(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107870964 A
(43)申请公布日 2018.04.03

(21)申请号 201710628098.9

(22)申请日 2017.07.28

(71)申请人 北京中科汇联科技股份有限公司
地址 100193 北京市海淀区东北旺西路8号9号楼二区305

(72)发明人 游世学 杜新凯

(74)专利代理机构 北京庆峰财智知识产权代理事务所(普通合伙) 11417
代理人 李文军

(51)Int.Cl.
G06F 17/30(2006.01)

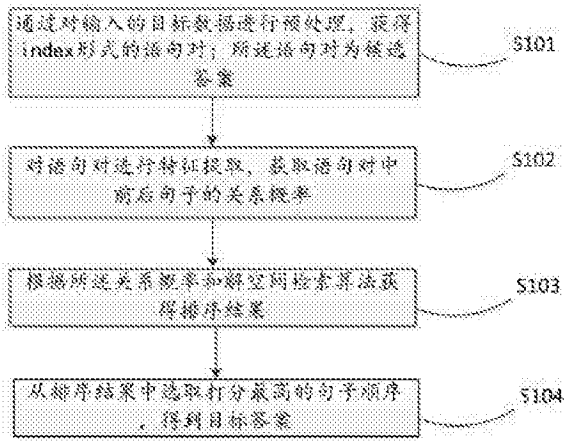
权利要求书2页 说明书8页 附图5页

(54)发明名称

一种应用于答案融合系统的语句排序方法及系统

(57)摘要

本发明公开了一种应用于答案融合系统的语句排序方法及系统，所述方法包括：通过对输入的目标数据进行预处理，获得index形式的语句对；所述语句对为候选答案；对语句对进行特征提取，获取语句对中前后句子的关系概率，包括：判断语句对中句子是否为前后顺序的关系，利用基于注意力机制的seq2seq模型，对语句对进行特征提取，将提取得到的特征输入到前馈神经网络中，得到句子前后顺序概率；根据所述关系概率和解空间检索算法获得排序结果，检索算法求解的目标函数是 $Score(s,o)=\sum_{i=1}^{k}\sum_{j=i+1}^{k}score(s,o,i,j)$；根据所述排序结果得到目标答案。本发明能够实现从不同答案中抽取的句子进行排序，从而使得获取的目标答案更加通顺，利于理解，且具有较强的可读性；从而提升了用户体验。



CN 107870964 A

1.一种应用于答案融合系统的语句排序方法,其特征在于,包括:

通过对输入的目标数据进行预处理,获得index形式的语句对;所述语句对为候选答案;

对语句对进行特征提取,获取语句对中前后句子的关系概率,包括:

判断语句对中句子是否为前后顺序的关系,利用基于注意力机制的seq2seq模型,对语句对进行特征提取,

将提取得到的特征输入到前馈神经网络中,得到所述语句对中句子的关系概率;

根据所述关系概率和解空间检索算法获得排序结果,所述解空间检索算法求解的目标函数是$Score(s,o)=\sum_{i=1}^{n}\sum_{j=i+1}^{n}score(s,o,i,j)$,s为所有目标语句组成的句子集,o为语句对构成的句子顺序,i、j为句子标号,表示第i句话,第j句话;排序结果为多个元素,每个元素包含句子顺序o以及该顺序的打分Score(s,o);

从排序结果中选取打分最高的句子顺序,得到目标答案。

2.根据权利要求1所述的方法,其特征在于,所述通过对输入的目标数据进行预处理,获得index形式的语句对,包括:

根据分词工具对所有的目标数据进行分词,所述目标数据中包括多个目标语句;

将多个目标语句两两构建成语句对;

通过语句转换功能将所述语句对转换为index形式的序列。

3.根据权利要求1所述的方法,其特征在于,所述采用基于注意力机制的seq2seq模型,对语句对进行特征提取,包括:

将index形式的语句对输入到seq2seq模型中;

将两个LSTM模型进行拼接得到所述seq2seq模型;

将attention机制应用到seq2seq模型上,通过前一LSTM模块LSTM1的隐藏层输出及后一LSTM模块LSTM2的每个节点的隐藏层输出,计算注意力分配权重,更新中间权重;

根据更新得到的中间权重,结合LSTM1隐藏层输出及LSTM2最后节点的隐藏层输出计算得到特征值。

4.根据权利要求3所述的方法,其特征在于,所述将index形式的语句对输入到seq2seq模型中,包括:

将转换得到的index形式的语句对通过embedding层的转换,使其以词向量特征的形式进行表示;

将语句对中分词得到的每个词输入到LSTM模块的节点中;

