

(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
2017 年 8 月 31 日 (31.08.2017)



(10) 国际公布号
WO 2017/143920 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2017/073445
- (22) 国际申请日: 2017 年 2 月 14 日 (14.02.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610107373.8 2016 年 2 月 26 日 (26.02.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 蔡宁 (CAI, Ning); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 张凯 (ZHANG, Kai); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 杨旭 (YANG, Xu); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, 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,

[见续页]

(54) Title: METHOD AND APPARATUS FOR DISTINGUISHING TOPICS

(54) 发明名称: 一种区分主题的方法及装置

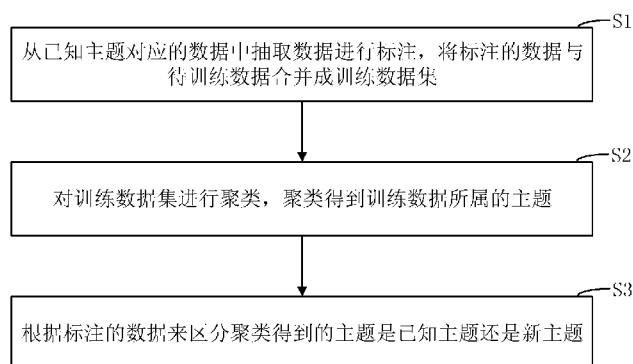


图 1

S1 EXTRACTING DATA FROM DATA CORRESPONDING TO A KNOWN TOPIC FOR MARKING, AND MERGING THE MARKED DATA AND DATA TO BE TRAINED INTO A TRAINING DATA SET

S2 CLUSTERING THE TRAINING DATA SET, AND OBTAINING, BY MEANS OF THE CLUSTERING, A TOPIC TRAINING DATA BELONGS TO

S3 ACCORDING TO THE MARKED DATA, PERFORMING DISTINCTION TO DETERMINE WHETHER THE TOPIC OBTAINED BY MEANS OF THE CLUSTERING IS A KNOWN TOPIC OR A NEW TOPIC

(57) Abstract: A method and apparatus for distinguishing topics, used for identifying whether data to be trained belongs to a known topic or a new topic. The method for distinguishing topics comprises: extracting data from data corresponding to the known topic for marking, and merging the marked data and the data to be trained into a training data set (S1); then, clustering the training data set, and obtaining, by means of the clustering, a topic training data belongs to (S2); and finally, according to the marked data, performing distinction to determine whether the topic obtained by means of the clustering is a known topic or a new topic (S3). The apparatus comprises a data extraction module, a clustering module and a topic distinguishing module. By means of the method and the apparatus, the difference between the understanding of a person regarding a problem and the understanding of a machine regarding the problem is reduced, and the problems proposed by a client can be distinguished accurately.

(57) 摘要: 一种区分主题的方法及装置, 用于识别待训练数据属于已知主题还是新主题, 所述区分主题的方法从已知主题对应的数据中抽取数据进行标注, 将标注的数据与待训练数据合并成训练数据集 (S1); 然后对训练数据集进行聚类, 聚类得到训练数据所属的主题 (S2); 最后根据标注的数据来区分聚类得到的主题是已知主题还是新主题 (S3)。所述装置包括数据抽取模块、聚类模块和主题区分模块。所述方法及装置减少了人对问题的

理解和机器对问题理解的差异, 能够准确地区分客户所提出的问题。

WO 2017/143920 A1



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

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)。

- (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,

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

一种区分主题的方法及装置

本申请要求 2016 年 02 月 26 日递交的申请号为 201610107373.8、发明名称为“一种区分主题的方法及装置”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。

5

技术领域

本发明属于数据处理技术领域，尤其涉及一种区分主题的方法及装置。

背景技术

10 用户在使用产品或者服务的时候经常会遇到自己无法处理的问题，或者需要咨询的问题，进而会寻求客服帮助。每天用户的咨询量非常大，咨询的角度也不同，很多用户都在问相同的问题，有些是客服已经知道的老问题，有一些是没有被发现的新问题。

了解用户提出的问题，可以对产品的设计和改进提供帮助。例如如果发现新问题，有可能是产品需要改进的点。如果老问题的咨询量变多或者变少，则可能是某个功能块
15 用户数减少，也是需要引起注意的。因此，如何从大规模的对话中找出问题，并且能区分出新问题和老问题，是非常必要的并且重要的。

