(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 7 月 16 日 (16.07.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/143320 A1

(51) 国际专利分类号:
G06F 17/00 (2019.01)

(21) 国际申请号: PCT/CN2019/118244

(22) 国际申请日: 2019 年 11 月 14 日 (14.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910012554.6 2019年1月7日 (07.01.2019) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: 郑立颖 (ZHENG, Liying); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 金戈 (JIN, Ge); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 徐亮 (XU, Liang); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市精英专利事务所 (SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田区深南中路6009号绿景广场B栋20层B, Guangdong 518000 (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,

(54) **Title:** METHOD AND APPARATUS FOR ACQUIRING WORD VECTORS OF TEXT, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) **发明名称**: 文本词向量获取方法、装置、计算机设备及存储介质

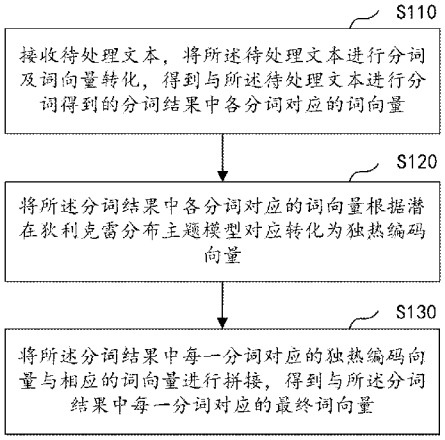

图 2

S110 Receive a text to be processed, and perform word segmentation and word vector conversion on the text, so as to obtain word vectors corresponding to respective word segments in a word segmentation result, the word segmentation result being obtained by means of performing the word segmentation on the text

S120 Correspondingly convert, according to a latent Dirichlet allocation topic model, the word vectors corresponding to the respective word segments in the word segmentation result into one-hot encoding vectors

S130 Combine the one-hot encoding vectors corresponding to the respective word segments in the word segmentation result with the corresponding word vectors, so as to obtain final word vectors corresponding to the respective word segments in the word segmentation result

(57) **Abstract:** The present application discloses a method and apparatus for acquiring word vectors of a text, a computer device, and a storage medium. The method comprises: receiving a text to be processed, and performing word segmentation and word vector conversion on the text, so as to obtain word vectors corresponding to respective word segments in a word segmentation result, the word segmentation result being obtained by means of performing the word segmentation on the text; correspondingly converting, according to a latent Dirichlet allocation topic model, the word vectors corresponding to the respective word segments in the word segmentation result into one-hot encoding vectors; and combining the one-hot encoding vectors corresponding to the respective word segments in the word segmentation result with the corresponding word vectors, so as to obtain final word vectors corresponding to the respective word segments in the word segmentation result.

WO 2020/143320 A1

[见续页]

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)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：本申请公开了文本词向量获取方法、装置、计算机设备及存储介质。该方法包括：接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量；将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量；以及将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

文本词向量获取方法、装置、计算机设备及存储介质

本申请要求于2019年1月7日提交中国专利局、申请号为201910012554.6、申请名称为“文本词向量获取方法、装置、计算机设备及存储介质”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

本申请涉及语义解析技术领域，尤其涉及一种文本词向量获取方法、装置、计算机设备及存储介质。

背景技术

目前，一般使用深度学习处理文本类数据的时候需要对文本进行分词，进而将词表示为词向量方式作为特征输入；目前已有的词向量表征算法中只考虑到词的相邻词信息，故基于目前已有的词向量表征算法信息量少，识别准确率不高。

发明内容

本申请实施例提供了一种文本词向量获取方法、装置、计算机设备及存储介质，旨在解决现有技术中使用深度学习处理文本类数据的时对文本进行分词，进而将词表示为词向量方式作为特征输入，词向量表征算法信息量少，识别准确率不高的问题。

第一方面，本申请实施例提供了一种文本词向量获取方法，其包括：

接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量；

将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量；以及

将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

第二方面，本申请实施例提供了一种文本词向量获取装置，其包括：

词向量获取单元，用于接收待处理文本，将所述待处理文本进行分词及词

向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量；

独热编码单元，用于将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量；以及

最终词向量获取单元，用于将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

第三方面，本申请实施例又提供了一种计算机设备，其包括存储器、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上述第一方面所述的文本词向量获取方法。

第四方面，本申请实施例还提供了一种计算机可读存储介质，其中所述计算机可读存储介质存储有计算机程序，所述计算机程序当被处理器执行时使所述处理器执行上述第一方面所述的文本词向量获取方法。

附图说明

为了更清楚地说明本申请实施例技术方案，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的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请实施例提供的文本词向量获取方法的应用场景示意图；

图 2 为本申请实施例提供的文本词向量获取方法的流程示意图；

图 3 为本申请实施例提供的文本词向量获取方法的另一流程示意图；

图 4 为本申请实施例提供的文本词向量获取方法的子流程示意图；

图 5 为本申请实施例提供的文本词向量获取方法的另一子流程示意图；

图 6 为本申请实施例提供的文本词向量获取装置的示意性框图；

图 7 为本申请实施例提供的文本词向量获取装置的另一示意性框图；

图 8 为本申请实施例提供的文本词向量获取装置的子单元示意性框图；

图 9 为本申请实施例提供的文本词向量获取装置的另一子单元示意性框图；

图 10 为本申请实施例提供的计算机设备的示意性框图。

具体实施方式

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。

应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和 “包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

还应当理解，在此本申请说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本申请。如在本申请说明书和所附权利要求书中所使用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。

还应当进一步理解，在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

请参阅图 1 和图 2，图 1 是本申请实施例提供的文本词向量获取方法的应用场景示意图，图 2 是本申请实施例提供的文本词向量获取方法的流程示意图，该文本词向量获取方法应用于服务器中，该方法通过安装于服务器中的应用软件进行执行。

