词袋模型，将定场景词扩展为包含N个词的关键词集合。其中，在信息检索中，词袋(Bag of words，BOW）模型假定对于一个文本，忽略其词序和语法、句法，将其仅仅看做是一个词集合，或者说是词的一个组合，文本中每个词的出现都是独立的，不依赖于其他词是否出现，或者说当这篇文章的作者在任意一个位置选择一个词汇都不受前面句子的影响而独立选择的。

而且，在本申请实施例中，给定场景词的关键词集合中也可以包含相应的给定场景词在内，当然也可以不包含相应的给定场景词，具体可以根据需求进行预先设置，对此本申请实施例不加以限定。

可选地，在本申请实施例中，所述步骤 110 进一步可以包括:

子步骤 111，获取给定场景词，并通过 word2vec 获取与所述给定场景词的相关度最高的 N 个词构建所述给定场景词的关键词集合。

Word2vec 是 Google 在 2013 年年中开源的一款将词表征为实数值向量的高效工具，其利用深度学习的思想，可以通过训练，把对文本内容的处理简化为 K 维向量空间中的向量运算，而向量空间上的相似度可以用来表示文本语义上的相似度。Word2vec 输出的词向量可以被用来做很多 NLP（Natural Language Processing 自然语言处理）相关的工作，比如聚类、找同义词、词性分析等等。如果换个思路， 把词当做特征，那么 Word2vec 就可以把特征映射到 K 维向量空间，可以为文本数据寻求更加深层次的特征表示 。

Word2vec 使用的是Distributed representation（分布式表示）的词向量表示方式。Distributed representation 最早由 Hinton 在 1986 年提出。其基本思想是 通过训练将每个词映射成 K 维实数向量（K 一般为模型中的超参数），通过词之间的距离（比如余弦相似度、欧氏距离等）来判断它们之间的语义相似度。其采用一个三层的神经网络，输入层-隐层-输出层。有个核心的技术是根据词频用 Huffman（哈夫曼）编码，使得所有词频相似的词隐藏层激活的内容基本一致，出现频率越高的词语，他们激活的隐藏层数目越少，这样有效的降低了计算的复杂度。而 Word2vec 大受欢迎的一个原因正是其高效性。这个三层神经网络本身是对语言模型进行建模，但也同时获得一种单词在向量空间上的表示，而这个副作用才是 Word2vec 的真正目标。Word2Vec 的基本思想是把自然语言中的每一个词，表示成一个统一意义统一维度的短向量。

在本申请实施例中，可以直接使用 word2vec 官方文档中计算给定任意词的 Top N 个词的公式直接计算得到给定场景词的相似度最高的前 N 个相似词，进而可以构建得到给定场景词的关键词集合。

步骤 120，基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家。

关键词集合中的每个词都是一个关键词，那么此时为了扩大召回商家的数量以及范围，可以分别基于关键词集合中的每个关键词，从平台商家中确

定初始种子商家。其中的平台商家可以包括进行商家召回所涉及的平台中所包含的全部商家。

具体的，可以基于关键词集合中的每个关键词与平台商家的匹配度，确定初始种子商家；或者，可以将关键词集合中的全部关键词作为一个整体，然后基于全部关键词与平台商家的匹配度，确定初始种子商家；等等。而且，与平台商家的匹配度可以包括与平台商家的各个属性的匹配度。其中的属性可以包括商家名、推荐菜、团购信息、评论，等等。在本申请实施例中，可以分别从平台商家中筛选出与每个关键词的匹配度大于等于第二预设匹配度的商家作为相应关键词对应的初始召回商家，或者是分别从平台商家中筛选出与每个关键词的匹配度最高的前S个商家作为相应关键词对应的初始召回商家，等等。而且如果分别获取得到各个关键词对应的初始召回商家，那么还需要将各个关键词对应的初始召回商家汇总，得到相应的关键词集合所对应的初始种子商家。其中，S 为大于 1 的整数，第二预设匹配度以及 S 的具体取值可以根据需求进行预先设置，对此本申请实施例不加以限定。

或者，在本申请实施例中，根据相关场景在语料中的稀疏度，在确定初始种子商家时可以直接使用关键词集合中的各个关键词在商家评论中直接命中，并通过命中计数来筛选初始种子商家。具体的，如果在平台商家的任意一条商家评论中包含关键词集合中的任意一个关键词，则确认相应的关键词命中相应的平台商家一次，进而可以将关键词集合中的各个关键词对同一平台商家的全部商家评论的命中次数进行加权求和，得到关键词集合对相应的平台商家的总命中次数。进而可以基于关键词集合对各个平台商家的总命中次数从高到低的顺序，选取前 R 个平台商家作为初始种子商家。其中，R 的具体取值可以根据需求进行预先设置，对此本申请实施例不加以限定。各个关键词对应的命中次数的权重可以为相应关键字与给定场景词的相关度权重。

步骤 130，根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家。

如前述步骤可知，为了提高召回率会将给定场景词扩展为一个关键词集合。但是并不能完全保证经关键词集合中的每个关键词得到的初始种子商家的精确度，而且初始种子商家到最终召回结果的算法具有极强的传播能力，如果一个初始种子商家的精确度不够，那么在后续步骤中其会被继续放大，从而影响商家召回的精确度，所以在本申请实施例中对初始种子商家的精度要求很高，但是对初始种子商家的召回率可以无任何要求。

因此，在本申请实施例中，为了能够在提高商家召回率的同时提高召回准确性，在获取得到初始召回商家之后，还需要根据相应的关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家。具体的，根据关键词集合，对所述初始种子商家进行过滤的过滤策略，可以根据需求进行预先设置，对此本申请实施例不加以限定。

例如，可以设置过滤策略为如果初始种子商家的商家名中包含相应的关键词集合中的至少一个关键词，则确认相应的初始种子商家为高精度种子商家，而如果初始种子商家的商家名中不包含相应的关键词集合中的任意一个关键词，则可以将相应的初始种子商家过滤掉，那么经过滤后保留的初始种子商家则为高精度种子商家。

步骤 140，根据所述高精度种子商家，确定所述给定场景词的最终召回商家。

如前述，确定的高精度种子商家的精度较高，但是数量有限，因此在确定了高精度种子商家之后，还需要进一步扩大召回范围，也即确定给定场景词的最终召回商家。

具体的，可以首先确定各个平台商家与相应的每个高精度种子商家的相似度。很明显，某一平台商家与高精度种子商家的相似度越高，则说明该平台商家与高精度种子商家对应同一场景的可能性就越大。因此，在本申请实施例中，可以基于各个平台商家与高精度种子商家之间的相似度，确定相应的给定场景词的最终召回商家。而且，具体的可以根据任何可用方法确认平台商家与高精度种子商家之间的相似度，对此本申请实施例不加以限定。

例如，可以通过比对平台商家与高精度种子商家各个属性之间的匹配程度，进而确定平台商家与高精度种子商家的相似度；或者是分别将平台商家与高精度种子商家的各个属性表征为向量形式，进而分别比较同一属性对应的各个向量之间的相似度，从而确定平台商家与高精度种子商家的相似度；等等。其中的属性可以包括商家名、评论、特色菜，等等中的至少一种，具体的上述属性可以根据需求进行预先设置，对此本申请实施例不加以限定。

而且，需要说明的是，其中基于给定场景词确定的高精度商家可以默认为相应的给定场景词对应的最终召回商家，当然也可以根据需求设置不将基于给定场景词确定的高精度商家默认为最终召回商家，对此本申请实施例不加以限定。

可选地，在本申请实施例中，在所述步骤 140 之后，还可以包括：

步骤 150，根据用户搜索记录确定每个场景的搜索词集合。

经上述的步骤生成的最终召回商家是基于给定场景词的召回商家，但是在实际应用中，针对不同场景不同用户的查询词并不完全相同，那么在确定了各个场景的最终召回商家之后，为了能够根据不同用户的个性化查询词返回相应的查询结果，还需要进一步根据场景具体覆盖的查询词以及业务要求，对每个具体的查询词覆盖的召回商家进行规则化过滤，产生与具体查询词以及业务高度相关的召回商家集合。