LDA (Latent Dirichlet Allocation) 方法是一种文档主题生成模型，非常适合从大量的对话中获取问题，每一篇文档代表了一些主题所构成的一个概率分布，而每一个主题又代表了很多单词所构成的一个概率分布。语料库中的每一篇文档与 T (通过反复
20 试验等方法事先给定) 个主题的一个多项分布相对应，将该多项分布记为 θ 。每个主题又与词汇表 (vocabulary) 中的 V 个单词的一个多项分布相对应，将这个多项分布记为 ϕ 。上述词汇表是由语料库中所有文档中的所有互异单词组成，但实际建模的时候要剔除一些停用词 (stopword)，还要进行一些词干化 (stemming) 处理等。 θ 和 ϕ 分别有一个带有超参数 (hyperparameter) α 和 β 的 Dirichlet 先验分布。对于一篇文档 d
25 中的每一个单词，从该文档所对应的多项分布 θ 中抽取一个主题 z ，然后再从主题 z 所对应的多项分布 ϕ 中抽取一个单词 w 。将这个过程重复 N_d 次，就产生了文档 d ，这里的 N_d 是文档 d 的单词总数。

LDA 方法是一种非监督机器学习技术，LDA 可以用来识别大规模文档集 (document collection) 或语料库 (corpus) 中潜藏的主题信息，通过聚类发现问题。但是 LDA 方
30 法本身不能区分哪些是新问题哪些是老问题，而且人对问题的理解和机器的理解存在差

异，有些老问题可能被打散变成新问题，聚类出来的问题并不一定是想要的问题。

发明内容

5 本发明的目的是提供一种区分主题的方法及装置，以解决现在技术无法区分聚类得到的主题是新主题还是老主题，或者聚类得到的主题并不一定是想要的主题等技术问题。

为了实现上述目的，本发明技术方案如下：

一种区分主题的方法，用于识别待训练数据属于已知主题还是新主题，所述区分主题的方法包括：

10 从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集；

对训练数据集进行聚类，聚类得到训练数据所属的主题；

根据标注的数据来区分聚类得到的主题是已知主题还是新主题。

进一步地，所述对训练数据集进行聚类，所采用的聚类方法为 LDA 聚类方法。

15 进一步地，所述 LDA 聚类方法聚类的主题数目大于已知主题数目。

进一步地，所述从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集，所标注的标注数据少于待训练数据。

进一步地，所述根据标注的数据来区分聚类得到的主题是已知主题还是新主题，包括：

20 如果一个已知主题的所有标注数据只出现在一个主题中，那么这个主题是已知主题；

如果一个主题中没有出现已知主题的标注数据，则该主题是新主题。

进一步地，所述对训练数据集进行聚类，聚类得到训练数据所属的主题，还包括：聚类出对应到每个主题下每一个关键词的概率。

25 进一步地，所述根据标注的数据来区分聚类得到的主题是已知主题还是新主题，还包括：

根据每个聚类得到的主题对应的关键词，进一步判断聚类得到的主题是已知主题还是新主题。

30 本发明还提出了一种区分主题的装置，用于识别待训练数据属于已知主题还是新主题，所述区分主题的装置包括：

数据抽取模块，用于从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集；

聚类模块，用于对训练数据集进行聚类，聚类得到训练数据所属的主题；

主题区分模块，用于根据标注的数据来区分聚类得到的主题是已知主题还是新主

5 题。

优选地，所述聚类模块对训练数据集进行聚类，所采用的聚类方法为 LDA 聚类方法。

进一步地，所述 LDA 聚类方法聚类的主题数目大于已知主题数目。

进一步地，所述数据抽取模块从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集，所标注的标注数据少于待训练数据。

10 进一步地，所述主题区分模块根据标注的数据来区分聚类得到的主题是已知主题还是新主题，执行如下操作：

如果一个已知主题的所有标注数据只出现在一个主题中，那么这个主题是已知主题；

如果一个主题中没有出现已知主题的标注数据，则该主题是新主题。

15 进一步地，所述聚类模块对训练数据集进行聚类，聚类得到训练数据所属的主题，还包括：

聚类出对应到每个主题下每一个关键词的概率。

进一步地，所述主题区分模块根据标注的数据来区分聚类得到的主题是已知主题还是新主题，还执行如下步骤：

20 根据每个聚类得到的主题对应的关键词，进一步判断聚类得到的主题是已知主题还是新主题。

本发明提出的一种区分主题的方法及装置，使用半监督的方法，利用少量标注的数据区分 LDA 聚类出来的主题是否是客服已知的问题，并且减少人对问题的理解和机器对问题理解的差异，能够准确地区分客户所提出的问题。