如图 2 所示，该方法包括步骤 S110~S130。

S110、接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量。

在本实施例中，当用户端上传了待处理文本至服务器时，服务器接收所述待处理文本。之后通过服务器获取所述待处理文本所对应的词向量， 以便于后续进行语义分析。

在一实施例中，如图 4 所示，步骤 S110 包括：

S111、将所述待处理文本通过基于概率统计分词模型进行分词，得到与所述待处理文本对应的分词结果；

S112、通过用于将单词转化为向量的 Word2Vec 模型获取所述分词结果中各分词对应的词向量。

在本实施例中，对所述待处理文本进行分词时，是通过基于概率统计模型的分词方法进行分词。例如，令 C=C1C2...Cm，C 是待切分的汉字串，令 W=W1W2...Wn，W 是切分的结果，Wa，Wb，……，Wk 是 C 的所有可能的切分方案。那么，基于概率统计的切分模型就是能够找到目的词串 W ，使得 W 满足：P(W|C)=MAX(P(Wa|C),P(Wb|C)...P(Wk|C))的分词模型，上述分词模型得到的词串 W 即估计概率为最大之词串。

即对一个待分词的子串 S，按照从左到右的顺序取出全部候选词 w_1、w_2、…、w_i、…、w_n；在词典中查出每个候选词的概率值 P(wi)，并记录每个候选词的全部左邻词；计算每个候选词的累积概率，同时比较得到每个候选词的最佳左邻词；如果当前词 w_n 是字串 S 的尾词，且累积概率 P(wn)最大，则 w_n 就是 S 的终点词；从 w_n 开始，按照从右到左顺序，依次将每个词的最佳左邻词输出，即 S 的分词结果。

当获取了与所述待处理文本对应的分词结果，通过用于将词语转化为向量的 Word2Vec 模型对所述分词结果中每一分词进行转化，得到与每一分词对应的词向量。其中，Word2Vec 是从大量文本语料中以无监督的方式学习语义知识的一种模型，能将分词结果中每一词语转化成对应的词向量，具体可将分词结果每一词语都转化为一个 k 维的行向量。

S120、将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量。

在本实施例中，潜在狄利克雷分布主题模型（Latent Dirichlet Allocation，简记为 LDA）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。所谓生成模型，就是说，可认为一篇文章的每个词都是通过“以一定概率选择了某个主题，并从这个主题中以一定概率选择某个词语”这样一个过程得到。文档到主题服从多项式分布，主题到词服从多项式分布。

对于语料库中的每篇文档，LDA 定义了如下生成过程：

1）对每一篇文档，从主题分布中抽取一个主题；

2）从上述被抽到的主题所对应的单词分布中抽取一个单词；

3）重复上述步骤 1）-2）过程直至遍历文档中的每一个单词。

语料库中的每一篇文档与 T（通过反复试验等方法事先给定）个主题的一个

多项分布 （multinomialdistribution）相对应，将该多项分布记为θ。每个主题又与词汇表中的V个单词的一个多项分布相对应，将这个多项分布记为φ。

在一实施例中，如图5所示，步骤S120包括：

S121、获取所述分词结果，将所述分词结果中每一分词作为根据语料库预先训练所得到的潜在狄利克雷分布主题模型的输入，得到与所述分词结果中每一分词对应的主题；

S122、将所述分词结果中每一分词的主题分别进行独热编码，得到与所述分词结果中各分词一一对应的独热编码向量。

在本实施例中，独热编码即one-hot编码，其将离散型特征的每一种取值都看成一种状态，若某一特征中有N个不相同的取值，那么就可以将该特征抽象成N种不同的状态，one-hot编码保证了每一个取值只会使得一种状态处于“激活态”，也就是说这N种状态中只有一个状态位值为1，其他状态位都是0。举个例子，假设以学历为例，想要研究的类别为小学、中学、大学、硕士、博士五种类别，使用one-hot对其编码就会得到：小学->[1,0,0,0,0];中学->[0,1,0,0,0];大学->[0,0,1,0,0];硕士->[0,0,0,1,0];博士->[0,0,0,0,1]。

故在确定了每个分词对应主题后，可以以主题在词汇表大集合中所对应的词语而得到独热编码向量。

S130、将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

在一实施例中，步骤S130中具体包括：

将所述分词结果中每一分词对应的独热编码向量拼接至相应的词向量的头部或尾部，得到与所述分词结果中每一分词对应的最终词向量。

在本实施例中，将所述分词结果中每一分词对应的独热编码向量拼接至相应的词向量的头部或尾部，而扩展得到的最终词向量，充分考虑到了主题信息并融入到词向量中，有助于后续自然语言处理任务准确率的提高。例如每一主题对应的词向量为1*300的向量，而每一主题对应的独热编码向量为1*10的向量，则将两向量直接拼接后得到一个1*310的向量。一般是主题对应的词向量在前，主题对应的独热编码向量在后。

在一实施例中，如图3所示，步骤S130之后还包括：

S141、将所述分词结果中每一分词对应的最终词向量从上至下进行组合得

到文本矩阵；

S142、将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量；

S143、将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果。

在本实施例中，通过服务器获取所述待处理文本所对应的词向量及与主题对应的独热编码向量，以组成与所述待处理文本所对应的最终词向量时，将由多个最终词向量组成的文本矩阵输入至已训练的卷积神经网络模型，得到文本识别向量。通过将文本转化为文本识别向量，有效的提取了待进行情感识别文本的文本特征，便于根据文本特征进行情感识别。

在根据与所述待处理文本对应的最终词向量进行情感识别时，具体如下：将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵；将所述文本作为所述已训练的卷积神经网络模型中输入层的输入，得到多个特征图；将多个特征图均输入池化层，得到每一特征图的最大值所对应一维行向量；将每一特征图的最大值所对应一维行向量输入至全连接层，得到与所述待进行情感识别文本对应的文本识别向量。

在本实施例中，将分词结果中各分词对应的词向量从上至下依次排列以得到输入矩阵，将输入矩阵作为已训练的卷积神经网络模型中输入层的输入，得到多个特征图；输入层通过卷积操作得到若干个 Feature Map（Feature Map 可以理解为特征图），卷积窗口的大小为 h ×k ，其中 h 表示纵向词语的个数，而 k 表示向量的维数。通过这样一个大型的卷积窗口，将得到若干个列数为 1 的 Feature Map。

在已训练的卷积神经网络模型的池化层中，可采用从上述多个一维的 Feature Map 中提出最大的值。通过这种池化方式可以解决可变长度的句子输入问题（因为不管 Feature Map 中有多少个值，只需要提取其中的最大值），最终池化层的输出为各个 Feature Map 的最大值，即一个一维的向量。

