(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
2022 年 5 月 12 日 (12.05.2022)

WIPO | PCT

(10) 国际公布号
WO 2022/095370 A1

(51) 国际专利分类号:
G06F 16/33 (2019.01) *G06F 40/30* (2020.01)

(21) 国际申请号: PCT/CN2021/090952

(22) 国际申请日: 2021 年 4 月 29 日 (29.04.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202011228709.9 2020年11月6日 (06.11.2020) 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: 谢舒翼 (XIE, Shuyi); 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: 深圳中一联合知识产权代理有限公司 (SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市福田区园岭街道深南中路1014号报春大厦9楼（5号信箱）, Guangdong 518028 (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, IT, 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,

(54) **Title:** TEXT MATCHING METHOD AND APPARATUS, TERMINAL DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种文本匹配方法、装置、终端设备和存储介质

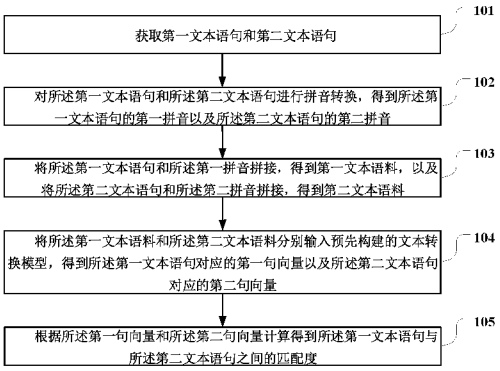


图 1

101 OBTAIN A FIRST TEXT STATEMENT AND A SECOND TEXT STATEMENT
102 PERFORM PINYIN CONVERSION ON THE FIRST TEXT STATEMENT AND THE SECOND TEXT STATEMENT TO OBTAIN A FIRST PINYIN OF THE FIRST TEXT STATEMENT AND A SECOND PINYIN OF THE SECOND TEXT STATEMENT
103 SPLICE THE FIRST TEXT STATEMENT AND THE FIRST PINYIN TO OBTAIN A FIRST TEXT CORPUS, AND SPLICE THE SECOND TEXT STATEMENT AND THE SECOND PINYIN TO OBTAIN A SECOND TEXT CORPUS
104 RESPECTIVELY INPUT THE FIRST TEXT CORPUS AND THE SECOND TEXT CORPUS INTO A PRECONSTRUCTED TEXT CONVERSION MODEL TO OBTAIN A FIRST SENTENCE VECTOR CORRESPONDING TO THE FIRST TEXT STATEMENT AND A SECOND SENTENCE VECTOR CORRESPONDING TO THE SECOND TEXT STATEMENT
105 CALCULATE A MATCHING DEGREE BETWEEN THE FIRST TEXT STATEMENT AND THE SECOND TEXT STATEMENT ACCORDING TO THE FIRST SENTENCE VECTOR AND THE SECOND SENTENCE VECTOR

(57) **Abstract:** A text matching method and apparatus, a terminal device, and a storage medium, suitable for the technical field of artificial intelligence. The method comprises: obtaining a first text statement and a second text statement (101); performing Pinyin conversion on the first text statement and the second text statement to obtain a first Pinyin of the first text statement and a second Pinyin of the second text statement (102); splicing the first text statement and the first Pinyin to obtain a first text corpus, and splicing the second text statement and the second Pinyin to obtain a second text corpus (103); respectively inputting the first text corpus and the

WO 2022/095370 A1

[见续页]

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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

second text corpus into a preconstructed text conversion model to obtain a first sentence vector corresponding to the first text statement and a second sentence vector corresponding to the second text statement (104); and calculating a matching degree between the first text statement and the second text statement according to the first sentence vector and the second sentence vector (105). By adoption of the method, the accuracy of text matching can be improved.

(57) 摘要：一种文本匹配方法、装置、终端设备和存储介质，适用于人工智能技术领域。该方法包括：获取第一文本语句和第二文本语句（101）；对第一文本语句和第二文本语句进行拼音转换，得到第一文本语句的第一拼音以及第二文本语句的第二拼音（102）；将第一文本语句和第一拼音拼接，得到第一文本语料，以及将第二文本语句和第二拼音拼接，得到第二文本语料（103）；将第一文本语料和第二文本语料分别输入预先构建的文本转换模型，得到第一文本语句对应的第一句向量以及第二文本语句对应的第二句向量（104）；根据第一句向量和第二句向量计算得到第一文本语句与第二文本语句之间的匹配度（105）。采用该方法能够提高文本匹配的准确率。

一种文本匹配方法、装置、终端设备和存储介质

[0001] 本申请要求于2020年11月06日提交中国专利局、申请号为202011228709.9，发明名称为“一种文本匹配方法、装置、终端设备和存储介质”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

[0002] 本申请属于人工智能技术领域，尤其涉及一种文本匹配方法、装置、终端设备和存储介质。

背景技术

[0003] 文本匹配是自然语言处理的一个核心问题，很多自然语言处理的任务都可以抽象成文本匹配问题，例如，信息检索系统、问答系统、对话系统、阅读理解系统等。随着语音系统的不断发展，越来越多的系统支持语音输入，发明人意识到，由于语音转换成文本存在一定的差异性及人工输入的误差性，使得输入的文本存在一定的差错，而传统的文本匹配模型没办法正确进行文本匹配，导致意图识别错误，给用户带来不好的体验。

发明概述

技术问题

[0004] 有鉴于此，本申请提出一种文本匹配方法、装置、终端设备和存储介质，能够提高文本匹配的准确率。

问题的解决方案

技术解决方案

[0005] 第一方面，本申请实施例提供了一种文本匹配方法，包括：

[0006] 获取第一文本语句和第二文本语句；

[0007] 对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

[0008] 将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

[0009] 将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

[0010] 根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[0011] 第二方面，本申请实施例提供了一种文本匹配装置，包括：

[0012] 文本语句获取模块，用于获取第一文本语句和第二文本语句；

[0013] 拼音转换模块，用于对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

[0014] 文本拼接模块，用于将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

[0015] 向量转换模块，用于将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

[0016] 匹配度计算模块，用于根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[0017] 第三方面，本申请实施例提供了一种终端设备，包括存储器、处理器以及存储在所述存储器中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如本申请实施例第一方面提出的文本匹配方法的步骤。

[0018] 第四方面，本申请实施例提供了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，所述计算机程序被处理器执行时实现如本申请实施例第一方面提出的文本匹配方法的步骤。

发明的有益效果

有益效果

[0019] 本申请实施例将拼音作为一个新特征，与文本语句进行拼接，形成新的语料输入到文本转换模型中，使模型学习到包含拼音信息的更多语义信息。这样，即使出现语音转换文本错误时，模型也可以根据学习到的拼音信息，对输入信息进行纠正，返回正确的匹配信息，从而能够提高文本匹配的准确率。