那么首先则需要根据用户搜索记录确定每个场景的搜索词集合。例如，根据用户搜索记录筛选出的用户针对不同的场景可能会搜索或查询的词，从而得到各个场景的搜索词集合。对于每个场景，都可以通过用户搜索记录给定一个属于相应的各个场景的搜索词集合，这个集合是闭集合，通常集合里都会包含相应场景的给定场景词，只要用户输入搜索词集合里的任意至少一个词就会触发对相应场景的搜索流程。

步骤 160，根据所述搜索词集合以及业务需求，确定所述搜索词集合的过滤策略。

在实际应用中，由于搜索目的、搜索发起者等不同，相应的业务需求也会有所不同。例如，可以分为对公搜索业务需求、个人搜索业务需求、高精度业务需求、高召回率业务需求，等等。

在本申请实施例中，其中的过滤策略、业务需求的具体内容，以及过滤策略与搜索词集合和业务需求的对应关系可以根据需求进行预先设置，对此本申请实施例不加以限定。当然，在本申请实施例中，也可以仅根据所述搜索词集合或者是业务需求，确定所述搜索词集合的过滤策略，对此本申请实施例不加以限定。

步骤 170，基于所述过滤策略对所述搜索词集合中的各个搜索词的最终召回商家进行过滤，得到与所述搜索词以及所述业务需求匹配的召回商家集合。

在确定了过滤策略之后，则可以基于所述过滤策略对所述搜索词集合中的各个搜索词的最终召回商家进行过滤，得到与所述搜索词以及所述业务需求匹配的召回商家集合。由于所述搜索词集合中包含相应场景的给定场景词，因此可以基于所述过滤策略对所述给定场景词的最终召回商家进行过滤，得到与所述给定场景词以及所述业务需求匹配的召回商家集合。例如，用户搜索“老北京火锅”时，需要分别首先将“火锅”改写成“涮肉”、“铜锅涮肉”、“涮羊肉”，然后同时使用“火锅”、“涮肉”、“涮羊肉”、“铜锅涮肉”从老北京的最终召回商家里面过滤出“火锅”的结果。

另外，在本申请实施例中，为了避免经由上述的步骤 110-140 得到的最终召回商家相对于现有的场景化商家召回方法有遗漏，还可以在步骤 140 得到最终召回商家的基础上进行补召回。具体的可以将传统基于关键词命中过滤等方法召回的商家与由上述步骤 110-140 召回的商家进行比较，提取出经传统方法额外召回的商家，并与由上述步骤 110-140 召回的商家进行合并生成最终召回结果。

根据本申请的场景化商家召回方法，可以获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数；基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家；根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家；根据所述高精度种子商家，确定所述给定场景词的最终召回商家。由此降低了商家召回的时间成本以及人力成本，同时兼顾了商家召回的精度和召回率。

实施例二

详细介绍本申请实施例提供的一种场景化商家召回方法。

参照图 2，示出了本申请实施例中一种场景化商家召回方法的步骤流程图。

步骤 210，通过预选的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵。

为了能够生成高精度且高召回的结果，首先需要从全方位去描述给定场景，需要将给定场景词扩展为带有相关度权重的关键词集合，所以我们选择了考虑全局上下文且发散性更强的 LDA（Latent Dirichlet Allocation）模型。LDA 模型是一种文档主题生成模型，也称为一个三层贝叶斯概率模型，包含词、主题和文档三层结构。那么，通过预选的高质量商家评论对 LDA 模型进行训练则可以生成相应的 word-topic（主题词）矩阵。

其中的高质量商家评论是预先根据需求选定的不同商家的商家评论。例如，对于美团点评而言，高质量商家评论可以包括在存储之前已经进行过多道工序进行过人为筛选或者自动筛选的美团点评的评论，所以高质量商家评论的质量是有保证的。

可选地，在本申请实施例中，所述步骤 210 进一步可以包括：

子步骤 211，对所述预选的高质量商家评论进行预处理。

在实际应用中，用户在进行评价时可能会输入无关紧要的内容，例如一些特殊符号、停用词等等，而这些内容都可能影响最终的训练结果。因此，在本申请实施例中，为了提高训练效果，可以先对预选的高质量商家评论进

行预处理。而且，其中的预处理可以包括但不限于去除特殊字符处理、分词处理中的至少一种。其中特殊字符可以包括但不限于特殊符号、低频词、停用词等等。

其中，特殊符号可以包括相对于传统或常用的符号外，使用频率较少字符且难以直接输入的符号，比如：数学符号，单位符号，制表符等。低频词是指出现次数较少的词，在本申请实施例中可以预先设置一出现频率阈值，如果评论中某一词的出现次数少于该出现频率阈值，则可以确认其为低频词。停用词可以包括在信息检索中，为节省存储空间和提高搜索效率，在处理自然语言数据（或文本）之前或之后会自动过滤掉某些字或词，这些字或词即被称为 Stop Words（停用词）。在本申请实施例中，可以基于现有的停用词表执行去除停用词处理，和/或可以预先设置不同业务的停用词。例如，对于美食业务商家名中“餐厅”就是停用词，对于团购字段文本中“满 XXX 减 XXX”就是停用词，比如写真业务中“写真”就是停用词，等等。

另外，在本申请实施例中，可以通过任何可用分词处理方法对高质量商家评论进行分词处理，对此本申请实施例不加以限定。

子步骤 212，利用预处理后的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵。

在对高质量商家评论进行预处理之后，则可以进一步，利用预处理后的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵，矩阵中包括标识主题的词。其中，在本申请实施例中，还可以预先设置 LDA 模型中的隐含话题数 K。其中的隐含话题数可以理解为训练后得到的 word-topic 矩阵中的 topic（主题）数量。在本申请实施例中，基于多轮实验结果，优选地可以设置 K 为 2500。

步骤 220，基于所述 LDA 模型将所述给定场景词表示为多个原始主题的混合成分。

利用训练好的 LDA 模型可以对任意一篇给定文档生成一个基于原始主题的混合成分。其中的原始主题也即上述的训练后的 LDA 模型得到的 topic。

此时可以将给定场景词输入训练后的 LDA 模型，进而得到给定场景词在各个原始主题下的概率，也即将给定场景词表示为多个原始主题的混合成分。

例如，假设给定场景词是“动漫”，那么基于 LDA 模型可以得到“动漫”在各个原始主题下的概率，也即“动漫”的主题分布如下：

1297: 0.89444275499683

1320: 0.044182015916109305
617: 0.013558739479591326
1024: 0.010889265702068057
1885: 0.009441767594398784
182: 0.005473920591330183
.....

其中，左侧一列代表原生主题编号，右侧一列代表给定场景词在相应原生主题下的概率。

步骤 230，基于所述混合成分所表征的条件概率，确定所述给定场景词与所述 word-topic 矩阵中各个词的相关度权重。

其中，上述给定场景词在原始主题下的概率可以表示为 $P(topic_k|d)$，其中 $topic_k$ 表示为第 k 个原始主题，d 表示给定场景词。而且如果给定场景词具体包含多个词，那么 d 则可以表示给定场景词构成的文档。

进而可以基于上述的 $P(topic_k|d)$ 得到给定场景词与所述 word-topic 矩阵中各个词的相关度权重为:

$$P(word|d)=\sum_{z=1}^{K} P(topic_k|d)\times P(word|topic_k)$$

其中，word 表示为 word-topic 矩阵中的任意一个词，K 表示原始主题的总数。$P(word|topic_k)$ 的取值可以在获取得到 word-topic 矩阵时同时获取得到。

步骤 240，从所述 word-topic 矩阵的各个词中选取与所述给定场景词的相关度权重最高的 N 个词构建所述给定场景词的关键词集合。

在计算得到给定场景词与所述 word-topic 矩阵中各个词的相关度权重之后，很明显相关度权重越大则说明相应的词与给定场景词越接近，越可能对应同一场景。因此，在本申请实施例中，可以进一步从 word-topic 矩阵的各个词中选取与给定场景词的相关度权重最高的N个词构建所述给定场景词的关键词集合。其中 N 的取值可以根据需求进行预先设置，对此本申请实施例不加以限定。

例如，对于前述的给定场景词“动漫”，根据其与 word-topic 矩阵中每个词的相关度权重，可以得到其关键词集合为{可爱,卡通,漫画,动画,海贼王,榻榻米,趣味,玩偶,桌游,女仆,日式,居酒屋,二次元,玩具,公仔,任务,贴纸,彩绘,机器猫,卡哇伊,奇幻}，此时 N 的取值为 21。