在已训练的卷积神经网络模型的全连接层中，全连接层的每一个结点都与上一层的所有结点相连，用于将池化层所提取到的特征综合起来，得到一个 1*n 的文本识别向量，例如得到一个 1*310 的文本识别向量。通过卷积神经网络模型能有效提取出待进行情感识别文本中更深层次的特征，使得后续的文本情感

识别更加准确。

在一实施例中，步骤S143之后还包括：

判断所述待处理文本的文本情感识别结果的情感识别结果；其中，所述情感识别结果为正向情感结果、中性情感结果或负向情感结果；

若所述待处理文本的文本情感识别结果为正向情感结果，将所述待处理文本对应的文本文件标题设置为第一显示颜色；

若所述待处理文本的文本情感识别结果为中性情感结果，将所述待处理文本对应的文本文件标题设置为第二显示颜色；

若所述待处理文本的文本情感识别结果为负向情感结果，将所述待处理文本对应的文本文件标题设置为第三显示颜色。

在本实施例中，当完成了对所述待处理文本的情感识别后，此时可以获取各待处理文本的情感识别结果，其中所述情感识别结果为正向情感结果、中性情感结果或负向情感结果，通过这一方式能有效将不同的情感识别结果的待处理文本的文件标题设置为不同的颜色，便于用户直观的根据文件标题的颜色进行区分。

若所述待处理文本的文本情感识别结果为正向情感结果，将所述待处理文本对应的文本文件标题设置为第一显示颜色；例如，所述第一显示颜色为红色。

若所述待处理文本的文本情感识别结果为中性情感结果，将所述待处理文本对应的文本文件标题设置为第二显示颜色；例如，所述第二显示颜色为黑色。

若所述待处理文本的文本情感识别结果为负向情感结果，将所述待处理文本对应的文本文件标题设置为第三显示颜色；例如，所述第三显示颜色绿色。

该方法实现了将词的主题信息融入到其向量表征中，进一步丰富其信息量，有助于后续自然语言处理任务准确率的提高。

本申请实施例还提供一种文本词向量获取装置，该文本词向量获取装置用于执行前述文本词向量获取方法的任一实施例。具体地，请参阅图6，图6是本申请实施例提供的文本词向量获取装置的示意性框图。该文本词向量获取装置100可以配置于服务器中。

如图6所示，文本词向量获取装置100包括词向量获取单元110、独热编码单元120、最终词向量获取单元130。

词向量获取单元110，用于接收待处理文本，将所述待处理文本进行分词及

词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量。

在本实施例中，当用户端上传了待处理文本至服务器时，服务器接收所述待处理文本。之后通过服务器获取所述待处理文本所对应的词向量， 以便于后续进行语义分析。

在一实施例中，如图 8 所示，词向量获取单元 110 包括：

分词单元 111，用于将所述待处理文本通过基于概率统计分词模型进行分词，得到与所述待处理文本对应的分词结果；

词向量转化单元 112，用于通过用于将单词转化为向量的 Word2Vec 模型获取所述分词结果中各分词对应的词向量。

在本实施例中，对所述待处理文本进行分词时，是通过基于概率统计模型的分词方法进行分词。例如，令 C=C1C2...Cm，C 是待切分的汉字串，令 W=W1W2...Wn，W 是切分的结果，Wa，Wb，……，Wk 是 C 的所有可能的切分方案。那么，基于概率统计的切分模型就是能够找到目的词串 W ，使得 W 满足：P(W|C)=MAX(P(Wa|C),P(Wb|C)...P(Wk|C))的分词模型，上述分词模型得到的词串 W 即估计概率为最大之词串。

即对一个待分词的子串 S，按照从左到右的顺序取出全部候选词 w_1、w_2、…、w_i、…、w_n；在词典中查出每个候选词的概率值 P(wi)，并记录每个候选词的全部左邻词；计算每个候选词的累积概率，同时比较得到每个候选词的最佳左邻词；如果当前词 w_n 是字串 S 的尾词，且累积概率 P(wn)最大，则 w_n 就是 S 的终点词；从 w_n 开始，按照从右到左顺序，依次将每个词的最佳左邻词输出，即 S 的分词结果。

当获取了与所述待处理文本对应的分词结果，通过用于将词语转化为向量的 Word2Vec 模型对所述分词结果中每一分词进行转化，得到与每一分词对应的词向量。其中，Word2Vec 是从大量文本语料中以无监督的方式学习语义知识的一种模型，能将分词结果中每一词语转化成对应的词向量，具体可将分词结果每一词语都转化为一个 k 维的行向量。

独热编码单元 120，用于将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量。

在本实施例中，潜在狄利克雷分布主题模型（Latent Dirichlet Allocation，简

记为LDA）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。所谓生成模型，就是说，可认为一篇文章的每个词都是通过“以一定概率选择了某个主题，并从这个主题中以一定概率选择某个词语”这样一个过程得到。文档到主题服从多项式分布，主题到词服从多项式分布。

对于语料库中的每篇文档，LDA定义了如下生成过程：

1）对每一篇文档，从主题分布中抽取一个主题；

2）从上述被抽到的主题所对应的单词分布中抽取一个单词；

3）重复上述步骤1）-2）过程直至遍历文档中的每一个单词。

语料库中的每一篇文档与T（通过反复试验等方法事先给定）个主题的一个多项分布 （multinomialdistribution）相对应，将该多项分布记为θ。每个主题又与词汇表中的V个单词的一个多项分布相对应，将这个多项分布记为φ。

在一实施例中，如图9所示，独热编码单元120包括：

主题获取单元121，用于获取所述分词结果，将所述分词结果中每一分词作为根据语料库预先训练所得到的潜在狄利克雷分布主题模型的输入，得到与所述分词结果中每一分词对应的主题；