对附图的简要说明

附图说明

[0020] 图1是本申请实施例提供的一种文本匹配方法的第一个实施例的流程图；

[0021] 图2是本申请实施例提供的一种文本匹配方法的第二个实施例的流程图；

[0022] 图3是本申请实施例提供的一种文本匹配方法的第三个实施例的流程图；

[0023] 图4是本申请实施例提供的一种文本匹配装置的一个实施例的结构图；

[0024] 图5是本申请实施例提供的一种终端设备的示意图。

发明实施例

本发明的实施方式

[0025] 以下描述中，为了说明而不是为了限定，提出了诸如特定系统结构、技术之类的具体细节，以便透彻理解本申请实施例。然而，本领域的技术人员应当清楚，在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中，省略对众所周知的系统、装置、电路以及方法的详细说明，以免不必要的细节妨碍本申请的描述。另外，在本申请说明书和所附权利要求书的描述中，术语“第一”、“第二”、“第三”等仅用于区分描述，而不能理解为指示或暗示相对重要性。

[0026] 文本匹配是自然语言处理的一个核心问题，很多自然语言处理的任务都可以抽象成文本匹配问题，例如，信息检索系统、问答系统、对话系统、阅读理解系统等。由于文本匹配存在匹配对象差异巨大、优质训练样本难以大量获取及文本匹配的层次性等问题，针对不同的任务选取合适的匹配模型，提高匹配的准确率成为自然语言处理任务的重要挑战。

[0027] 目前，基于神经网络进行文本匹配的方式主要可以分为两大类：Sentence interaction（SI）类，该类模型可以更好的把握语义焦点，能对上下文重要性进行更好的建模，但忽视了句法、句间对照等全局性信息，无法由局部匹配信息刻画全局匹配信息；还有一类为Sentence encoding(Sentence representation)即SE类，它可以对文本进行预处理，构建索引，大幅度降低在线计算耗时、使用方便简单，且效果也比较好，孪生网络就是典型的SE结构，因此很多人都会采用孪生网络去进行文本匹配。然而，随着语音系统的不断发

展，越来越多的系统支持语音输入，由于语音转换成文本存在一定的差异性及人工输入的误差性，使得输入的文本存在一定的差错，传统的文本匹配模型没办法正确进行文本匹配，导致意图识别错误，给用户带来不好的体验。

[0028] 针对上述问题，本申请提出一种文本匹配方法、装置、终端设备和存储介质，能够提高文本匹配的准确率。

[0029] 本申请实施例提供的文本匹配方法可以应用于手机、平板电脑、可穿戴设备、车载设备、增强现实(augmented reality，AR)/虚拟现实(virtual reality，VR)设备、笔记本电脑、超级移动个人计算机(ultra-mobile personal computer，UMPC)、上网本、个人数字助理(personal digital assistant，PDA)、大屏电视等终端设备或者服务器上，本申请实施例对终端设备和服务器的具体类型不作任何限制。

[0030] 请参阅图1，本申请实施例中一种文本匹配方法的第一个实施例包括：

[0031] 101、获取第一文本语句和第二文本语句；

[0032] 首先，获取待匹配的第一文本语句和第二文本语句。第一文本语句和第二文本语句是需要执行文本匹配的两个汉字语句，每个文本语句都包含多个汉字。其中，第一文本语句和第二文本语句包含的汉字数量可以相同，也可以不同。

[0033] 102、对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

[0034] 在获取第一文本语句和第二文本语句后，对这两个文本语句分别进行拼音转换，得到第一文本语句的拼音以及第二文本语句的拼音。具体的，可以采用拼音转换器将输入的汉字转换成拼音，拼音转换器是可以将汉字转换成拼音的一种工具，比如可以采用python 中的xpinyin库进行转换，直接将句子中的每个汉字都转换成拼音，例如：“我爱人工智能”经过转换处理后得到“wo ai ren gong zhi neng”。

[0035] 103、将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

[0036] 在获得两个文本语句的拼音之后，分别将每个文本语句与对应的拼音进行拼接，得到两个新的文本语料。具体的，在拼接时，可以按照指定的字符拼接顺序

，将汉字和拼音依次拼接。例如，在将我们是一家人”转换为“wo men shi yi jia ren”后，可以按照汉字在前，拼音在后的顺序，拼接成“我们是一家人 wo men shi yi jia ren”；可以按照拼音在前，汉字在后的顺序，拼接成“wo men shi yi jia ren我们是一家人”；也可以按照“汉字-拼音-汉字”的顺序，拼接成“我wo们men是shi一yi家jia人 ren”。

[0037] 104、将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

[0038] 在得到第一文本语料和第二文本语料之后，将这两个文本语料分别输入某个预先构建的文本转换模型中。通过该文本转换模型，可以将文本语料转换为对应的句向量，也即可以将第一文本语料转换为第一句向量，将第二文本语料转换为第二句向量。

[0039] 具体的，该文本转换模型是预先训练完成的神经网络模型。可以包含Transformer编码层和pooling层，首先利用Transformer编码层对输入的文本预料进行编码，得到一系列数字表示的二维向量，类似于以下结构形式：

[0040] [[0 2 1 0],

[0041] [3 1 2 1],

[0042] [4 0 5 1]]

[0043] 然后，利用pooling层对编码后得到的二维向量执行文本特征提取操作。pooling层提取特征主要有max-pooling（最大池化层）和mean-pooling（均值池化层），max-pooling是对领域内特征点取最大值，而mean-pooling是对领域内特征点求平均，本申请优选采用的是max-pooling层进行特征选取，例如pooling层大小为2*2，则对上面向量从左到右，从上到下依次以2*2的窗口进行移动，抽取窗口内特征最大值，最后经过最大池化后变成以下二维向量结构：

[0044] [[3 2 2],

[0045] [4 5 5]]

[0046] 接着，对上述二维向量结构进行降维，得到一维向量（3 2 2 4 5 5），即为文本语句对应的句向量。

[0047] 105、根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[0048] 在获得第一文本语句对应的第一句向量和第二文本语句对应的第二句向量之后，即可根据这两个句向量计算得到第一文本语句和第二文本语句之间的匹配度，从而完成文本匹配的操作。

[0049] 进一步的，根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度，可以包括：

[0050] （1）计算所述第一句向量和所述第二句向量之间的余弦相似度；

[0051] （2）根据所述余弦相似度确定所述第一文本语句与所述第二文本语句之间的匹配度。

[0052] 余弦相似度通过测量两个向量的夹角的余弦值来度量它们之间的相似性，故通过计算第一句向量和第二句向量之间的余弦相似度，可以用于确定两个文本语句的匹配度。