步骤 250，基于预设的重排算法重新确定所述 word-topic 矩阵中各个

词与所述给定场景词的相关度权重。

步骤 260，根据重新确定的相关度权重更新所述关键词集合。

在实际应用中，基于上述的相关度权重得到的给定场景词的关键词集合中可能会包含一些实际相关性不大的关键词，那么此时则可以基于预设的重排算法重新确定所述 word-topic 矩阵中各个词与所述给定场景词的相关度权重。其中的重排算法可以根据需求进行预先设置，对此本申请实施例不加以限定。

例如，可以基于给定场景词与 word-topic 矩阵中各个词的词义、出现场景等等多个属性重新确定给定场景词与 word-topic 矩阵中各个词的相关度权重，等等。

在重新确定了 word-topic 矩阵中各个词与所述给定场景词的相关度权重之后，则可以基于重新确定的相关度权重更新给定场景词对应的关键词集合。

而且，如果重排后得到的给定场景词的关键词集合仍达不到要求精度，则可以通过少量人工筛选出高精度的关键词集合。

步骤 270，确定各个平台商家的向量表示，以及所述关键词集合的向量表示。

为了进一步基于关键词集合中的各个关键词从平台商家中筛选出初始种子商家，则需要确定各个平台商家与关键词集合的相似度。但是基于文本形式的平台商家和关键词集合很难准确地确定相似度。因此在本申请实施例中，可以首先确定各个平台商家的向量表示，以及所述关键词集合的向量表示。在本申请实施例中，可以通过任何可用方法确定各个平台商家的向量表示，以及所述关键词集合的向量表示，对此本申请实施例不加以限定。

可选地，在本申请实施例中，所述步骤 270 进一步可以包括:

子步骤 271，将同一平台商家的评论聚合为一个评论文本，并将所述关键词集合中的全部关键词聚合为一个关键词文本。

在本申请实施例中，为了充分考虑同一商家平台的各个评论，以及关键词集合中的全部关键词，进而确定初始种子商家。在本申请实施例中，可以先将同一平台商家的评论聚合为一个评论文本，并将所述关键词集合中的全部关键词聚合为一个关键词文本。

其中，聚合为一个评论文本的属于同一平台商家的评论可以为经过预先筛选的较高质量的且属于相应平台商家的评论，当然也可以为获取到的相应平台商家的全部评论，具体的可以根据需求进行预先设置，对此本申请实施

例不加以限定。

子步骤 272，基于所述评论文本确定所述平台商家的向量表示，并基于所述关键词文本确定所述关键词集合的向量表示。

在获取得到评论文本以及关键词文本之后，则可以分别基于评论文本确定所述平台商家的向量表示，以及基于关键词文本确定所述平台商家的向量表示。

可选地，在本申请实施例中，所述子步骤 272 进一步可以包括：通过所述 LDA 模型，根据所述评论文本确定所述平台商家的向量表示，并通过所述 LDA 模型，根据所述关键词文本确定所述关键词集合的向量表示；或者，通过 word2vec 模型，根据所述评论文本确定所述平台商家的向量表示，并根据所述关键词文本确定所述关键词集合的向量表示。其中，通过 LDA 模型得到的向量表示也可以称为主题向量。

由于 LDA 模型只对长文本有好的表征效果，而经聚合后得到的评论文本以及关键词文本可以作为长文本，因此在本申请实施例中，可以使用前述训练后的 LDA 模型根据评论文本推断出相应平台商家的向量表示，接着将关键词文本作为长文本同样通过 LDA 模型推断出相应的关键词集合的向量表示。

或者，在本申请实施例中，根据场景在语料中的稀疏度，也可以通过 word2vec 模型，根据所述评论文本确定所述平台商家的向量表示，并通过 word2vec 模型，根据所述关键词文本确定所述关键词集合的向量表示。word2vec 也叫 word embeddings，中文名“词向量”，作用就是将自然语言中的字词转为计算机可以理解的稠密向量（Dense Vector）。

在本申请实施例中，如果关键词文本和/或评论文本为长文本，则可以优选地利用 LDA 模型确定向量表示，而如果关键词文本和/或评论文本为短文本，则可以优选地利用 word2vec 模型确定向量表示。当然，具体的可以根据需求进预先设置，对此本申请实施例不加以限定。

步骤 280，获取各个所述平台商家的向量表示与所述关键词集合的向量表示之间的相似度。

进一步的，则可以基于向量表示确定各个平台商家的向量表示与所述关键词集合的向量表示之间的相似度。其中的相似度可以包括但不限于余弦相似度、欧氏距离相似度、交叉熵相似度、相对熵距离相似度中的任意一种。

而且，根据相似度的具体种类，则可以采用相应的任何可用方法确定相应的相似度，对此本申请实施例不加以限定。

步骤 290，选取相似度最高的前 M 个平台商家作为所述给定场景词的初

始种子商家。

很明显，平台商家与关键词集合的相似度越高，则说明相应的平台商家与关键词集合对应同一场景的可能性越高，因此，在确定了各个平台商家与关键词集合的相似度之后，则可以选取相似度最高的前M个平台商家作为相应给定场景词的初始种子商家。其中M为大于1的整数，M的取值可以根据需求进行预先设置，对此本申请实施例不加以限定。

步骤 2110，根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家。

步骤 2120，根据所述高精度种子商家，进行多次迭代运算，并以每次迭代运算得到的精确召回商家的总和作为所述给定场景词的最终召回商家。

为了进一步扩大召回商家数量，以提高召回率，在本申请实施例中，在确定得到高精度种子商家之后，还可以基于高精度种子商家，进行多次迭代运算，并以每次迭代运算得到的精确召回商家的总和作为所述给定场景词的最终召回商家。同样地，最终召回商家中还可以包括初始获得的高精度种子商家，当然也可以不包括初始获得的高精度种子商家，具体的可以根据需求进行预先设置，对此本申请实施例不加以限定。

其中，每次迭代运算的过程包括:

步骤 S1，根据所述平台商家的各个属性以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度;

在实际应用中，每个平台商家又可以包括多个不同的属性，例如商家名、特色菜、评价等等属性，而且不同的属性在确定相似度时的重要程度并不完全一致，因此在本申请实施例中，可以预先设置平台商家的各个属性的预设权重，进而可以根据所述平台商家的各个属性以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度。其中，某一给定场景词对应的其他商家可以包括平台商家中除去相应的给定场景词对应的高精度种子商家之外的其他平台商家。

具体的，可以分别确定高精度种子商家与其他商家在不同属性下的相似度，进而基于各个属性的预设权重，对不同属性的相似度进行加权求和，从而得到高精度种子商家与相应的其他商家的最终相似度。

而且，在本申请实施例中，为了提高相似度的准确性，同样可以先对各个属性对应的文本数据进行预处理。此时的预处理可以包括用特定停用词对文本数据进行清洗，其中的特定停用词可以包括基于业务需求预先设定的停用词，例如美食业务商家名中“餐厅”就是特定停用词，比如团购字段文本

中“满 XXX 减 XXX”就是特定停用词，比如写真业务中“写真”就是特定停用词，等等。

步骤 S2，基于预设的相似度阈值，确定所述高精度种子商家对所述其他商家所投的票数；所述票数的权重为所述高精度种子商家与所述平台商家的最终相似度；

其中的相似度阈值可以根据需求进行预先设置，而且相似度预设的设置可以与相似度的种类有关，对此本申请实施例不加以限定。具体的，如果某一其他商家与一高精度种子商家的最终相似度超过预设的相似度阈值，则确定相应的高精度种子商家会对相应的其他商家投票，而且所投票数的权重为相应平台商家与相应的高精度种子商家的最终相似度；而如果某一其他商家与一高精度种子商家的最终相似度未超过预设的相似度阈值，则确定相应的高精度种子商家不会对相应的其他商家投票。