独热编码向量获取单元122，用于将所述分词结果中每一分词的主题分别进行独热编码，得到与所述分词结果中各分词一一对应的独热编码向量。

在本实施例中，独热编码即one-hot编码，其将离散型特征的每一种取值都看成一种状态，若某一特征中有N个不相同的取值，那么就可以将该特征抽象成N种不同的状态，one-hot编码保证了每一个取值只会使得一种状态处于“激活态”，也就是说这N种状态中只有一个状态位值为1，其他状态位都是0。举个例子，假设以学历为例，想要研究的类别为小学、中学、大学、硕士、博士五种类别，使用one-hot对其编码就会得到：小学->[1,0,0,0,0];中学->[0,1,0,0,0];大学->[0,0,1,0,0];硕士->[0,0,0,1,0];博士->[0,0,0,0,1]。

故在确定了每个分词对应主题后，可以以主题在词汇表大集合中所对应的词语而得到独热编码向量。

最终词向量获取单元130，用于将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

在一实施例中，最终词向量获取单元 130 具体用于：

将所述分词结果中每一分词对应的独热编码向量拼接至相应的词向量的头部或尾部，得到与所述分词结果中每一分词对应的最终词向量。

在本实施例中，将所述分词结果中每一分词对应的独热编码向量拼接至相应的词向量的头部或尾部，而扩展得到的最终词向量，充分考虑到了主题信息并融入到词向量中，有助于后续自然语言处理任务准确率的提高。例如每一主题对应的词向量为 1*300 的向量，而每一主题对应的独热编码向量为 1*10 的向量，则将两向量直接拼接后得到一个 1*310 的向量。一般是主题对应的词向量在前，主题对应的独热编码向量在后。

在一实施例中，如图 7 所示文本词向量获取装置 100 还包括：

文本矩阵获取单元 141，用于将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵；

文本识别向量获取单元 142，用于将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量；

情感分类单元 143，用于将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果。

在本实施例中，通过服务器获取所述待处理文本所对应的词向量及与主题对应的独热编码向量，以组成与所述待处理文本所对应的最终词向量时，将由多个最终词向量组成的文本矩阵输入至已训练的卷积神经网络模型，得到文本识别向量。通过将文本转化为文本识别向量，有效的提取了待进行情感识别文本的文本特征，便于根据文本特征进行情感识别。

在根据与所述待处理文本对应的最终词向量进行情感识别时，具体如下：将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵；将所述文本作为所述已训练的卷积神经网络模型中输入层的输入，得到多个特征图；将多个特征图均输入池化层，得到每一特征图的最大值所对应一维行向量；将每一特征图的最大值所对应一维行向量输入至全连接层，得到与所述待进行情感识别文本对应的文本识别向量。

在本实施例中，将分词结果中各分词对应的词向量从上至下依次排列以得到输入矩阵，将输入矩阵作为已训练的卷积神经网络模型中输入层的输入，得到多个特征图；输入层通过卷积操作得到若干个 Feature Map（Feature Map 可以

理解为特征图），卷积窗口的大小为 h ×k ，其中 h 表示纵向词语的个数，而 k 表示向量的维数。通过这样一个大型的卷积窗口，将得到若干个列数为 1 的 Feature Map。

在已训练的卷积神经网络模型的池化层中，可采用从上述多个一维的 Feature Map 中提出最大的值。通过这种池化方式可以解决可变长度的句子输入问题（因为不管 Feature Map 中有多少个值，只需要提取其中的最大值），最终池化层的输出为各个 Feature Map 的最大值，即一个一维的向量。

在已训练的卷积神经网络模型的全连接层中，全连接层的每一个结点都与上一层的所有结点相连，用于将池化层所提取到的特征综合起来，得到一个 1*n 的文本识别向量，例如得到一个 1*310 的文本识别向量。通过卷积神经网络模型能有效提取出待进行情感识别文本中更深层次的特征，使得后续的文本情感识别更加准确。

该装置实现了将词的主题信息融入到其向量表征中，进一步丰富其信息量，有助于后续自然语言处理任务准确率的提高。

上述文本词向量获取装置可以实现为计算机程序的形式，该计算机程序可以在如图 10 所示的计算机设备上运行。

请参阅图 10，图 10 是本申请实施例提供的计算机设备的示意性框图。该计算机设备 500 是服务器。其中，服务器可以是独立的服务器，也可以是多个服务器组成的服务器集群。

参阅图 10，该计算机设备 500 包括通过系统总线 501 连接的处理器 502、存储器和网络接口 505，其中，存储器可以包括非易失性存储介质 503 和内存储器 504。

该非易失性存储介质 503 可存储操作系统 5031 和计算机程序 5032。该计算机程序 5032 被执行时，可使得处理器 502 执行文本词向量获取方法。

该处理器 502 用于提供计算和控制能力，支撑整个计算机设备 500 的运行。

该内存储器 504 为非易失性存储介质 503 中的计算机程序 5032 的运行提供环境，该计算机程序 5032 被处理器 502 执行时，可使得处理器 502 执行文本词向量获取方法。

该网络接口 505 用于进行网络通信，如提供数据信息的传输等。本领域技术人员可以理解，图 10 中示出的结构，仅仅是与本申请方案相关的部分结构的

框图，并不构成对本申请方案所应用于其上的计算机设备 500 的限定，具体的计算机设备 500 可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

其中，所述处理器 502 用于运行存储在存储器中的计算机程序 5032，以实现本申请实施例中的文本词向量获取方法。

本领域技术人员可以理解，图 10 中示出的计算机设备的实施例并不构成对计算机设备具体构成的限定，在其他实施例中，计算机设备可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。例如，在一些实施例中，计算机设备可以仅包括存储器及处理器，在这样的实施例中，存储器及处理器的结构及功能与图 10 所示实施例一致，在此不再赘述。

应当理解，在本申请实施例中，处理器 502 可以是中央处理单元 (Central Processing Unit，CPU)，该处理器 502 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor，DSP)、专用集成电路 (Application Specific Integrated Circuit，ASIC)、现成可编程门阵列 (Field-Programmable Gate Array，FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。其中，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。

在本申请的另一实施例中提供计算机可读存储介质。该计算机可读存储介质可以为非易失性的计算机可读存储介质。该计算机可读存储介质存储有计算机程序，其中计算机程序被处理器执行时实现本申请实施例中的文本词向量获取方法。

所述存储介质为实体的、非瞬时性的存储介质，例如可以是 U 盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、磁碟或者光盘等各种可以存储程序代码的实体存储介质。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的设备、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权利要求书