[0053] 本申请实施例将拼音作为一个新特征，与文本语句进行拼接，形成新的语料输入到文本转换模型中，使模型学习到包含拼音信息的更多语义信息。这样，即使出现语音转换文本错误时，模型也可以根据学习到的拼音信息，对输入信息进行纠正，返回正确的匹配信息，从而能够提高文本匹配的准确率。

[0054] 请参阅图2，本申请实施例中一种文本匹配方法的第二个实施例包括：

[0055] 201、获取第一文本语句和第二文本语句；

[0056] 202、对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

[0057] 203、将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

[0058] 204、将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

[0059] 步骤201-204与步骤101-104相同，具体可参照步骤101-104的相关说明。

[0060] 205、基于所述第一句向量和所述第二句向量构建目标特征向量，所述目标特

征向量用于衡量所述第一句向量和所述第二句向量的匹配关系；

[0061] 在获得第一文本语句对应的第一句向量和第二文本语句对应的第二句向量之后，即可根据这两个句向量计算得到第一文本语句和第二文本语句之间的匹配度，具体可以通过构建目标特征向量的方式计算匹配度，该目标特征向量用于衡量第一句向量和第二句向量之间的匹配关系。

[0062] 具体的，基于所述第一句向量和所述第二句向量构建目标特征向量，可以包括：

[0063] （1）将所述第一句向量和所述第二句向量首尾相接，得到第一特征向量；

[0064] （2）将所述第一句向量和所述第二句向量对应位置的元素值相乘，得到第二特征向量；

[0065] （3）将所述第一句向量和所述第二句向量对应位置的元素值相减并取绝对值，得到第三特征向量；

[0066] （4）将所述第一特征向量、所述第二特征向量和所述第三特征向量降维度拼接，得到所述目标特征向量。

[0067] 以举例的方式说明上述步骤（1）至（4），假设第一句向量为u=[1 2 3 1 1 0]，第二句向量为v = [1 0 1 2 3 0]，则可以对u和v进行首尾相接，得到第一特征向量（u，v）=[[1 2 3 1 1 0]，[1 0 1 2 3 0]]；将u和v对应位置的元素值分别相乘，得到第二特征向量u*v=[1 0 3 2 3 0]；将u和v对应位置的元素值相减并取绝对值，得到第三特征向量|u-v|=[0 2 2 1 2 0]。然后，将这3个特征向量在-1维度进行拼接，得到以下目标特征向量：

[0068] [[1 1 1 0]

[0069] [2 0 0 2]

[0070] [3 1 3 2]

[0071] [1 2 2 1]

[0072] [1 3 2 2]

[0073] [0 0 0 0]]

[0074] 该目标特征向量可以用于衡量第一句向量u和第二句向量v 之间的匹配关系，可以通过该目标特征向量计算得到第一文本语句和第二文本语句之间的匹配度

。

[0075] 206、采用多个全连接层对所述目标特征向量进行处理，得到所述第一文本语句与所述第二文本语句之间的匹配度。

[0076] 在获得目标特征向量之后，可以将该目标特征向量输入多个全连接层进行处理，从而得到两个文本语句之间的匹配度。全连接层（fully connected layers）在整个神经网络中起到“分类器”的作用，如果说卷积层、池化层和激活函数层等操作是将原始数据映射到隐层特征空间的话，全连接层则起到将学到的“分布式特征表示”映射到样本标记空间的作用。

[0077] 具体的，全连接层中的每个神经元与其前一层的所有神经元进行全连接．可以整合卷积层或者池化层中具有类别区分性的局部信息．也即可以将目标特征向量具有的文本特征整合成一个用于评价两个文本之间相似程度的数值参数，即匹配度。

[0078] 进一步的，在得到所述第一文本语句与所述第二文本语句之间的匹配度之后，还可以包括：

[0079] 采用softmax层对所述匹配度进行处理，得到所述第一文本语句与所述第二文本语句是否匹配的结果。

[0080] 最后，还可以采用softmax层完成分类，也即根据输入的匹配度获得一个第一文本语句和第二文本语句是否匹配的结果。例如，若输入的匹配度大于某个阈值，则输出第一文本语句和第二文本语句匹配的结果；否则输出第一文本语句和第二文本语句不匹配的结果。

[0081] 本申请实施例将拼音作为一个新特征，与汉字语料进行拼接，形成新的语料输入到文本转换模型中，使模型学习到包含拼音信息的更多语义信息。这样，即使出现语音转换文本错误时，模型也可以根据学习到的拼音信息，对输入信息进行纠正，返回正确的匹配信息，从而能够提高文本匹配的准确率。与本申请的第一个实施例相比，本实施例提出一种根据第一句向量和第二句向量计算得到两个文本语句之间的匹配度的具体方式。

[0082] 请参阅图3，本申请实施例中一种文本匹配方法的第三个实施例包括：

[0083] 301、获取第一文本语句和第二文本语句；

[0084] 302、分别检测所述第一文本语句的长度和所述第二文本语句的长度；

[0085] 本申请实施例在获取到两个文本语句之后，会分别检测这两个文本语句的长度，具体可以是检测文本语句具有的字符串的个数。

[0086] 303、若所述第一文本语句的长度超过预设长度值，则对所述第一文本语句进行截断，得到长度为所述预设长度值的第一文本语句；

[0087] 304、若所述第二文本语句的长度超过所述预设长度值，则对所述第二文本语句进行截断，得到长度为所述预设长度值的第二文本语句；

[0088] 305、若所述第一文本语句的长度小于所述预设长度值，则采用指定字符对所述第一文本语句进行填充，得到长度为所述预设长度值的第一文本语句；

[0089] 306、若所述第二文本语句的长度小于所述预设长度值，则采用所述指定字符对所述第二文本语句进行填充，得到长度为所述预设长度值的第二文本语句；

[0090] 后续步骤中采用的文本转换模型，其针对不同长度文本语句的处理效果是有区别的，因此在将文本语句输入模型之前，可以先将文本语句截断或者填充至某个指定长度，以提高后续模型的处理效果。比如，可以设置语句长度值为60个字符，当某个文本语句的长度超过60个字符时，按照设定的方式对该文本语句进行截断，使得该文本语句的长度为60个字符。当某个文本语句的长度不足60个字符时，可以采用某个指定字符（比如字符“0”）对该文本语句进行填充，使得该文本语句的长度为60个字符。

[0091] 另外，也可以设置一个优选的文本长度范围区间，比如[30，60]，其中30是文本语句的长度下限，60是文本语句的长度上限。当检测到某个文本语句的长度超过60时，对该文本语句进行截断，使得该文本语句的长度为60；当某个文本语句的长度小于30时，采用指定字符对该文本语句进行填充，使得该文本语句的长度为30。