例如，假设其他商家 B、C 和 D 与高精度种子商家 A1 之间的最终相似度分别为 0.6、0.8、0.3，其他商家 B、C 和 D 与高精度种子商家 A2 之间的最终相似度分别为 0.7、0.4、0.3，且相似度数值越大表示越相似。如果预设的相似度阈为 0.5，那么可以确定高精度种子商家 A1 对其他商家 B 所投票数为 0.6，高精度种子商家 A1 对其他商家 C 所投票数为 0.8，高精度种子商家 A1 不会对其他商家 D 投票，高精度种子商家 A2 对其他商家 B 所投票数为 0.7，高精度种子商家 A2 不会对其他商家 C、D 投票。那么可以得到此时的其他商家 B 得到的票数为 1.3，其他商家 C 得到的票数为 0.8，其他商家 D 得到的票数为 0。

步骤 S3，选取票数最高的前 L 个其他商家作为所述初始召回商家，L 为自然数；

很明显，得到票数越高则说明相应的其他商家与高精度种子商家的相似度越高，因此在本申请实施例中可以选取票数最高的前 L 个其他商家作为相应的给定场景词的初始召回商家。其中 L 的取值可以根据需求进行预先设置，对此本申请实施例不加以限定。

如前述的其他商家 B、C 和 D 与高精度种子商家 A1、A2，如果此时设置 L 的取值为 2，那么则可以选取得到初始召回商家依次为其他商家 B、C。

步骤 S4，根据所述关键词集合，对所述初始召回商家进行过滤，得到精确召回商家；

为了提高召回精确率，同样地可以根据关键词集合，对所述初始召回商家进行过滤，得到精确召回商家。具体的过滤过程可以与前述的对初始种子

商家的过滤过程类似，在此不加以赘述。

步骤 S5，从所述精确召回商家中选取票数最高的第一数量个商家作为新的高精度种子商家，并更新本次迭代的所述高精度种子商家，直至所述精确召回商家的召回精度下降或者所述精确召回商家的数量收敛。

为了进一步扩大召回范围，则可以从经本次迭代得到的精确召回商家中选取第一数量个商家作为新的高精度种子商家，并通过所述新的高精度种子商家更新本次迭代的所述高精度种子商家，然后以更新后的高精度种子商家进入下一次迭代过程，直至得到的精确召回商家的召回精度下降或者是精确召回商家的数量收敛。其中第一数量的具体取值可以根据需求进行预先设置，对此本申请实施例不加以限定。而在更新高精度种子商家之后，则可以基于更新后的高精度种子商家进入下一次迭代过程，也即依次执行上述的步骤 S1-S5，以更新后的高精度种子商家，确定每个高精度种子商家与所述平台商家中每个其他商家的最终相似度，进而基于预设的相似度阈值，确定更新后的高精度种子商家对相应的其他商家所投的票数，其中票数的权重为相应的高精度种子商家与平台商家的最终相似度，重新得到初始召回商家并且经过滤得到新的精确召回商家，然后再次更新高精度种子商家，直至当前次迭代的精确召回商家的召回精度下降或者所述精确召回商家的数量收敛。

其中，更新本次迭代的所述高精度种子商家的过程可以为将最新选取的第一数量个高精度种子商家添加到本次迭代过程中所用的高精度种子商家中，或者是以最新选取的第一数量个高精度种子商家替换本次迭代过程中所用的高精度种子商家，等等。具体的可以根据需求进行预先设置，对此本申请实施例不加以限定。

而且，经步骤 S4 之后，还可以进一步判断所述精确召回商家的召回精度是否下降，或者是所述精确召回商家的数量是否收敛，如果所述精确召回商家的召回精度下降，或者是所述精确召回商家的数量收敛，则可以停止迭代，也无需执行步骤 S5；而如果精确召回商家的召回精度未下降，且所述精确召回商家的数量未收敛，则可以执行 S5 并继续进行迭代。

可选地，在本申请实施例中，所述属性包括商家名、推荐菜、团购信息、评论中的至少一种。

可选地，在本申请实施例中，所述步骤 S2 进一步可以包括：

子步骤 S21，将所述平台商家的各个属性表征为向量，得到所述平台商家的属性向量；

在本申请实施例中，为了计算高精度种子商家与每个其他商家的最终相

似度，而且是基于平台商家的各个属性确定其中的高精度种子商家与所述平台商家中每个其他商家的最终相似度，那么同样地则需要将平台商家的各个属性表征为向量，得到所述平台商家的属性向量。而且，在本申请实施例中，可以通过任何可用方法将属性表征为向量，对此本申请实施例不加以限定。例如，通过前述的 LDA 模型，word2vec 模型，或者是 doc2vec 模型，等等。

子步骤 S22，基于所述属性向量以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度。

在得到各个属性的属性向量，则可以进一步基于属性向量以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度。其中各个属性的预设权重可以根据需求进行预先设置，对此本申请实施例不加以限定。

可选地，在本申请实施例中，所述子步骤 S21 进一步可以包括:

子步骤 S211，基于各个属性的特定停用词对所述属性对应的文本数据进行数据清洗;

在本申请实施例中，可以针对不同的属性设置不同的特定停用词。例如，将美食业务商家名中“餐厅”设置为特定停用词，将团购字段文本中“满 XXX 减 XXX”设置为特定停用词，将写真业务中“写真”设置为特定停用词，等等。那么则可以基于各个属性的特定停用词对所述属性对应的文本数据进行数据清洗。

子步骤 S212，通过 word2vec 将数据清洗后的所述平台商家的商家名表征为向量;

如前述，在本申请实施例中的属性可以包括商家名，而且由于商家名一般字数不会过多，那么此时则可以适用于通过 word2vec 将数据清洗后的所述平台商家的商家名表征为向量，进而得到商家名对应的属性向量。

子步骤 S213，通过 doc2vec 将数据清洗后的所述平台商家的参考信息分别表征为向量; 或者，通过 word2vec 将数据清洗后的所述平台商家的参考信息中的每个分词分别表征为向量，并获取每个参考信息中的每个分词的向量平均值作为所述参考信息的向量。

而对于推荐菜、团购信息、评论等参考信息而言，其篇幅一般会大于商家名，而且表述上更倾向于文档形式，因此在本申请实施例中，可以通过 doc2vec 将数据清洗后的所述平台商家的参考信息分别表征为向量，进而得到各个参考信息对应的属性向量。

或者，为了方便利用 word2vec 将数据清洗后的文本数据表征为向量，

还可以对平台商家的参考信息进行分词处理，进而将其处理为多个分词，那么则可以通过 word2vec 将数据清洗后的所述平台商家的参考信息中的每个分词分别表征为向量，并获取每个参考信息中的每个分词的向量平均值作为所述参考信息的向量。

例如，对于推荐菜而言，则可以对数据清洗后的推荐菜进行分词处理，进而通过 word2vec 将推荐菜的每个分词表征为向量，并获取推荐菜的全部分词的向量平均值，进而得到推荐菜所对应的文本数据的属性向量。而对于其他的参考信息也类似。

可选地，在本申请实施例中，所述子步骤 S22 进一步可以包括:

子步骤 S221，基于所述属性向量，分别获取所述高精度种子商家与所述其他商家在每个属性下的相似度;

在本申请实施例中，可以基于平台商家的属性向量，分别获取平台商家中的与当前的给定场景词对应的高精度种子商家与其他商家在每个属性下的相似度。例如商家名相似度，特色菜相似度、团购信息相似度，等等。

子步骤 S222，按照所述属性的预设权重，将各个属性的相似度加权求和，得到所述高精度种子商家与所述其他商家的最终相似度。

在获取得到各个属性下的相似度之后，则可以基于各个属性的预设权重，将各个属性的相似度进行加权求和，进而得到相应的高精度种子商家与相应的其他商家的最终相似度。

可选地，在本申请实施例中，所述相似度可以包括但不限于余弦相似度、欧氏距离相似度、交叉熵相似度、相对熵距离相似度中的任意一种。

可选地，在本申请实施例中，在所述步骤 S5 之前，还可以包括:

步骤 S6，根据本次迭代运算得到的精确召回商家的数量，以及本次迭代运算的迭代次序，确定所述第一数量的取值。

在实际应用中，对于每一次迭代对应的第一数量可以相同，也可以不完全相同，那么如果每次迭代所对应的第一数量不完全相同，则可以根据迭代过程中设置第一数量的取值，从而进一步提高得到的最终召回商家的准确率以及召回率。因此，在本申请实施例中，可以根据本次迭代运算得到的精确召回商家的数量，以及本次迭代运算的迭代次序，确定所述第一数量的取值。具体的本次迭代运算得到的精确召回商家的数量以及本次迭代运算的迭代次序与本次迭代对应的第一数量之间的对应关系可以根据需求进行预先设置，对此本申请实施例不加以限定。

例如，可以设置第一数量的取值为 T* step * n_iteration，step 为经

本次迭代运算确定的精确召回商家的数量，n_iteration 为本次迭代运算的迭代次序，T 为预设参数。其中 T 的取值可以根据需求进行设置，对此本申请实施例不加以限定，例如可以直接设置 T 为 1，等等。

第一方面，根据本申请的场景化商家召回方法，可以获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合；基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家；根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家；根据所述高精度种子商家，确定所述给定场景词的最终召回商家。由此降低了商家召回的时间成本以及人力成本，同时兼顾了商家召回的精度和召回率。

第二方面，根据本申请的场景化商家召回方法，可以通过预选的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵；基于所述 LDA 模型将所述给定场景词表示为多个原始主题的混合成分；基于所述混合成分所表征的条件概率，确定所述给定场景词与所述 word-topic 矩阵中各个词的相关度权重；从所述 word-topic 矩阵的各个词中选取与所述给定场景词的相关度权重最高的 N 个词构建所述给定场景词的关键词集合。并且，基于预设的重排算法重新确定所述 word-topic 矩阵中各个词与所述给定场景词的相关度权重；根据重新确定的相关度权重更新所述关键词集合。以及，对所述预选的高质量商家评论进行预处理；利用预处理后的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵；其中，所述预处理包括去除特殊字符处理、分词处理；所述特殊字符包括特殊符号、低频词、停用词中的至少一种。从而可以进一步提高商家召回的准确率以及召回率。

第三方面，根据本申请的场景化商家召回方法，可以确定各个平台商家的向量表示，以及所述关键词集合的向量表示；获取各个所述平台商家的向量表示与所述关键词集合的向量表示之间的相似度；选取相似度最高的前 M 个平台商家作为所述给定场景词的初始种子商家。将同一平台商家的评论聚合为一个评论文本，并将所述关键词集合中的全部关键词聚合为一个关键词文本；基于所述评论文本确定所述平台商家的向量表示，并基于所述关键词文本确定所述关键词集合的向量表示。同样可以进一步提高商家召回的准确率以及召回率。

第四方面，根据本申请的场景化商家召回方法，可以根据所述高精度种子商家，进行多次迭代运算，并以每次迭代运算得到的精确召回商家的总和作为所述给定场景词的最终召回商家；其中，每次迭代运算的过程包括：根据所述平台商家的各个属性以及所述属性的预设权重，确定每个所述高精度

种子商家与所述平台商家中每个其他商家的最终相似度；基于预设的相似度阈值，确定所述高精度种子商家对所述其他商家所投的票数；所述票数的权重为所述高精度种子商家与所述平台商家的最终相似度；选取票数最高的前L个其他商家作为所述初始召回商家；根据所述关键词集合，对所述初始召回商家进行过滤，得到精确召回商家；从所述精确召回商家中选取票数最高的第一数量个商家作为新的高精度种子商家，并更新本次迭代的所述高精度种子商家，直至所述精确召回商家的召回精度下降或者所述精确召回商家的数量收敛。并且，将所述平台商家的各个属性表征为向量，得到所述平台商家的属性向量；基于所述属性向量以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度。基于各个属性的特定停用词对所述属性对应的文本数据进行数据清洗；通过 word2vec 将数据清洗后的所述平台商家的商家名表征为向量；通过 doc2vec 将数据清洗后的所述平台商家的参考信息分别表征为向量；或者，通过 word2vec 将数据清洗后的所述平台商家的参考信息中的每个分词分别表征为向量，并获取每个参考信息中的每个分词的向量平均值作为所述参考信息的向量；其中，所述参考信息包括推荐菜、团购信息、评论中的至少一种。基于所述属性向量，分别获取所述高精度种子商家与所述其他商家在每个属性下的相似度；按照所述属性的预设权重，将各个属性的相似度加权求和，得到所述高精度种子商家与所述其他商家的最终相似度。根据本次迭代运算得到的精确召回商家的数量，以及本次迭代运算的迭代次序，确定所述第一数量的取值。同样可以进一步降低商家召回的时间成本以及人力成本，同时提高商家召回的精度和召回率。

对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本申请实施例所必须的。

实施例三

详细介绍本申请实施例提供的一种场景化商家召回装置。

参照图 3，示出了本申请实施例中一种场景化商家召回装置的结构示意图。

关键词集合扩展模块 310，用于获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数；

初始种子商家确定模块320，用于基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家；

高精度种子商家确定模块330，用于根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家；

最终召回商家确认模块340，用于根据所述高精度种子商家，确定所述给定场景词的最终召回商家。

第一方面，根据本申请的场景化商家召回装置，可以获取给定场景词，并将所述给定场景词扩展为包含N个词的关键词集合；基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家；根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家；根据所述高精度种子商家，确定所述给定场景词的最终召回商家。由此降低了商家召回的时间成本以及人力成本，同时兼顾了商家召回的精度和召回率。

本发明本申请实施例中还公开了一种电子设备，包括：

处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的获取机程序，所述处理器执行所述程序时实现前述的场景化商家召回方法。如图4所示，电子设备400包括：存储器410和处理器420，存储器410与处理器420之间通过总线通信连接，存储器410中存储有计算机程序，该计算机程序可在处理器420上运行，进而实现本申请上述任一实施例所述的方法中的步骤。

本申请实施例中还公开了一种可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行前述的场景化商家召回方法。

本申请实施例公开的场景化商家召回装置能够实现前述场景化商家召回方法，对于装置实施例而言，由于其与方法实施例基本相似，装置的各个模块的具体实施方式与相应的方法步骤对应，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。

在此提供的算法和显示不与任何特定获取机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述，构造这类系统所要求的结构是显而易见的。此外，本申请也不针对任何特定编程语言。应当明白，可以利用各种编程语言实现在此描述的本申请的内容，并且上面对特定语言所做的描述是为了披露本申请的最佳实施方式。

在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本申请的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未

详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

类似地，应当理解，为了精简本申请并帮助理解各个发明方面中的一个或多个，在上面对本申请的示例性实施例的描述中，本申请的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释成反映如下意图：即所要求保护的本申请要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，申请方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本申请的单独实施例。

本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它们分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本申请的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

本申请的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本申请实施例的场景化商家召回设备中的一些或者全部部件的一些或者全部功能。本申请还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序（例如，获取机程序和获取机程序产品）。这样的实现本申请的程序可以存储在获取机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

例如，图 5 示出了可以实现根据本申请的方法的电子设备，该电子设备

传统上包括处理器520和以存储器510形式的计算机程序产品或者计算机可读存储介质。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。所述电子设备为计算处理设备。所述存储器510可以是诸如闪存、EEPROM（电可擦除可编程只读存储器）、EPROM、硬盘或者ROM之类的电子存储器。存储器510具有用于执行上述方法中的任何方法步骤的程序代码的存储空间5101。例如，用于程序代码的存储空间5101可以包括分别用于实现上面的方法中的各种步骤的程序代码5102。所述计算机程序产品或者计算机可读存储介质，其上存储有计算机程序的程序代码，该程序代码被处理器520执行时实现如本申请实施例一和实施例二所述的数据读取方法。这些程序代码可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。这些计算机程序产品包括诸如硬盘，紧致盘（CD）、存储卡或者软盘之类的程序代码载体。这样的计算机程序产品或者计算机可读存储介质通常为如参考图6所述的便携式或者固定存储单元。该存储单元用于保持或者携带实现根据本申请的方法的程序代码，该存储单元可以具有与图5的电子设备中的存储器510类似布置的存储段、存储空间等。程序代码可以例如以适当形式进行压缩。通常，存储单元包括计算机可读代码6102'，即可以由例如诸如520之类的处理器读取的代码，这些代码当电子设备运行时，导致该电子设备执行上面所描述的方法中的各个步骤。