25

附图说明

图 1 为本发明区分主题的方法流程图；

图 2 为本发明区分主题的装置结构示意图。

30 具体实施方式

下面结合附图和实施例对本发明技术方案做进一步详细说明，以下实施例不构成对本发明的限定。

本实施例以客户服务过程中经常遇到的客户咨询为例，一般客服人员需要根据与客户的对话，来判别客户所咨询的问题是什么。而进一步判断用户所咨询的问题属于新问题还是老问题，对产品的改进和发展也是非常重要的。在本实施例中，将客户与客服人员之间的对话作为训练数据，通过 LDA 聚类从大量的对话中获知客户所要咨询的问题，客户所要咨询的问题即通过 LDA 聚类得到的主题，并进一步判断客户所要咨询的问题是新问题还是老问题。

如图 1 所示，本实施例一种区分主题的方法，包括如下步骤：

10 步骤 S1、从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集。

本实施例根据历史经验数据总结出一些老问题，将这些老问题作为已知主题。客服人员在平时的工作中会积累经验，根据接听到的与客户的对话数据，得到一些已知主题。从这些已知主题对应的对话数据中选取部分数据进行标注，对每个已知主题标注少量数据，一般为 3-5 条对话数，标注数据的量级比待训练数据要小很多，不会对训练数据的聚类结果造成影响。

例如：

- A. 我的资格够了，为什么开通不了。标注：开通不了。
- B. 我已经实名认证了，怎么还开通不了啊。标注：开通不了。
- 20 C. 我朋友都开通了，为什么开通不了。标注：开通不了。
- D. 为什么开通不了啊。标注：开通不了。

将标注的数据 A、B、C、D 与待训练数据合并成新的训练数据集，用于进行后续的聚类。

步骤 S2、对训练数据集进行聚类，聚类得到训练数据所属的主题。

25 本实施例进行聚类的方法采用 LDA 聚类方法，LDA 聚类方法是一种非监督机器学习技术，LDA 可以用来识别大规模文档集（document collection）或语料库（corpus）中潜藏的主题信息。

LDA 聚类简单来说就是为一堆文档进行聚类（所以是非监督学习），一种主题（topic）就是一类，要聚成的主题（topic）数目是事先指定的，topic 数目一般是根据经验指定一个值，在本实施例中可以用老问题数目的 3 倍作为 topic 数目。聚类的结果是一个概

30

率。以下通过一个例子进行说明，假设有以下几组句子：

- 1、我喜欢吃西兰花和香蕉；
- 2、我吃了有香蕉和菠菜汁的早餐；
- 3、龙猫和小猫很可爱；
- 4、我妹妹昨天收养了一只小猫；
- 5、看看这个可爱的仓鼠嚼一块西兰花。

如果对这些句子进行 LDA 聚类，聚类的 topic 有两个，分别是 TopicA 和 TopicB，那么经过 LDA 聚类后得到：

句子 1 和 2：100% 属于 Topic A；

句子 3 和 4：100% 属于 Topic B；

句子 5：60%属于 Topic A，40% 属于 Topic B；

Topic A：30% 西兰花，15% 香蕉，10% 早餐，10%咀嚼，...（对于任何一个单词，可以得知 TopicA 都关系到一个食物的主题）；

Topic B：20% 龙猫，20% 小猫，20%可爱，15%仓鼠，...（对于任何一个单词，可以得知 TopicB 都关系到一个可爱的动物的主题。

上面关于句子 5 的结果，可以看出来是一个明显的概率类型的聚类结果，在本实施例中会认为句子 5 属于 topicA。句子 1 和 2 正好都是 100%的确定性结果。

再看例子中的结果，除了为每句话得出了一个概率的聚类结果，而且对每个 Topic，都有代表性的词以及一个比例。以 Topic A 为例，就是说所有对应到 Topic A 的词里面，有 30%的词是西兰花。在 LDA 算法中，会把每一个文档中的每一个词对应到一个 Topic。

本实施例通过 LDA 聚类方法可以从训练数据集中识别出训练数据所属的主题，以及对应的概率。例如：句子 5：60%属于 Topic A，40% 属于 Topic B。还可以进一步聚类出对应到每个主题下每一个关键词的概率，从而可以通过这些关键词来判断主题是已知问题还是新问题。