1.一种文本词向量获取方法，包括：

接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量；

将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量；以及

将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

2. 根据权利要求 1 所述的文本词向量获取方法，其中，所述将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量之后，还包括：

将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵；

将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量；

将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果。

3. 根据权利要求 1 所述的文本词向量获取方法，其中，所述将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量，包括：

将所述待处理文本通过基于概率统计分词模型进行分词，得到与所述待处理文本对应的分词结果；

通过用于将单词转化为向量的 Word2Vec 模型获取所述分词结果中各分词对应的词向量。

4. 根据权利要求 1 所述的文本词向量获取方法，其中，所述将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量，包括：

获取所述分词结果，将所述分词结果中每一分词作为根据语料库预先训练所得到的潜在狄利克雷分布主题模型的输入，得到与所述分词结果中每一分词

对应的主题；

将所述分词结果中每一分词的主题分别进行独热编码，得到与所述分词结果中各分词一一对应的独热编码向量。

5. 根据权利要求 1 所述的文本词向量获取方法，其中，所述将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量，包括：

将所述分词结果中每一分词对应的独热编码向量拼接至相应的词向量的头部或尾部，得到与所述分词结果中每一分词对应的最终词向量。

6. 根据权利要求 2 所述的文本词向量获取方法，其中，所述将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量，包括：

将所述文本作为所述已训练的卷积神经网络模型中输入层的输入，得到多个特征图；

将多个特征图均输入池化层，得到每一特征图的最大值所对应一维行向量；

将每一特征图的最大值所对应一维行向量输入至全连接层，得到与所述待进行情感识别文本对应的文本识别向量。

7. 根据权利要求 2 所述的文本词向量获取方法，其中，所述将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果之后，还包括：

判断所述待处理文本的文本情感识别结果的情感识别结果；其中，所述情感识别结果为正向情感结果、中性情感结果或负向情感结果；

若所述待处理文本的文本情感识别结果为正向情感结果，将所述待处理文本对应的文本文件标题设置为第一显示颜色；

若所述待处理文本的文本情感识别结果为中性情感结果，将所述待处理文本对应的文本文件标题设置为第二显示颜色；

若所述待处理文本的文本情感识别结果为负向情感结果，将所述待处理文本对应的文本文件标题设置为第三显示颜色。

8.一种文本词向量获取装置，包括：

词向量获取单元，用于接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词

向量；

独热编码单元，用于将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量；以及

最终词向量获取单元，用于将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

9.根据权利要求 8 所述的文本词向量获取装置，其中，还包括：

文本矩阵获取单元，用于将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵；

文本识别向量获取单元，用于将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量；

情感分类单元，用于将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果。

10.根据权利要求 8 所述的文本词向量获取装置，其中，所述独热编码单元，包括：

主题获取单元，用于获取所述分词结果，将所述分词结果中每一分词作为根据语料库预先训练所得到的潜在狄利克雷分布主题模型的输入，得到与所述分词结果中每一分词对应的主题；

独热编码向量获取单元，用于将所述分词结果中每一分词的主题分别进行独热编码，得到与所述分词结果中各分词一一对应的独热编码向量。

11.一种计算机设备，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以下步骤：

接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量；

将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量；以及

将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

12. 根据权利要求 11 所述的计算机设备，其中，所述将所述分词结果中每

一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量之后，还包括：

将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵；

将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量；

将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果。

13. 根据权利要求 11 所述的计算机设备，其中，所述将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量，包括：

将所述待处理文本通过基于概率统计分词模型进行分词，得到与所述待处理文本对应的分词结果；

通过用于将单词转化为向量的 Word2Vec 模型获取所述分词结果中各分词对应的词向量。

14. 根据权利要求 11 所述的计算机设备，其中，所述将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量，包括：

获取所述分词结果，将所述分词结果中每一分词作为根据语料库预先训练所得到的潜在狄利克雷分布主题模型的输入，得到与所述分词结果中每一分词对应的主题；

将所述分词结果中每一分词的主题分别进行独热编码，得到与所述分词结果中各分词一一对应的独热编码向量。

15. 根据权利要求 11 所述的计算机设备，其中，所述将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量，包括：

将所述分词结果中每一分词对应的独热编码向量拼接至相应的词向量的头部或尾部，得到与所述分词结果中每一分词对应的最终词向量。

16. 根据权利要求 12 所述的计算机设备，其中，所述将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量，包

括：

将所述文本作为所述已训练的卷积神经网络模型中输入层的输入，得到多个特征图；

将多个特征图均输入池化层，得到每一特征图的最大值所对应一维行向量；

将每一特征图的最大值所对应一维行向量输入至全连接层，得到与所述待进行情感识别文本对应的文本识别向量。

17. 根据权利要求 12 所述的计算机设备，其中，所述将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果之后，还包括：

判断所述待处理文本的文本情感识别结果的情感识别结果；其中，所述情感识别结果为正向情感结果、中性情感结果或负向情感结果；

若所述待处理文本的文本情感识别结果为正向情感结果，将所述待处理文本对应的文本文件标题设置为第一显示颜色；

若所述待处理文本的文本情感识别结果为中性情感结果，将所述待处理文本对应的文本文件标题设置为第二显示颜色；

若所述待处理文本的文本情感识别结果为负向情感结果，将所述待处理文本对应的文本文件标题设置为第三显示颜色。

18.一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序当被处理器执行以下操作：

接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量；

将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量；以及

将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量。

19. 根据权利要求 18 所述的计算机可读存储介质，其中，所述将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量之后，还包括：

将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵；

将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本

对应的文本识别向量；

将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果。

20. 根据权利要求 18 所述的计算机可读存储介质，其中，所述将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量，包括：