本文中所称的“一个实施例”、“实施例”或者“一个或者多个实施例”意味着，结合实施例描述的特定特征、结构或者特性包括在本公开的至少一个实施例中。此外，请注意，这里“在一个实施例中”的词语例子不一定全指同一个实施例。

在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本申请的实施例可以在没有这些具体细节的情况下被实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本申请可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

应该注意的是上述实施例对本申请进行说明而不是对本申请进行限制，

并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本申请可以借助于包括有若干不同元件的硬件以及借助于适当编程的获取机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

权 利 要 求

1.一种场景化商家召回方法，包括:

获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数;

基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家;

根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家;

根据所述高精度种子商家，确定所述给定场景词的最终召回商家。

2.根据权利要求 1 所述的方法，所述获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合的步骤，包括:

通过预选的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵;

基于所述 LDA 模型将所述给定场景词表示为多个原始主题的混合成分;

基于所述混合成分所表征的条件概率，确定所述给定场景词与所述 word-topic 矩阵中各个词的相关度权重;

从所述 word-topic 矩阵的各个词中选取与所述给定场景词的相关度权重最高的 N 个词构建所述给定场景词的关键词集合。

3.根据权利要求 2 所述的方法，在所述基于所述条件概率以及预设的概率阈值，确定所述给定场景词的关键词集合的步骤之后，还包括:

基于预设的重排算法重新确定所述 word-topic 矩阵中各个词与所述给定场景词的相关度权重;

根据重新确定的相关度权重更新所述关键词集合。

4.根据权利要求 2 所述的方法，所述通过预选的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵的步骤，包括:

对所述预选的高质量商家评论进行预处理;

利用预处理后的高质量商家评论训练 LDA 模型并生成 word-topic 矩阵;

其中，所述预处理包括去除特殊字符处理、分词处理; 所述特殊字符包括特殊符号、低频词、停用词中的至少一种。

5.根据权利要求 1 所述的方法，所述基于所述关键词集合中的各个关键词，确定初始种子商家的步骤，包括:

确定各个平台商家的向量表示，以及所述关键词集合的向量表示；

获取各个所述平台商家的向量表示与所述关键词集合的向量表示之间的相似度；

选取相似度最高的前 M 个平台商家作为所述给定场景词的初始种子商家，M 为自然数。

6.根据权利要求 5 所述的方法，所述确定各个商家的向量表示，以及所述关键词集合的向量表示的步骤，包括：

将同一平台商家的评论聚合为一个评论文本，并将所述关键词集合中的全部关键词聚合为一个关键词文本；

基于所述评论文本确定所述平台商家的向量表示，并基于所述关键词文本确定所述关键词集合的向量表示。

7.根据权利要求 6 所述的方法，所述基于所述评论文本确定所述平台商家的向量表示，并基于所述关键词文本确定所述关键词集合的向量表示的步骤，包括：

通过 LDA 模型，根据所述评论文本确定所述平台商家的向量表示和根据所述关键词文本确定所述关键词集合的向量表示；

或者，通过 word2vec 模型，根据所述评论文本确定所述平台商家的向量表示和根据所述关键词文本确定所述关键词集合的向量表示。

8.根据权利要求 1 所述的方法，所述根据所述高精度种子商家，确定所述给定场景词的最终召回商家的步骤，包括：

根据所述高精度种子商家，进行多次迭代运算，并以每次迭代运算得到的精确召回商家的总和作为所述给定场景词的最终召回商家；

其中，每次迭代运算的过程包括：

根据所述平台商家的各个属性以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度；

基于预设的相似度阈值，确定所述高精度种子商家对所述其他商家所投的票数；所述票数的权重为所述高精度种子商家与所述平台商家的最终相似度；

选取票数最高的前 L 个其他商家作为所述初始召回商家，L 为自然数；

根据所述关键词集合，对所述初始召回商家进行过滤，得到精确召回商家；

从所述精确召回商家中选取票数最高的第一数量个商家作为新的高精

度种子商家，并更新本次迭代的所述高精度种子商家，直至所述精确召回商家的召回精度下降或者所述精确召回商家的数量收敛。

9.根据权利要求 8 所述的方法，所述根据所述平台商家的各个属性以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度的步骤，包括:

将所述平台商家的各个属性表征为向量，得到所述平台商家的属性向量;

基于所述属性向量以及所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度。

10.根据权利要求 9 所述的方法，所述将所述平台商家的各个属性表征为向量，得到所述平台商家的属性向量的步骤，包括:

基于各个属性的特定停用词对所述属性对应的文本数据进行数据清洗;

通过 word2vec 将数据清洗后的所述平台商家的商家名表征为向量;

通过 doc2vec 将数据清洗后的所述平台商家的参考信息分别表征为向量，或者，通过 word2vec 将数据清洗后的所述平台商家的参考信息中的每个分词分别表征为向量，并获取每个参考信息中的每个分词的向量平均值作为所述参考信息的向量。

11.根据权利要求 9 所述的方法，所述基于所述属性向量以及预设的所述属性的预设权重，确定每个所述高精度种子商家与所述平台商家中每个其他商家的最终相似度的步骤，包括:

基于所述属性向量，分别获取所述高精度种子商家与所述其他商家在每个属性下的相似度;

按照所述属性的预设权重，将各个属性的相似度加权求和，得到所述高精度种子商家与所述其他商家的最终相似度。

12.根据权利要求 8 所述的方法，在所述从所述精确召回商家中选取票数最高的第一数量个商家作为新的高精度种子商家，并更新本次迭代的所述高精度种子商家的步骤之前，还包括:

根据本次迭代运算得到的精确召回商家的数量，以及本次迭代运算的迭代次序，确定所述第一数量的取值。

13.根据权利要求 1 所述的方法，其特征在于，在所述根据所述高精度种子商家，确定所述给定场景词的最终召回商家的步骤之后，还包括:

根据用户搜索记录确定每个场景的搜索词集合;

根据所述搜索词集合以及业务需求，确定所述搜索词集合的过滤策略；

基于所述过滤策略对所述搜索词集合中的各个搜索词的最终召回商家进行过滤，得到与所述搜索词以及所述业务需求匹配的召回商家集合。

14.根据权利要求 1 所述的方法，所述获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合的步骤，包括：

获取给定场景词，并通过 word2vec 获取与所述给定场景词的相关度最高的 N 个词构建所述给定场景词的关键词集合。

15.一种场景化商家召回装置，包括：

关键词集合扩展模块，用于获取给定场景词，并将所述给定场景词扩展为包含 N 个词的关键词集合，N 为自然数；

初始种子商家确定模块，用于基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家；

高精度种子商家确定模块，用于根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家；

最终召回商家确认模块，用于根据所述高精度种子商家，确定所述给定场景词的最终召回商家。

16.一种电子设备，包括：

处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的获取机程序，所述处理器执行所述获取机程序时实现如权利要求 1-14 中的任一项所述的场景化商家召回方法。

17.一种可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如权利要求 1-14 中的任一项所述的场景化商家召回方法。

18. 一种计算机程序，包括计算机可读代码，当所述计算机可读代码在电子设备上运行时，导致所述电子设备执行根据权利要求 1-14 中的任一项所述的场景化商家召回方法。

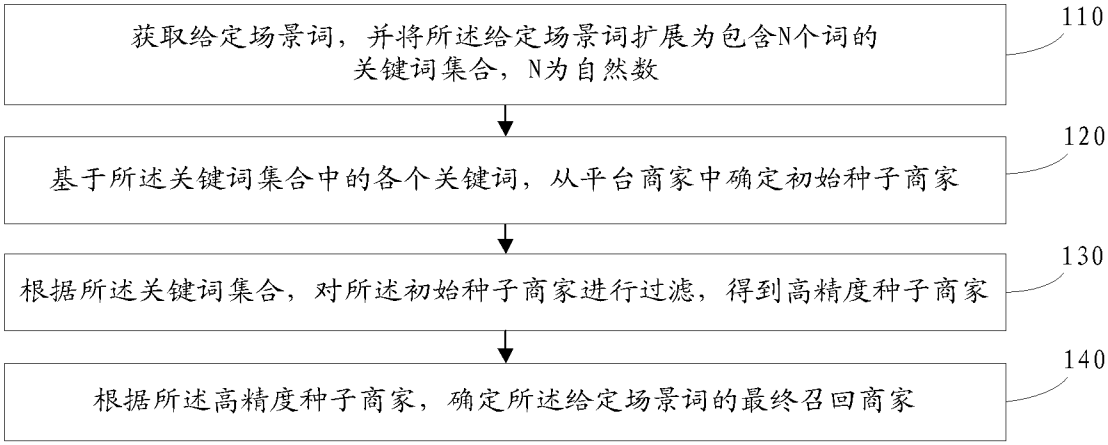

获取给定场景词，并将所述给定场景词扩展为包含N个词的关键词集合，N为自然数
110
基于所述关键词集合中的各个关键词，从平台商家中确定初始种子商家
120
根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家
130
根据所述高精度种子商家，确定所述给定场景词的最终召回商家
140