需要说明的是，本发明不限于所采用的聚类方法，例如可以采用 LDA 聚类方法，也可以采用 Kmeans 聚类方法，优选地采用 LDA 聚类方法，LDA 聚类方法能够确定训练数据对应的主题，以及主题中各个关键词的概率，有利于进一步对主题进行分析判断。

步骤 S3、根据标注的数据来区分聚类得到的主题是已知主题还是新的主题。

在通过 LDA 聚类方法识别出训练数据所属的主题后，可以根据标注的数据来区分聚类得到的主题是已知主题还是新主题。

具体判断方法包括：

1)、如果一个已知主题的所有标注数据只出现在一个主题中，那么这个主题就是个已知主题；

2)、如果一个主题中没有出现已知主题的标注数据，则该主题就是新主题；

5 3)、如果一个已知主题的标注数据出现在不同的主题里，则出现有标注数据的主题有可能是同一个已知主题的细化，需要进一步进行判断是已知主题还是新主题。具体的判断可以根据每个主题中出现的关键词来人工判断，即根据关键词所属的主题来进行判断，这里不再赘述。

例如：

10 如果标注的句子 A、B、C、D 都属于 topic1，那么就认为 topic1 是已知主题老问题开通不了。

如果 A、B 属于 topic1，C、D 属于 topic2，那么就认为 topic1，topic2 都可能是已知主题老问题开通不了，再通过关键词进一步确定。

如果 topic3 中都没有出现 A、B、C、D，则 topic3 是新主题。

15 需要说明的是，在根据标注的数据来区分聚类得到的主题是已知主题还是新主题，也可以根据标注数据在主题中出现的多少来进行判断，出现的多则认为是老问题，出现的多少可以根据实际情况来设定，并不一定要求所有标注的数据都出现在主题中才认定为是已知主题。

20 如图 2 所示，本实施例还提出了一种区分主题的装置，用于识别待训练数据属于已知主题还是新主题，该区分主题的装置包括：

数据抽取模块，用于从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集；

聚类模块，用于对训练数据集进行聚类，聚类得到训练数据所属的主题；

25 主题区分模块，用于根据标注的数据来区分聚类得到的主题是已知主题还是新主题。

优选地，聚类模块对训练数据集进行聚类，所采用的聚类方法为 LDA 聚类方法。

本实施例 LDA 聚类方法聚类的主题数目大于已知主题数目。

本实施例数据抽取模块从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集，所标注的标注数据少于待训练数据。

30 本实施例主题区分模块根据标注的数据来区分聚类得到的主题是已知主题还是新

主题，执行如下操作：

如果一个已知主题的所有标注数据只出现在一个主题中，那么这个主题是已知主题；

如果一个主题中没有出现已知主题的标注数据，则该主题是新主题。

- 5 本实施例聚类模块对训练数据集进行聚类，聚类得到训练数据所属的主题，还包括：聚类出对应到每个主题下每一个关键词的概率。

显然，本实施例主题区分模块根据标注的数据来区分聚类得到的主题是已知主题还是新主题，还执行如下步骤：

- 10 根据每个聚类得到的主题对应的关键词，进一步判断聚类得到的主题是已知主题还是新主题。

以上实施例仅用以说明本发明的技术方案而非对其进行限制，在不背离本发明精神及其实质的情况下，熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形，但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

权利要求书

1. 一种区分主题的方法，用于识别待训练数据属于已知主题还是新主题，其特征在于，所述区分主题的方法包括：

从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集；

对训练数据集进行聚类，聚类得到训练数据所属的主题；

根据标注的数据来区分聚类得到的主题是已知主题还是新主题。

2. 根据权利要求 1 所述的区分主题的方法，其特征在于，所述对训练数据集进行聚类，所采用的聚类方法为 LDA 聚类方法。

3. 根据权利要求 2 所述的区分主题的方法，其特征在于，所述 LDA 聚类方法聚类的主题数目大于已知主题数目。

4. 根据权利要求 1 所述的区分主题的方法，其特征在于，所述从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集，所标注的标注数据少于待训练数据。