将所述待处理文本通过基于概率统计分词模型进行分词，得到与所述待处理文本对应的分词结果；

通过用于将单词转化为向量的 Word2Vec 模型获取所述分词结果中各分词对应的词向量。

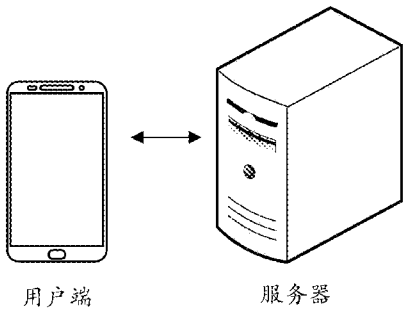
用户端
服务器

图 1

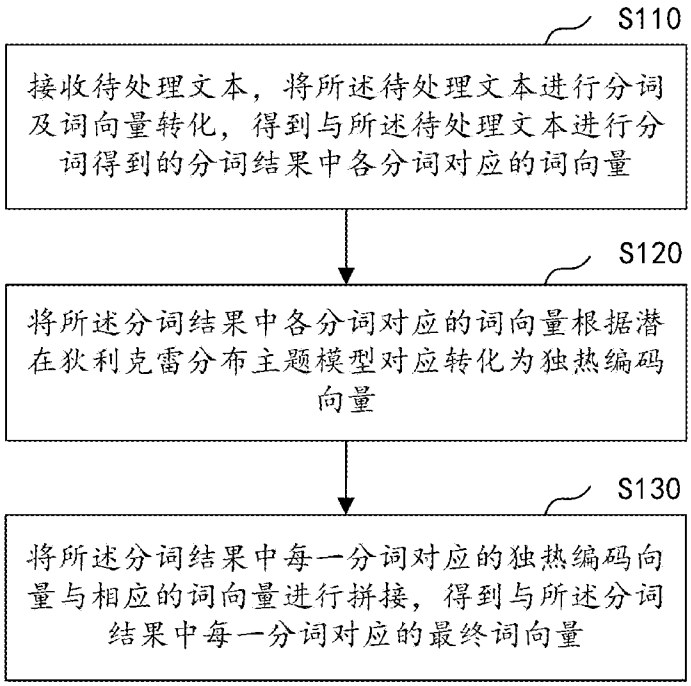
S110
接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量
S120
将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量
S130
将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量

图 2

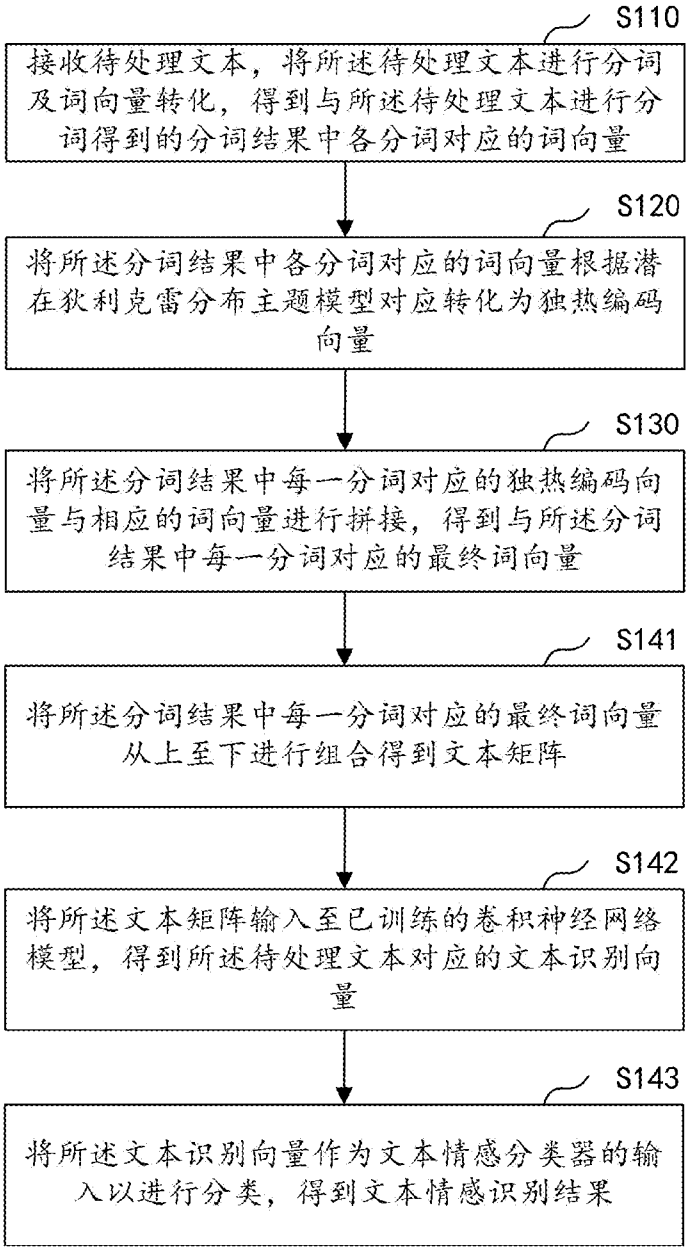
S110
接收待处理文本，将所述待处理文本进行分词及词向量转化，得到与所述待处理文本进行分词得到的分词结果中各分词对应的词向量
S120
将所述分词结果中各分词对应的词向量根据潜在狄利克雷分布主题模型对应转化为独热编码向量
S130
将所述分词结果中每一分词对应的独热编码向量与相应的词向量进行拼接，得到与所述分词结果中每一分词对应的最终词向量
S141
将所述分词结果中每一分词对应的最终词向量从上至下进行组合得到文本矩阵
S142
将所述文本矩阵输入至已训练的卷积神经网络模型，得到所述待处理文本对应的文本识别向量
S143
将所述文本识别向量作为文本情感分类器的输入以进行分类，得到文本情感识别结果

图 3

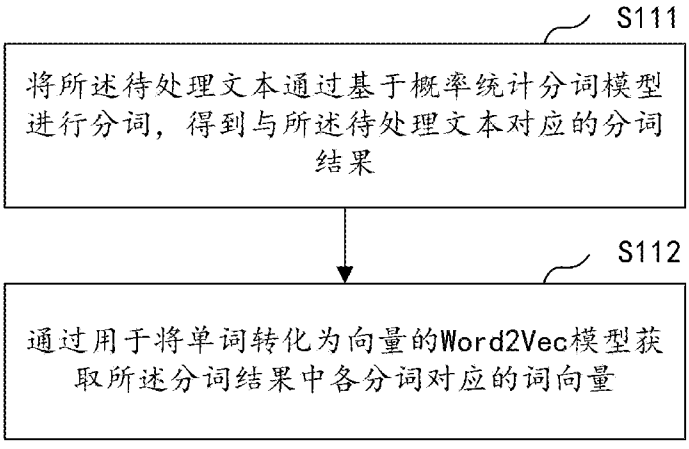
S111
将所述待处理文本通过基于概率统计分词模型进行分词，得到与所述待处理文本对应的分词结果
S112
通过用于将单词转化为向量的Word2Vec模型获取所述分词结果中各分词对应的词向量