图 1

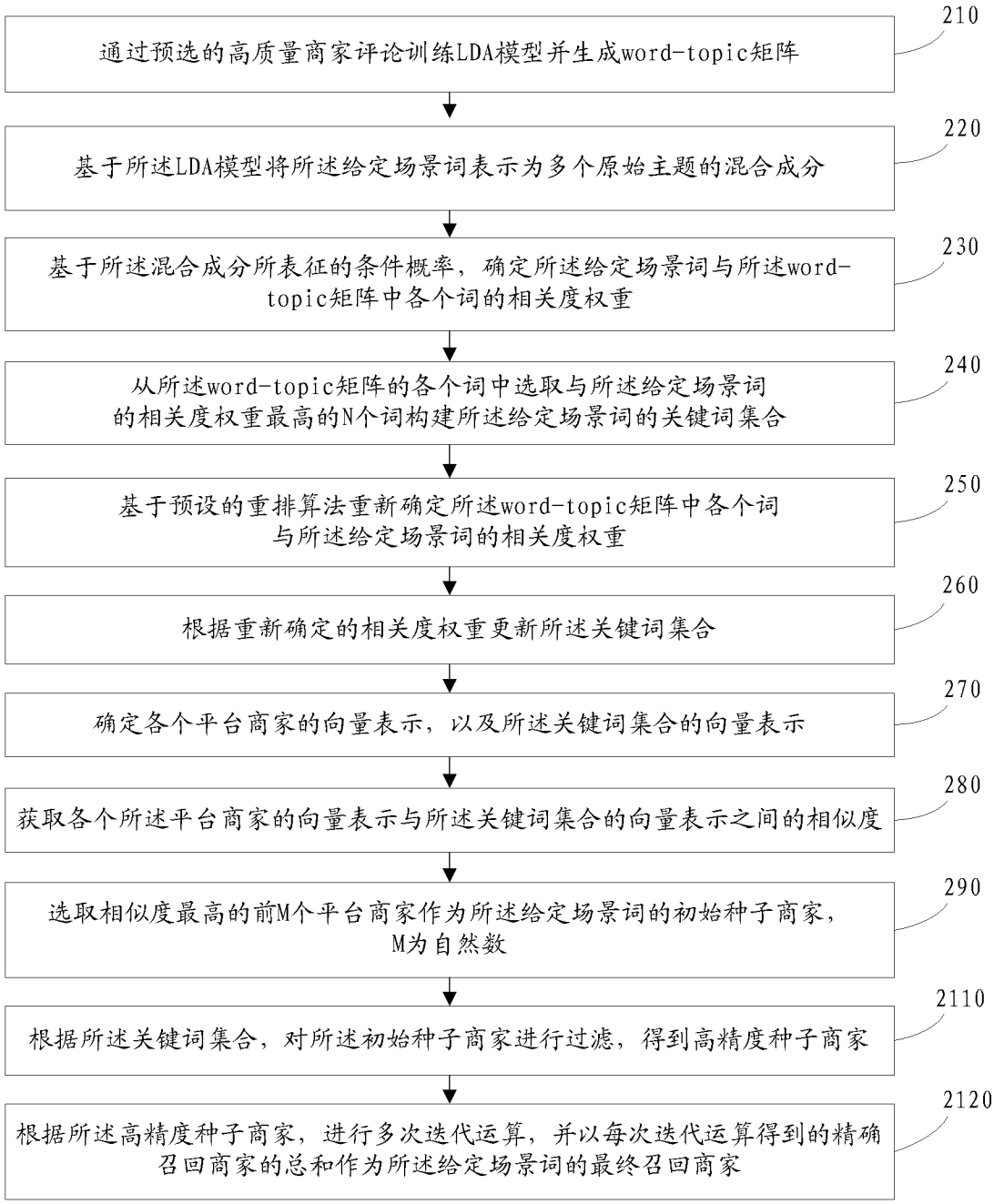

通过预选的高质量商家评论训练LDA模型并生成word-topic矩阵
210
基于所述LDA模型将所述给定场景词表示为多个原始主题的混合成分
220
基于所述混合成分所表征的条件概率，确定所述给定场景词与所述word-topic矩阵中各个词的相关度权重
230
从所述word-topic矩阵的各个词中选取与所述给定场景词的相关度权重最高的N个词构建所述给定场景词的关键词集合
240
基于预设的重排算法重新确定所述word-topic矩阵中各个词与所述给定场景词的相关度权重
250
根据重新确定的相关度权重更新所述关键词集合
260
确定各个平台商家的向量表示，以及所述关键词集合的向量表示
270
获取各个所述平台商家的向量表示与所述关键词集合的向量表示之间的相似度
280
选取相似度最高的前M个平台商家作为所述给定场景词的初始种子商家，M为自然数
290
根据所述关键词集合，对所述初始种子商家进行过滤，得到高精度种子商家
2110
根据所述高精度种子商家，进行多次迭代运算，并以每次迭代运算得到的精确召回商家的总和作为所述给定场景词的最终召回商家
2120