[0092] 307、对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

[0093] 308、将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

[0094] 309、将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换

模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

[0095] 310、根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[0096] 步骤307-310与步骤102-105相同，具体可参照步骤102-105的相关说明。

[0097] 本申请实施例将拼音作为一个新特征，与汉字语料进行拼接，形成新的语料输入到文本转换模型中，使模型学习到包含拼音信息的更多语义信息。这样，即使出现语音转换文本错误时，模型也可以根据学习到的拼音信息，对输入信息进行纠正，返回正确的匹配信息，从而能够提高文本匹配的准确率。另外，本申请实施例在将文本语句输入模型之前，会先将文本语句截断或者填充至某个指定长度，能够提高后续模型的处理效果。

[0098] 应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

[0099] 对应于上文实施例所述的文本匹配方法，图4示出了本申请实施例提供的一种文本匹配装置的结构框图，为了便于说明，仅示出了与本申请实施例相关的部分。

[0100] 参照图4，该装置包括：

[0101] 文本语句获取模块401，用于获取第一文本语句和第二文本语句；

[0102] 拼音转换模块402，用于对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

[0103] 文本拼接模块403，用于将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

[0104] 向量转换模块404，用于将所述第一文本语料和所述第二文本语料分别输入预先构建的文本向量转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

[0105] 匹配度计算模块405，用于根据所述第一句向量和所述第二句向量计算得到所

述第一文本语句与所述第二文本语句之间的匹配度。

[0106] 进一步的，所述匹配度计算模块可以包括：

[0107] 余弦相似度计算单元，用于计算所述第一句向量和所述第二句向量之间的余弦相似度；

[0108] 匹配度计算单元，用于根据所述余弦相似度确定所述第一文本语句与所述第二文本语句之间的匹配度。

[0109] 进一步的，所述匹配度计算模块可以包括：

[0110] 特征向量构建单元，用于基于所述第一句向量和所述第二句向量构建目标特征向量，所述目标特征向量用于衡量所述第一句向量和所述第二句向量的匹配关系；

[0111] 向量处理单元，用于采用多个全连接层对所述目标特征向量进行处理，得到所述第一文本语句与所述第二文本语句之间的匹配度。

[0112] 更进一步的，所述特征向量构建单元可以包括：

[0113] 向量拼接子单元，用于将所述第一句向量和所述第二句向量首尾相接，得到第一特征向量；

[0114] 向量相乘子单元，用于将所述第一句向量和所述第二句向量对应位置的元素值相乘，得到第二特征向量；

[0115] 向量作差子单元，用于将所述第一句向量和所述第二句向量对应位置的元素值相减并取绝对值，得到第三特征向量；

[0116] 向量降维子单元，用于将所述第一特征向量、所述第二特征向量和所述第三特征向量降维度拼接，得到所述目标特征向量。

[0117] 进一步的，所述匹配度计算模块还可以包括：

[0118] 分类单元，用于采用softmax层对所述匹配度进行分类处理，得到所述第一文本语句与所述第二文本语句是否匹配的结果。

[0119] 进一步的，所述文本匹配装置还可以包括：

[0120] 文本长度检测模块，用于分别检测所述第一文本语句的长度和所述第二文本语句的长度；

[0121] 第一文本截断模块，用于若所述第一文本语句的长度超过预设长度值，则对所

述第一文本语句进行截断，得到长度为所述预设长度值的第一文本语句；

[0122] 第二文本截断模块，用于若所述第二文本语句的长度超过所述预设长度值，则对所述第二文本语句进行截断，得到长度为所述预设长度值的第二文本语句。

[0123] 进一步的，所述文本匹配装置还可以包括：

[0124] 第一文本填充模块，用于若所述第一文本语句的长度小于所述预设长度值，则采用指定字符对所述第一文本语句进行填充，得到长度为所述预设长度值的第一文本语句；

[0125] 第二文本填充模块，用于若所述第二文本语句的长度小于所述预设长度值，则采用所述指定字符对所述第二文本语句进行填充，得到长度为所述预设长度值的第二文本语句。

[0126] 本申请实施例还提供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可读指令，所述计算机可读指令被处理器执行时实现如图1至图3表示的任意一种文本匹配方法的步骤。其中，所述计算机可读存储介质可以是非易失性，也可以是易失性。

[0127] 本申请实施例还提供一种计算机程序产品，当该计算机程序产品在服务器上运行时，使得服务器执行实现如图1至图3表示的任意一种文本匹配方法的步骤。

[0128] 图5是本申请一实施例提供的终端设备的示意图。如图5所示，该实施例的终端设备5包括：处理器50、存储器51以及存储在所述存储器51中并可在所述处理器50上运行的计算机可读指令52。所述处理器50执行所述计算机可读指令52时实现上述各个文本匹配方法实施例中的步骤，例如图1所示的步骤101至105。或者，所述处理器50执行所述计算机可读指令52时实现上述各装置实施例中各模块/单元的功能，例如图4所示模块401至405的功能。

[0129] 示例性的，所述计算机可读指令52可以被分割成一个或多个模块/单元，所述一个或者多个模块/单元被存储在所述存储器51中，并由所述处理器50执行，以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机可读指令段，该指令段用于描述所述计算机可读指令52在所述终端设备5中的执行过程。

[0130] 所述终端设备5可以是智能手机、笔记本、掌上电脑及云端终端设备等计算设

备。所述终端设备5可包括，但不仅限于，处理器50、存储器51。本领域技术人员可以理解，图5仅仅是终端设备5的示例，并不构成对终端设备5的限定，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，例如所述终端设备5还可以包括输入输出设备、网络接入设备、总线等。

[0131] 所述处理器50可以是中央处理单元(Central Processing Unit，CPU)，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。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。

[0132] 所述存储器51可以是所述终端设备5的内部存储单元，例如终端设备5的硬盘或内存。所述存储器51也可以是所述终端设备5的外部存储设备，例如所述终端设备5上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。进一步地，所述存储器51还可以既包括所述终端设备5的内部存储单元也包括外部存储设备。所述存储器51用于存储所述计算机可读指令以及所述终端设备所需的其他程序和数据。所述存储器51还可以用于暂时地存储已经输出或者将要输出的数据。

[0133] 需要说明的是，上述装置/单元之间的信息交互、执行过程等内容，由于与本申请方法实施例基于同一构思，其具体功能及带来的技术效果，具体可参见方法实施例部分，此处不再赘述。

[0134] 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，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中，上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。另外，各功能单元、模块的具体名称也只是为了便于相互区分，并不用于限制本申请的保护范围。上述系