图 4

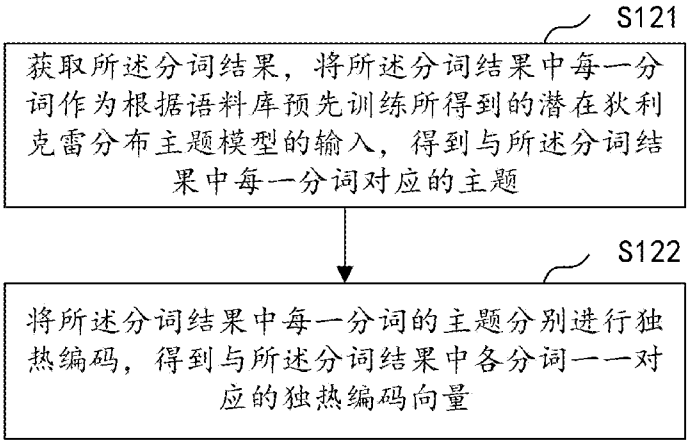
S121
获取所述分词结果，将所述分词结果中每一分词作为根据语料库预先训练所得到的潜在狄利克雷分布主题模型的输入，得到与所述分词结果中每一分词对应的主题
S122
将所述分词结果中每一分词的主题分别进行独热编码，得到与所述分词结果中各分词一一对应的独热编码向量

图 5

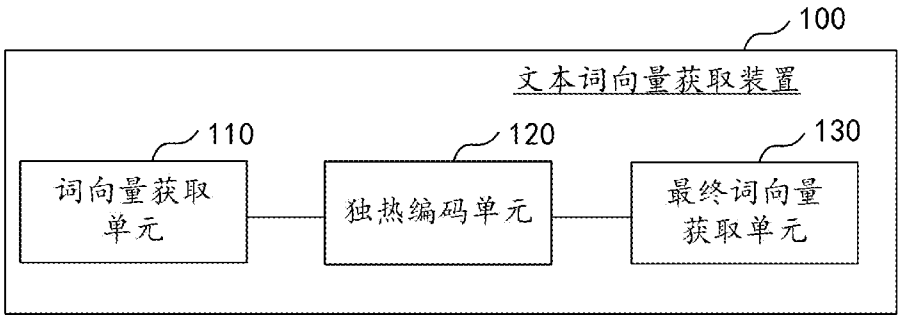
100
文本词向量获取装置
110
词向量获取单元
120
独热编码单元
130
最终词向量获取单元

图 6

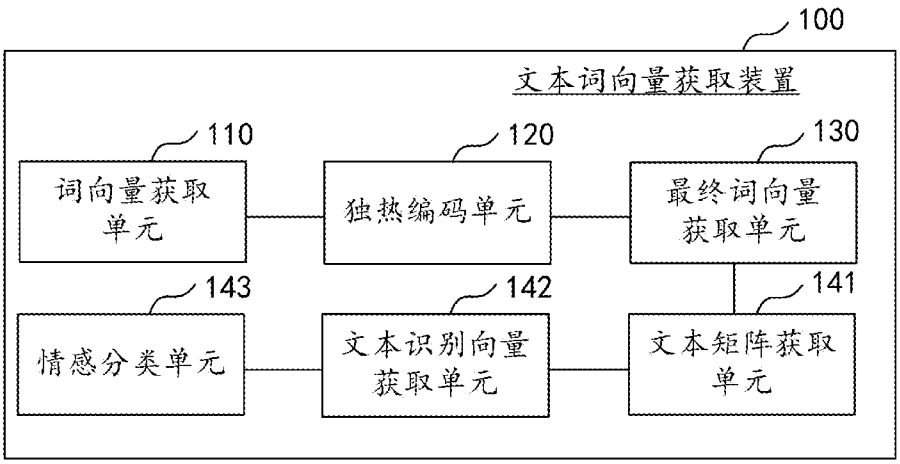
100
文本词向量获取装置
110
词向量获取单元
120
独热编码单元
130
最终词向量获取单元
143
情感分类单元
142
文本识别向量获取单元
141
文本矩阵获取单元