图 2

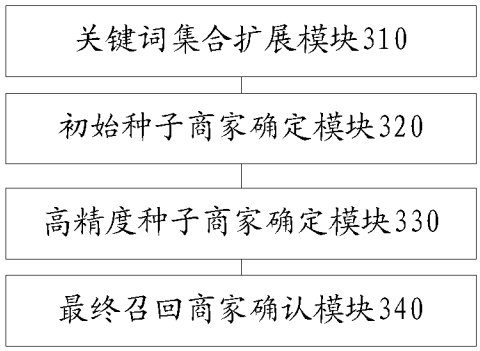

关键词集合扩展模块310
初始种子商家确定模块320
高精度种子商家确定模块330
最终召回商家确认模块340

图 3

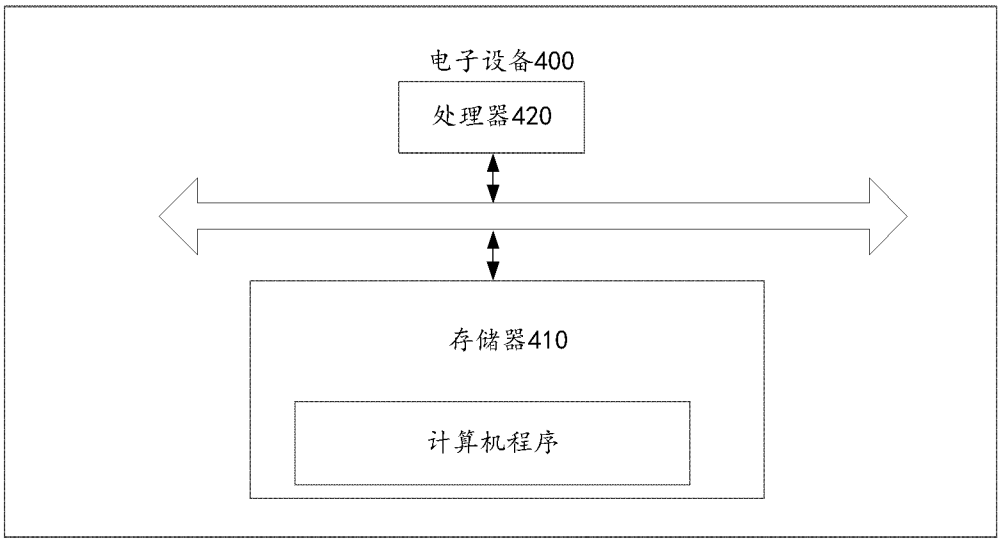

电子设备400
处理器420
存储器410
计算机程序

图 4

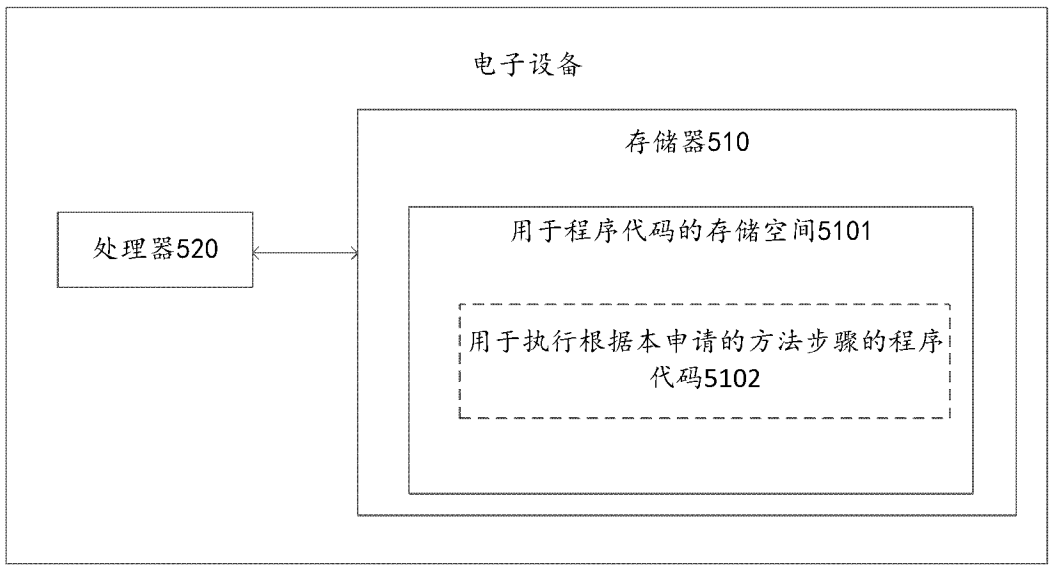

电子设备
存储器510
处理器520
用于程序代码的存储空间5101
用于执行根据本申请的方法步骤的程序代码5102

图 5

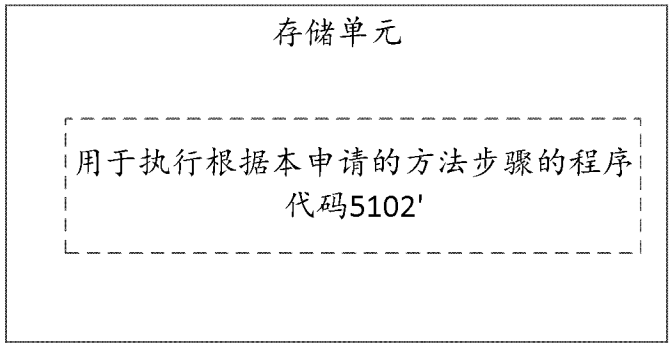


图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/098210

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/9535(2019.01)i; G06Q 30/00(2012.01)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; G06Q

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)

CNKI, CNPAT, WPI, EPODOC, IEEE: 场景, 关键词, 扩展, 商家, 商户, 检索, 搜索, 召回, 过滤, 筛选, 权重, scene, merchant, search, recall, keyword, expand, filtering, weight

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109815392 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 28 May 2019 (2019-05-28) entire document	1-18
X	CN 103577432 A (ALIBABA GROUP HOLDING LIMITED) 12 February 2014 (2014-02-12) description, paragraphs [0070]-[0223] and [0251]-[0282]	1-4, 8, 12-18
Y	CN 103577432 A (ALIBABA GROUP HOLDING LIMITED) 12 February 2014 (2014-02-12) description, paragraphs [0070]-[0223] and [0251]-[0282]	5-7, 9-11
Y	CN 108241699 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 03 July 2018 (2018-07-03) description, paragraphs [0005]-[0008]	5-7, 9-11
A	CN 107862027 A (BEIJING XIAODU INFORMATION TECHNOLOGY CO., LTD.) 30 March 2018 (2018-03-30) entire document	1-18

☑ 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	Date of mailing of the international search report
18 September 2019	**29 October 2019**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105069086 A (FOCUS TECHNOLOGY CO., LTD. et al.) 18 November 2015 (2015-11-18) entire document	1-18
A	CN 102629257 A (NANJING UNIVERSITY et al.) 08 August 2012 (2012-08-08) 1-18	entire document
A	CN 104063523 A (FOCUS TECHNOLOGY CO., LTD.) 24 September 2014 (2014-09-24) entire document	1-18
A	US 7831582 B1 (AMAZON TECHNOLOGIES, INC.) 09 November 2010 (2010-11-09) entire document	1-18

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/098210

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109815392	A	28 May 2019	None			
CN	103577432	A	12 February 2014	CN	103577432	B	14 July 2017
CN	108241699	A	03 July 2018	None			
CN	107862027	A	30 March 2018	CN	107862027	B	12 March 2019
				CN	109815308	A	28 May 2019
				WO	2019085236	A1	09 May 2019
CN	105069086	A	18 November 2015	CN	105069086	B	11 July 2017
CN	102629257	A	08 August 2012	CN	102629257	B	19 February 2014
CN	104063523	A	24 September 2014	CN	104063523	B	22 February 2017
US	7831582	B1	09 November 2010	US	7860895	B1	28 December 2010
				US	7774335	B1	10 August 2010

国际检索报告

国际申请号 PCT/CN2019/098210

A. 主题的分类

G06F 16/9535(2019.01)i；G06Q 30/00(2012.01)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；G06Q

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）)

CNKI,CNPAT,WPI,EPODOC,IEEE：场景，关键词，扩展，商家，商户，检索，搜索，召回，过滤，筛选，权重，scene, merchant, search, recall, keyword, expand, filtering, weight

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 109815392 A (北京三快在线科技有限公司) 2019年 5月 28日 (2019-05-28) 全文	1-18
X	CN 103577432 A (阿里巴巴集团控股有限公司) 2014年 2月 12日 (2014-02-12) 说明书第0070-0223，0251-0282段	1-4，8，12-18
Y	CN 103577432 A (阿里巴巴集团控股有限公司) 2014年 2月 12日 (2014-02-12) 说明书第0070-0223，0251-0282段	5-7，9-11
Y	CN 108241699 A (百度在线网络技术北京有限公司) 2018年 7月 3日 (2018-07-03) 说明书第0005-0008段	5-7，9-11
A	CN 107862027 A (北京小度信息科技有限公司) 2018年 3月 30日 (2018-03-30) 全文	1-18
A	CN 105069086 A (焦点科技股份有限公司 等) 2015年 11月 18日 (2015-11-18) 全文	1-18
A	CN 102629257 A (南京大学 等) 2012年 8月 8日 (2012-08-08) 1-18	全文
A	CN 104063523 A (焦点科技股份有限公司) 2014年 9月 24日 (2014-09-24) 全文	1-18

☑ 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	国际检索报告邮寄日期
2019年 9月 18日	2019年 10月 29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 张璇 电话号码 53961809

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告

国际申请号
PCT/CN2019/098210

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
A	US 7831582 B1 (AMAZON TECHNOLOGIES INC.) 2010年 11月 9日 (2010-11-09) 全文	1-18

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/098210

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	109815392	A	2019年 5月 28日		无		
CN	103577432	A	2014年 2月 12日	CN	103577432	B	2017年 7月 14日
CN	108241699	A	2018年 7月 3日		无		
CN	107862027	A	2018年 3月 30日	CN	107862027	B	2019年 3月 12日
				CN	109815308	A	2019年 5月 28日
				WO	2019085236	A1	2019年 5月 9日
CN	105069086	A	2015年 11月 18日	CN	105069086	B	2017年 7月 11日
CN	102629257	A	2012年 8月 8日	CN	102629257	B	2014年 2月 19日
CN	104063523	A	2014年 9月 24日	CN	104063523	B	2017年 2月 22日
US	7831582	B1	2010年 11月 9日	US	7860895	B1	2010年 12月 28日
				US	7774335	B1	2010年 8月 10日