5. 根据权利要求 1 所述的区分主题的方法，其特征在于，所述根据标注的数据来区分聚类得到的主题是已知主题还是新主题，包括：

如果一个已知主题的所有标注数据只出现在一个主题中，那么这个主题是已知主题；

如果一个主题中没有出现已知主题的标注数据，则该主题是新主题。

6. 根据权利要求 5 所述的区分主题的方法，其特征在于，所述对训练数据集进行聚类，聚类得到训练数据所属的主题，还包括：

聚类出对应到每个主题下每一个关键词的概率。

7. 根据权利要求 6 所述的区分主题的方法，其特征在于，所述根据标注的数据来区分聚类得到的主题是已知主题还是新主题，还包括：

根据每个聚类得到的主题对应的关键词，进一步判断聚类得到的主题是已知主题还是新主题。

8. 一种区分主题的装置，用于识别待训练数据属于已知主题还是新主题，其特征在于，所述区分主题的装置包括：

数据抽取模块，用于从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集；

聚类模块，用于对训练数据集进行聚类，聚类得到训练数据所属的主题；

主题区分模块，用于根据标注的数据来区分聚类得到的主题是已知主题还是新主题。

5 9. 根据权利要求 8 所述的区分主题的装置，其特征在于，所述聚类模块对训练数据集进行聚类，所采用的聚类方法为 LDA 聚类方法。

10. 根据权利要求 9 所述的区分主题的装置，其特征在于，所述 LDA 聚类方法聚类的主题数目大于已知主题数目。

11. 根据权利要求 8 所述的区分主题的装置，其特征在于，所述数据抽取模块从已知主题对应的数据中抽取数据进行标注，将标注的数据与待训练数据合并成训练数据集，所标注的标注数据少于待训练数据。

12. 根据权利要求 8 所述的区分主题的装置，其特征在于，所述主题区分模块根据标注的数据来区分聚类得到的主题是已知主题还是新主题，执行如下操作：

如果一个已知主题的所有标注数据只出现在一个主题中，那么这个主题是已知主题；

15 如果一个主题中没有出现已知主题的标注数据，则该主题是新主题。

13. 根据权利要求 12 所述的区分主题的装置，其特征在于，所述聚类模块对训练数据集进行聚类，聚类得到训练数据所属的主题，还包括：

聚类出对应到每个主题下每一个关键词的概率。

20 14. 根据权利要求 13 所述的区分主题的装置，其特征在于，所述主题区分模块根据标注的数据来区分聚类得到的主题是已知主题还是新主题，还执行如下步骤：

根据每个聚类得到的主题对应的关键词，进一步判断聚类得到的主题是已知主题还是新主题。

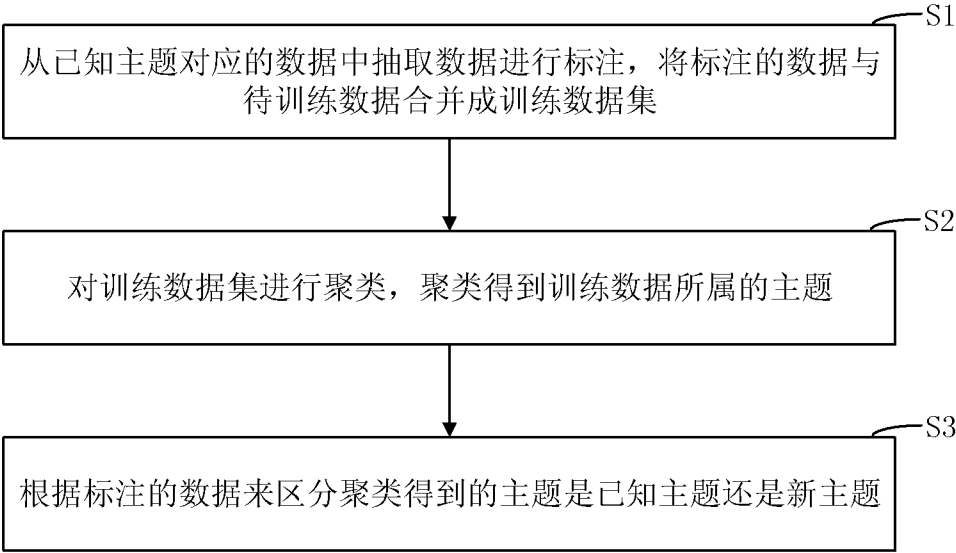


图 1

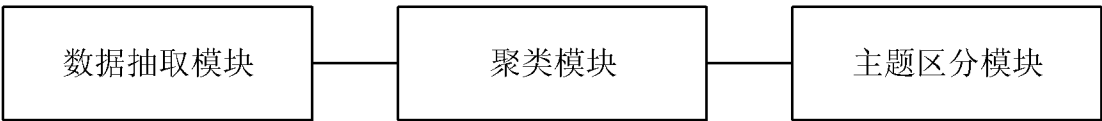


图 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/073445

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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)