图 7

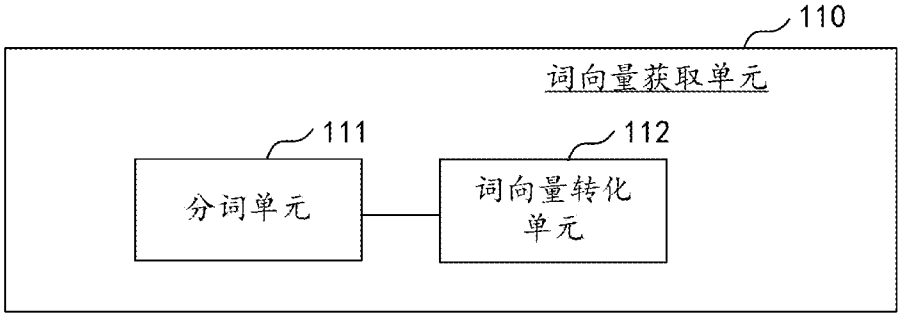
110
词向量获取单元
111
分词单元
112
词向量转化单元

图 8

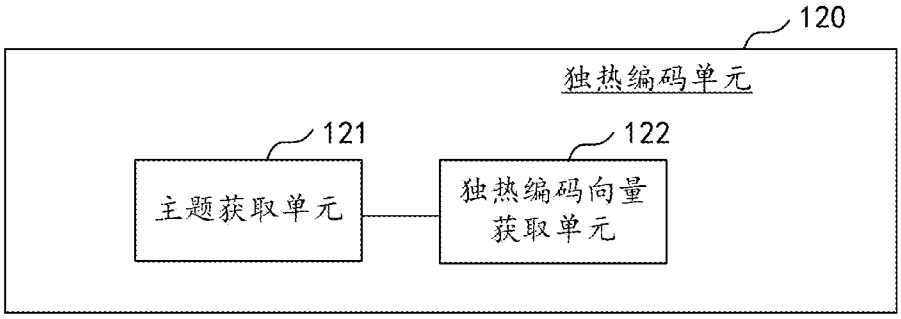
120
独热编码单元
121
主题获取单元
122
独热编码向量
获取单元

图 9

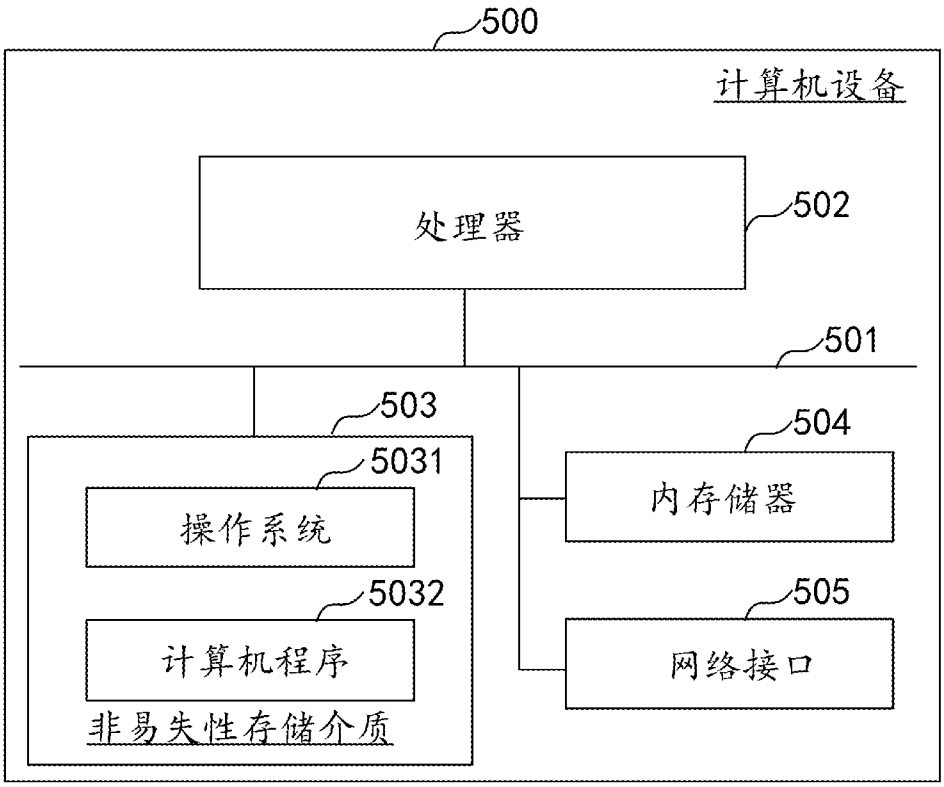
500
计算机设备
处理器
502
501
503
5031
操作系统
5032
计算机程序
非易失性存储介质
504
内存储器
505
网络接口

图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/118244

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/00(2019.01)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)

CNTXT; CNABS; CNKI; TWTXT; TWABS; USTXT; VEN; EPTXT; WOTXT: 文本, 文档, 分词, 向量, 潜在狄利克雷分布, 独热, 分类, text, word segmentation, vector, LDA, one-hot, classify

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109885826 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 14 June 2019 (2019-06-14) description, paragraphs 3-126	1-20
X	CN 108121699 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 05 June 2018 (2018-06-05) description, paragraphs 26-44 and 72-75, and figures 4 and 5	1-20
A	CN 108021546 A (BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.) 11 May 2018 (2018-05-11) entire document	1-20
A	CN 107220232 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 29 September 2017 (2017-09-29) entire document	1-20
A	CN 108595425 A (KUNMING UNIVERSITY OF SCIENCE AND TECHNOLOGY) 28 September 2018 (2018-09-28) entire document	1-20
A	US 2017116203 A1 (QBASE LLC) 27 April 2017 (2017-04-27) entire document	1-20

☐ 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
04 February 2020	**21 February 2020**

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.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/118244

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109885826 A	14 June 2019	None	
CN 108121699 A	05 June 2018	None	
CN 108021546 A	11 May 2018	None	
CN 107220232 A	29 September 2017	US 2018293507 A1	11 October 2018
CN 108595425 A	28 September 2018	None	
US 2017116203 A1	27 April 2017	None	

国际检索报告

国际申请号
PCT/CN2019/118244

A. 主题的分类

G06F 17/00(2019.01)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））

CNTXT;CNABS;CNKI;TWTXT;TWABS;USTXT;VEN;EPTXT;WOTXT:文本，文档，分词，向量，潜在狄利克雷分布，独热，分类，text, word segmentation, vector, LDA, one-hot, classify

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 109885826 A (平安科技深圳有限公司) 2019年 6月 14日 (2019-06-14) 说明书第3-126段	1-20
X	CN 108121699 A (北京百度网讯科技有限公司) 2018年 6月 5日 (2018-06-05) 说明书第26-44、72-75段，附图4-5	1-20
A	CN 108021546 A (北京嘀嘀无限科技发展有限公司) 2018年 5月 11日 (2018-05-11) 全文	1-20
A	CN 107220232 A (北京百度网讯科技有限公司) 2017年 9月 29日 (2017-09-29) 全文	1-20
A	CN 108595425 A (昆明理工大学) 2018年 9月 28日 (2018-09-28) 全文	1-20
A	US 2017116203 A1 (QBASE LLC) 2017年 4月 27日 (2017-04-27) 全文	1-20

☐ 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
2020年 2月 4日	2020年 2月 21日

ISA/CN的名称和邮寄地址	受权官员
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	杨牛
传真号 (86-10)62019451	电话号码 86-(20)-28950388

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/118244

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 109885826 A	2019年 6月 14日	无	
CN 108121699 A	2018年 6月 5日	无	
CN 108021546 A	2018年 5月 11日	无	
CN 107220232 A	2017年 9月 29日	US 2018293507 A1	2018年 10月 11日
CN 108595425 A	2018年 9月 28日	无	
US 2017116203 A1	2017年 4月 27日	无	