统中单元、模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0135] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，该计算机程序在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质至少可以包括：能够将计算机程序代码携带到拍照装置/终端设备的任何实体或装置、记录介质、计算机存储器、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、电载波信号、电信信号以及软件分发介质。例如U盘、移动硬盘、磁碟或者光盘等。

[0136] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述或记载的部分，可以参见其它实施例的相关描述。

[0137] 以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权利要求书

[权利要求 1] 一种文本匹配方法，其中，包括：

获取第一文本语句和第二文本语句；

对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 2] 如权利要求1所述的文本匹配方法，其中，根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度，包括：

计算所述第一句向量和所述第二句向量之间的余弦相似度；

根据所述余弦相似度确定所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 3] 如权利要求1文本匹配方法，其中，根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度，包括：

基于所述第一句向量和所述第二句向量构建目标特征向量，所述目标特征向量用于衡量所述第一句向量和所述第二句向量的匹配关系；

采用多个全连接层对所述目标特征向量进行处理，得到所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 4] 如权利要求3所述的文本匹配方法，其中，基于所述第一句向量和所述第二句向量构建目标特征向量，包括：

将所述第一句向量和所述第二句向量首尾相接，得到第一特征向量；

将所述第一句向量和所述第二句向量对应位置的元素值相乘，得到第二特征向量；

将所述第一句向量和所述第二句向量对应位置的元素值相减并取绝对值，得到第三特征向量；

将所述第一特征向量、所述第二特征向量和所述第三特征向量降维度拼接，得到所述目标特征向量。

[权利要求 5] 如权利要求3所述的文本匹配方法，其中，在得到所述第一文本语句与所述第二文本语句之间的匹配度之后，还包括：

采用softmax层对所述匹配度进行分类处理，得到所述第一文本语句与所述第二文本语句是否匹配的结果。

[权利要求 6] 如权利要求1至5任一项所述的文本匹配方法，其中，在获取第一文本语句和第二文本语句之后，对所述第一文本语句和所述第二文本语句进行拼音转换之前，还包括：

分别检测所述第一文本语句的长度和所述第二文本语句的长度；

若所述第一文本语句的长度超过预设长度值，则对所述第一文本语句进行截断，得到长度为所述预设长度值的第一文本语句；

若所述第二文本语句的长度超过所述预设长度值，则对所述第二文本语句进行截断，得到长度为所述预设长度值的第二文本语句。

[权利要求 7] 如权利要求6所述的文本匹配方法，其中，在分别检测所述第一文本语句的长度和所述第二文本语句的长度之后，还包括：

若所述第一文本语句的长度小于所述预设长度值，则采用指定字符对所述第一文本语句进行填充，得到长度为所述预设长度值的第一文本语句；

若所述第二文本语句的长度小于所述预设长度值，则采用所述指定字符对所述第二文本语句进行填充，得到长度为所述预设长度值的第二文本语句。

[权利要求 8] 一种文本匹配装置，其中，包括：

文本语句获取模块，用于获取第一文本语句和第二文本语句；

拼音转换模块，用于对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

文本拼接模块，用于将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

向量转换模块，用于将所述第一文本语料和所述第二文本语料分别输入预先构建的文本向量转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

匹配度计算模块，用于根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 9] 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其中，所述处理器执行所述计算机程序时实现如下步骤：

获取第一文本语句和第二文本语句；

对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 10] 如权利要求9所述的终端设备，其中，根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度，包括：

计算所述第一句向量和所述第二句向量之间的余弦相似度；

根据所述余弦相似度确定所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 11] 如权利要求9所述的终端设备，其中，根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度，包括：

基于所述第一句向量和所述第二句向量构建目标特征向量，所述目标特征向量用于衡量所述第一句向量和所述第二句向量的匹配关系；

采用多个全连接层对所述目标特征向量进行处理，得到所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 12] 如权利要求11所述的终端设备，其中，基于所述第一句向量和所述第二句向量构建目标特征向量，包括：

将所述第一句向量和所述第二句向量首尾相接，得到第一特征向量；

将所述第一句向量和所述第二句向量对应位置的元素值相乘，得到第二特征向量；

将所述第一句向量和所述第二句向量对应位置的元素值相减并取绝对值，得到第三特征向量；

将所述第一特征向量、所述第二特征向量和所述第三特征向量降维度拼接，得到所述目标特征向量。

[权利要求 13] 如权利要求11所述的终端设备，其中，在得到所述第一文本语句与所述第二文本语句之间的匹配度之后，还包括：

采用softmax层对所述匹配度进行分类处理，得到所述第一文本语句与所述第二文本语句是否匹配的结果。

[权利要求 14] 如权利要求9至13任一项所述的终端设备，其中，在获取第一文本语句和第二文本语句之后，对所述第一文本语句和所述第二文本语句进行拼音转换之前，还包括：

分别检测所述第一文本语句的长度和所述第二文本语句的长度；

若所述第一文本语句的长度超过预设长度值，则对所述第一文本语句进行截断，得到长度为所述预设长度值的第一文本语句；

若所述第二文本语句的长度超过所述预设长度值，则对所述第二文本语句进行截断，得到长度为所述预设长度值的第二文本语句。

[权利要求 15] 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，其中，所述计算机程序被处理器执行时实现如下步骤：

获取第一文本语句和第二文本语句；

对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音；

将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料；

将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量；

根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 16] 如权利要求15所述的计算机可读存储介质，其中，根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度，包括：

计算所述第一句向量和所述第二句向量之间的余弦相似度；

根据所述余弦相似度确定所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 17] 如权利要求15所述的计算机可读存储介质，其中，根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度，包括：

基于所述第一句向量和所述第二句向量构建目标特征向量，所述目标特征向量用于衡量所述第一句向量和所述第二句向量的匹配关系；

采用多个全连接层对所述目标特征向量进行处理，得到所述第一文本语句与所述第二文本语句之间的匹配度。

[权利要求 18] 如权利要求17所述的计算机可读存储介质，其中，基于所述第一句向

量和所述第二句向量构建目标特征向量，包括：

将所述第一句向量和所述第二句向量首尾相接，得到第一特征向量；

将所述第一句向量和所述第二句向量对应位置的元素值相乘，得到第二特征向量；

将所述第一句向量和所述第二句向量对应位置的元素值相减并取绝对值，得到第三特征向量；

将所述第一特征向量、所述第二特征向量和所述第三特征向量降维度拼接，得到所述目标特征向量。

[权利要求 19] 如权利要求17所述的计算机可读存储介质，其中，在得到所述第一文本语句与所述第二文本语句之间的匹配度之后，还包括：

采用softmax层对所述匹配度进行分类处理，得到所述第一文本语句与所述第二文本语句是否匹配的结果。

[权利要求 20] 如权利要求15至19任一项所述的终端设备，其中，在获取第一文本语句和第二文本语句之后，对所述第一文本语句和所述第二文本语句进行拼音转换之前，还包括：