CNPAT, WPI, EPODOC, IEEE, CNKI, PATENTICS, GOOGLE: subject, topic, theme, new, cluster, train, merge, combine, label, sign, marked, keyword

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103559175 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 05 February 2014 (05.02.2014) description, paragraphs [0100]-[0118], claim 1, and figures 5-7	1-14
X	CN 104054072 A (IBM) 17 September 2014 (17.09.2014) description, paragraphs [0048]-[0055], and figure 4	1-14
A	CN 102902700 A (NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY) 30 January 2013 (30.01.2013) the whole document	1-14
A	US 2011252015 A1 (BARD, KRISTINA BUTVYDAS et al.) 13 October 2011 (13.10.2011) the whole document	1-14

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 18 April 2017	Date of mailing of the international search report 28 April 2017
Name and mailing address of the ISA 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 TAN, Jingchan Telephone No. (86-10) 53318982

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/073445

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103559175 A	05 February 2014	None	
CN 104054072 A	17 September 2014	GB 2509874 A	16 July 2014
		WO 2013086931 A1	20 June 2013
		DE 112012005197 T5	28 August 2014
		US 2013151522 A1	13 June 2013
CN 102902700 A	30 January 2013	None	
US 2011252015 A1	13 October 2011	US 2009037412 A1	05 February 2009

A. 主题的分类 G06F 17/30 (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)) CNPAT, WPI, EPODOC, IEEE, CNKI, PATENTICS, GOOGLE: 主题, 标题, 新, 聚类, 训练, 合并, 标注, 标记, 关键词, subject, topic, theme, new, cluster, train, merge, combine, label, sign, marked, keyword																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 103559175 A (华南理工大学) 2014年 2月 5日 (2014 - 02 - 05) 说明书第[0100]-[0118]段, 权利要求1, 图5-7</td> <td>1-14</td> </tr> <tr> <td>X</td> <td>CN 104054072 A (国际商业机器公司) 2014年 9月 17日 (2014 - 09 - 17) 说明书第[0048]-[0055]段, 图4</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 102902700 A (中国人民解放军国防科学技术大学) 2013年 1月 30日 (2013 - 01 - 30) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>US 2011252015 A1 (BARD, KRISTINA BUTVYDAS 等) 2011年 10月 13日 (2011 - 10 - 13) 全文</td> <td>1-14</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 103559175 A (华南理工大学) 2014年 2月 5日 (2014 - 02 - 05) 说明书第[0100]-[0118]段, 权利要求1, 图5-7	1-14	X	CN 104054072 A (国际商业机器公司) 2014年 9月 17日 (2014 - 09 - 17) 说明书第[0048]-[0055]段, 图4	1-14	A	CN 102902700 A (中国人民解放军国防科学技术大学) 2013年 1月 30日 (2013 - 01 - 30) 全文	1-14	A	US 2011252015 A1 (BARD, KRISTINA BUTVYDAS 等) 2011年 10月 13日 (2011 - 10 - 13) 全文	1-14
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 103559175 A (华南理工大学) 2014年 2月 5日 (2014 - 02 - 05) 说明书第[0100]-[0118]段, 权利要求1, 图5-7	1-14															
X	CN 104054072 A (国际商业机器公司) 2014年 9月 17日 (2014 - 09 - 17) 说明书第[0048]-[0055]段, 图4	1-14															
A	CN 102902700 A (中国人民解放军国防科学技术大学) 2013年 1月 30日 (2013 - 01 - 30) 全文	1-14															
A	US 2011252015 A1 (BARD, KRISTINA BUTVYDAS 等) 2011年 10月 13日 (2011 - 10 - 13) 全文	1-14															
☐ 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 2017年 4月 18日		国际检索报告邮寄日期 2017年 4月 28日															
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (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 覃婧婵 电话号码 (86-10) 53318982															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/073445

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	103559175	A	2014年 2月 5日	无	
CN	104054072	A	2014年 9月 17日	GB	2509874 A 2014年 7月 16日
				WO	2013086931 A1 2013年 6月 20日
				DE	112012005197 T5 2014年 8月 28日
				US	2013151522 A1 2013年 6月 13日
CN	102902700	A	2013年 1月 30日	无	
US	2011252015	A1	2011年 10月 13日	US	2009037412 A1 2009年 2月 5日