分别检测所述第一文本语句的长度和所述第二文本语句的长度；

若所述第一文本语句的长度超过预设长度值，则对所述第一文本语句进行截断，得到长度为所述预设长度值的第一文本语句；

若所述第二文本语句的长度超过所述预设长度值，则对所述第二文本语句进行截断，得到长度为所述预设长度值的第二文本语句。

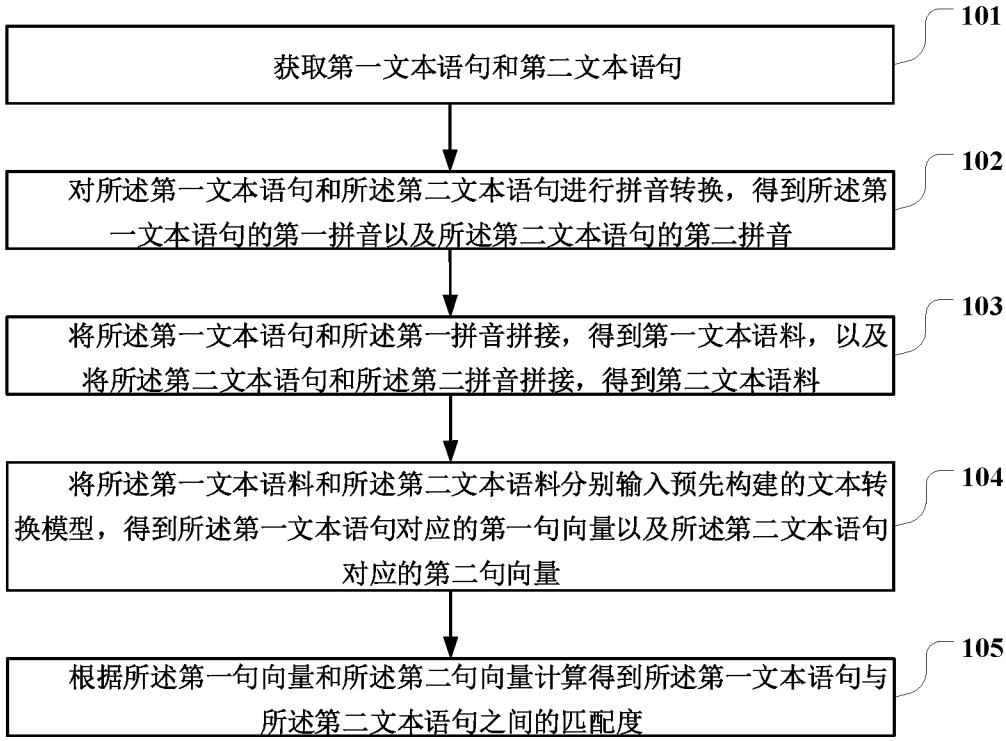

101
获取第一文本语句和第二文本语句
102
对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音
103
将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料
104
将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量
105
根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度

图 1

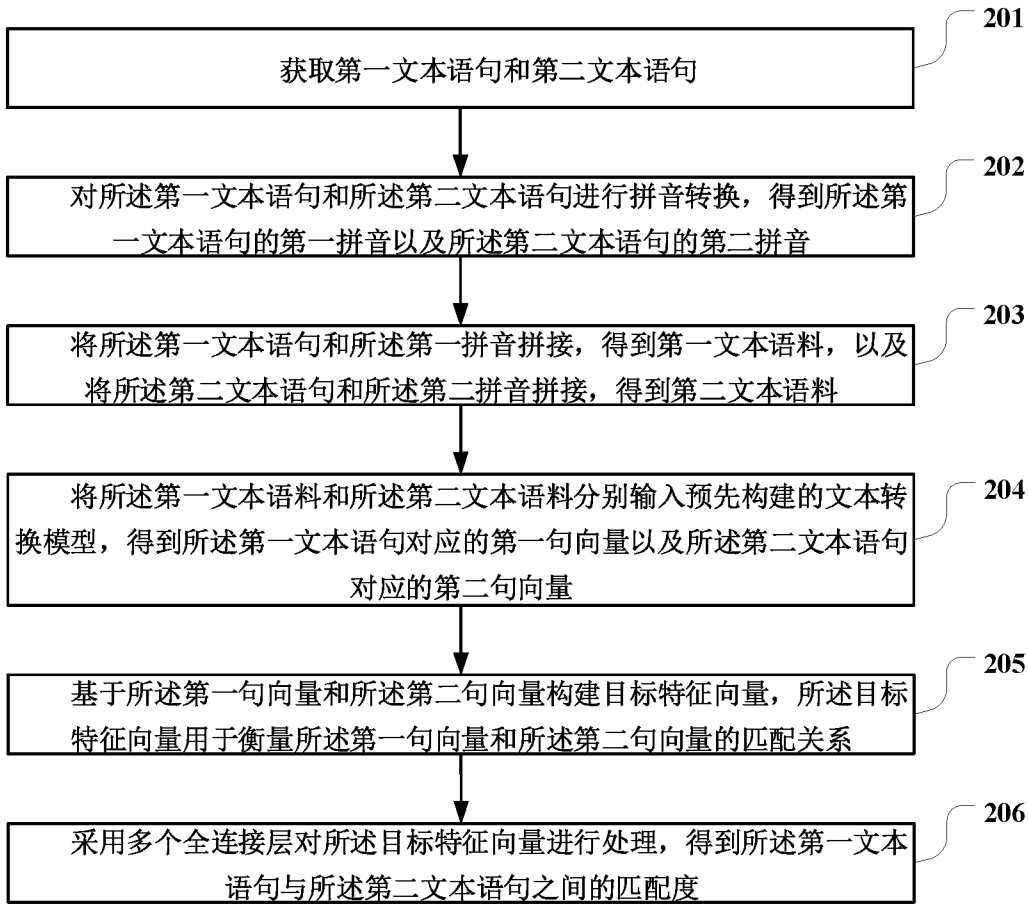

201
获取第一文本语句和第二文本语句
202
对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音
203
将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料
204
将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量
205
基于所述第一句向量和所述第二句向量构建目标特征向量，所述目标特征向量用于衡量所述第一句向量和所述第二句向量的匹配关系
206
采用多个全连接层对所述目标特征向量进行处理，得到所述第一文本语句与所述第二文本语句之间的匹配度

图 2

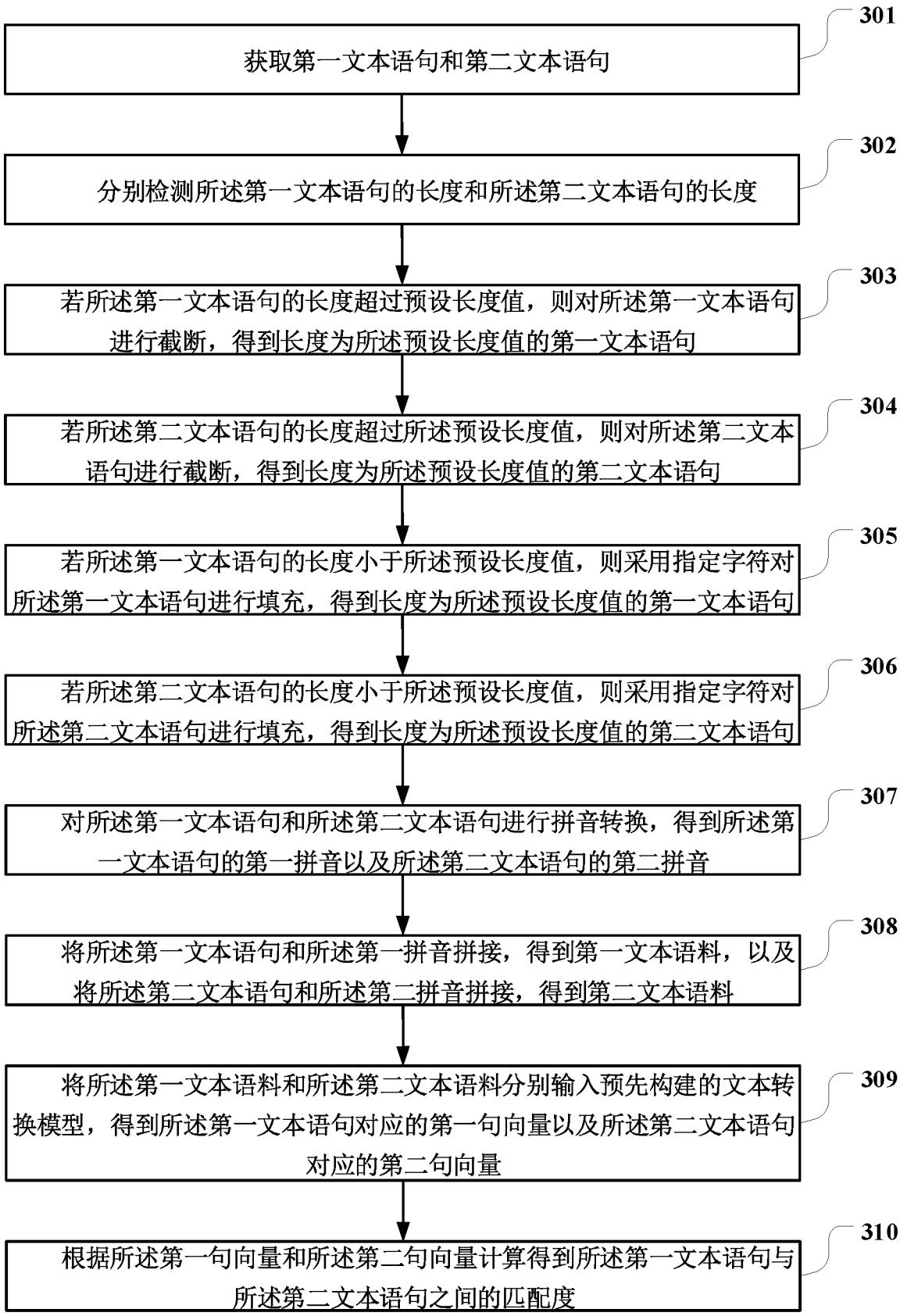

301
获取第一文本语句和第二文本语句
302
分别检测所述第一文本语句的长度和所述第二文本语句的长度
303
若所述第一文本语句的长度超过预设长度值，则对所述第一文本语句进行截断，得到长度为所述预设长度值的第一文本语句
304
若所述第二文本语句的长度超过所述预设长度值，则对所述第二文本语句进行截断，得到长度为所述预设长度值的第二文本语句
305
若所述第一文本语句的长度小于所述预设长度值，则采用指定字符对所述第一文本语句进行填充，得到长度为所述预设长度值的第一文本语句
306
若所述第二文本语句的长度小于所述预设长度值，则采用指定字符对所述第二文本语句进行填充，得到长度为所述预设长度值的第二文本语句
307
对所述第一文本语句和所述第二文本语句进行拼音转换，得到所述第一文本语句的第一拼音以及所述第二文本语句的第二拼音
308
将所述第一文本语句和所述第一拼音拼接，得到第一文本语料，以及将所述第二文本语句和所述第二拼音拼接，得到第二文本语料
309
将所述第一文本语料和所述第二文本语料分别输入预先构建的文本转换模型，得到所述第一文本语句对应的第一句向量以及所述第二文本语句对应的第二句向量
310
根据所述第一句向量和所述第二句向量计算得到所述第一文本语句与所述第二文本语句之间的匹配度

图 3

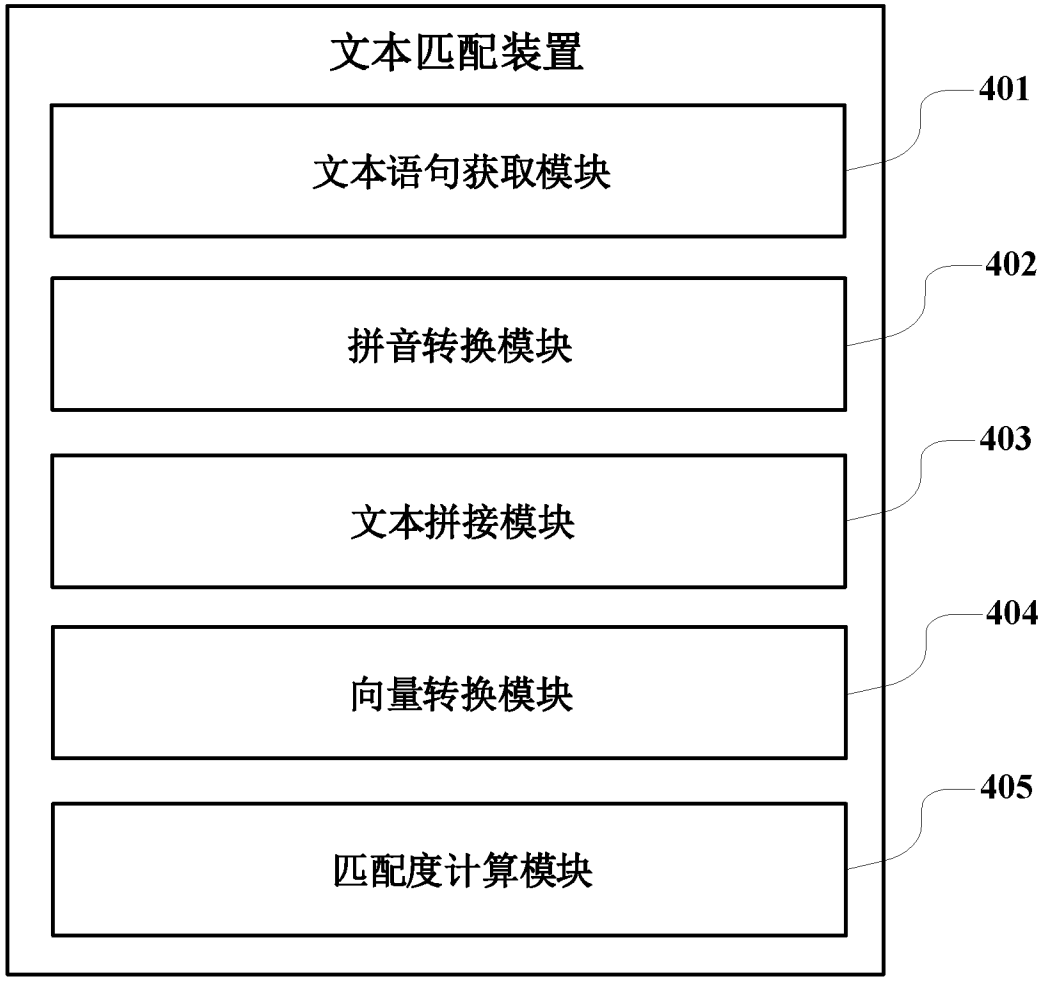

文本匹配装置
401
文本语句获取模块
402
拼音转换模块
403
文本拼接模块
404
向量转换模块
405
匹配度计算模块

图 4

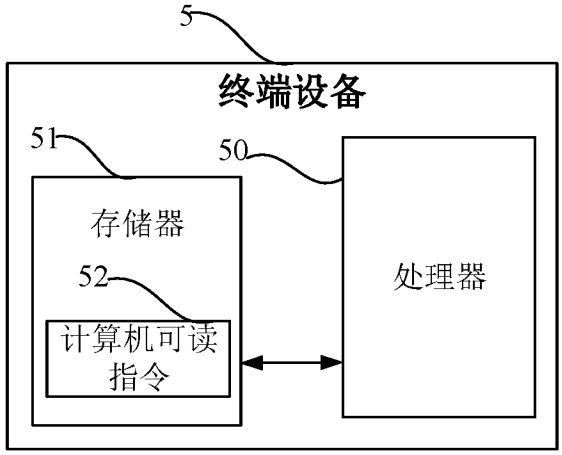

5
终端设备
51
存储器
52
计算机可读指令
50
处理器

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2021/090952

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/33(2019.01)i; G06F 40/30(2020.01)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: 文本, 匹配, 相似度, 人工智能, 语句, 句子, 拼音, 向量, 余弦, 准确率, text, matching, similarity, artificial intelligence, sentence, pinyin, vector, cosine, accuracy

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112182167 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 05 January 2021 (2021-01-05) claims 1-10, description paragraphs [0003]-[0042]	1-20
X	CN 111221960 A (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) 02 June 2020 (2020-06-02) description, paragraphs [0003]-[0031], [0120]	1-20
X	CN 109460461 A (SUZHOU AISPEECH INFORMATION TECHNOLOGY CO., LTD.) 12 March 2019 (2019-03-12) description, paragraphs [0005]-[0025] and [0119]-[0123]	1-20
A	CN 111159339 A (BEIJING ASIAINFO TECHNOLOGIES LIMITED) 15 May 2020 (2020-05-15) entire document	1-20
A	CN 106815197 A (BEIJING GRIDSUM TECHNOLOGY CO., LTD.) 09 June 2017 (2017-06-09) 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
02 June 2021	**25 June 2021**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China** Facsimile No. **(86-10)62019451**	Authorized officer 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 109658938 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 19 April 2019 (2019-04-19) entire document	1-20
A	US 2016217129 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 28 July 2016 (2016-07-28) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2021/090952

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 112182167 A	05 January 2021	None	
CN 111221960 A	02 June 2020	None	
CN 109460461 A	12 March 2019	None	
CN 111159339 A	15 May 2020	None	
CN 106815197 A	09 June 2017	None	
CN 109658938 A	19 April 2019	US 2020184978 A1	11 June 2020
US 2016217129 A1	28 July 2016	WO 2016082406 A1	02 June 2016
		CN 105701120 A	22 June 2016
		US 2020065388 A1	27 February 2020
		EP 3065068 A1	07 September 2016

国际检索报告

国际申请号
PCT/CN2021/090952

A. 主题的分类

G06F 16/33(2019.01)i; G06F 40/30(2020.01)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: 文本，匹配，相似度，人工智能，语句，句子，拼音，向量，余弦，准确率，text, matching, similarity, artificial intelligence, sentence, pinyin, vector, cosine, accuracy

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 112182167 A (平安科技深圳有限公司) 2021年 1月 5日 (2021-01-05) 权利要求1-10，说明书第[0003]-[0042]段	1-20
X	CN 111221960 A (支付宝杭州信息技术有限公司) 2020年 6月 2日 (2020-06-02) 说明书第[0003]-[0031]，[0120]段	1-20
X	CN 109460461 A (苏州思必驰信息科技有限公司) 2019年 3月 12日 (2019-03-12) 说明书第[0005]-[0025]，[0119]-[0123]段	1-20
A	CN 111159339 A (北京亚信数据有限公司) 2020年 5月 15日 (2020-05-15) 全文	1-20
A	CN 106815197 A (北京国双科技有限公司) 2017年 6月 9日 (2017-06-09) 全文	1-20
A	CN 109658938 A (百度在线网络技术北京有限公司) 2019年 4月 19日 (2019-04-19) 全文	1-20
A	US 2016217129 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 2016年 7月 28日 (2016-07-28) 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
2021年 6月 2日	2021年 6月 25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	石爽
传真号 (86-10)62019451	电话号码 86-(10)-53961413

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2021/090952

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	112182167	A	2021年 1月 5日	无			
CN	111221960	A	2020年 6月 2日	无			
CN	109460461	A	2019年 3月 12日	无			
CN	111159339	A	2020年 5月 15日	无			
CN	106815197	A	2017年 6月 9日	无			
CN	109658938	A	2019年 4月 19日	US	2020184978	A1	2020年 6月 11日
US	2016217129	A1	2016年 7月 28日	WO	2016082406	A1	2016年 6月 2日
				CN	105701120	A	2016年 6月 22日
				US	2020065388	A1	2020年 2月 27日
				EP	3065068	A1	2016年 9月 7日