对每个节点进行计算得到所述LSTM模块的隐藏层输出。

5.根据权利要求3所述的方法,其特征在于,在两个LSTM模型中,前一个LSTM模型的隐藏层输出作为后一LSTM隐藏层输入。

6.一种应用于答案融合系统中的语句排序系统,其特征在于,包括:

预处理模块,用于对输入的目标数据进行预处理,得到index形式的语句对,所述语句对为候选答案;

关系概率获取模块,用于对语句对进行特征提取,获取语句对中的前后句子的关系概率,包括:

特征提取单元，用于判断语句对中的句子是否为前后顺序的关系，采用基于注意力机制的seq2seq模型，对语句对进行特征提取；

关系概率获取单元，用于将提取得到的特征输入到前馈神经网络中，得到所述语句对中句子的关系概率；

排序结果获取模块，用于根据所述关系概率和解空间检索算法得到答案语句的排序结果；解空间检索算法求解的目标函数是$Score(s,o)=\sum_{i=1}^{k}\sum_{j=i+1}^{k}score(s,o,i,j)$；s为所有目标语句组成的句子集，o为语句对构成的句子顺序，i、j为句子标号，表示第i句话，第j句话；排序结果为多个元素，每个元素包含句子顺序o以及该顺序的打分Score(s,o)；

答案获取模块，用于排序结果中选取打分最高的句子顺序，得到目标答案。

7.根据权利要求1所述的系统，其特征在于，所述预处理模块，包括：

分词单元，用于根据分词工具对所有的目标数据进行分词，所述目标数据中包括多个目标语句；

语句对构建单元，用于将多个目标语句两两构建成语句对；

语句转换单元，用于通过语句转换功能将所述语句对转换为index形式的序列。

8.根据权利要求1所述的系统，其特征在于，所述特征提取单元，包括：

转换语句输入单元，用于将index形式的语句对输入到seq2seq模型中；

预设模型拼接单元，用于将两个LSTM模型进行拼接得到所述seq2seq模型；

权重获取单元，用于将attention机制应用到seq2seq模型上，通过前一LSTM模块LSTM1的隐藏层输出及后一LSTM模块LSTM2的每个节点的隐藏层输出，计算注意力分配权重，更新中间权重；

特征获取单元，用于根据更新得到的中间权重，结合LSTM1隐藏层输出及LSTM2最后节点的隐藏层输出计算得到特征值。

9.根据权利要求8所述的方法，其特征在于，所述转换语句输入单元，包括：

词向量形式子单元，用于将index形式的语句对通过embedding层的转换，使其以词向量特征的形式进行表示；

词语输入节点子单元，用于将语句对中分词得到的每个词输入到LSTM模块的节点中；

隐藏层输出子单元，用于对每个节点进行计算得到所述LSTM模块的隐藏层输出。

10.根据权利要求8所述的方法，其特征在于，在两个LSTM模型中，前一个LSTM模型的隐藏层输出作为后一LSTM模型隐藏层输入。

一种应用于答案融合系统的语句排序方法及系统

技术领域

[0001] 本发明涉及数据处理技术领域，尤其涉及一种应用于答案融合系统的语句排序方法及系统。

背景技术

[0002] 答案融合系统是问答系统中的一个部分，用来构建候选答案库。答案融合系统利用如百度知道、搜搜问问等平台提供的由用户生成的问答对，从中抽取相关答案并融合。答案融合系统从候选答案中抽取的答案是无序的，如果直接作为答案反馈给用户，可读性差，不利于理解。具体是，在答案融合系统中，从候选答案中抽取出的句子大多数是无序的，因此得到的答案对人们的阅读产生较大障碍。

[0003] 但抽取得到的句子大多与问题相关，一个问题的标准答案句子间存在较强的逻辑关系，如前后句间的逻辑关系；因此利用句子间的前后句关系对抽取得到的答案进行句子排序，提高句子间的连贯性，增加答案的可读性，增强用户体验，使答案更加通顺利于理解，对于用户来说将具有重要的意义。

[0004] 目前大多数答案融合系统中的句子排序多是根据答案在原候选答案中的相对位置进行排序，或者利用时间因素进行排序；而从不同答案中抽取的句子则无法进行排序。

发明内容

[0005] 为了解决上述技术问题，本发明提出了一种应用于答案融合系统的语句排序方法及系统。

[0006] 本发明是以如下技术方案实现的：

[0007] 第一方面提供了一种应用于答案融合系统的语句排序方法，包括：

[0008] 通过对输入的目标数据进行预处理，获得index形式的语句对；所述语句对为候选答案；

[0009] 对语句对进行特征提取，获取语句对中前后句子的关系概率，包括：

[0010] 判断语句对中句子是否为前后顺序的关系，利用基于注意力机制的seq2seq模型，对语句对进行特征提取，

[0011] 将提取得到的特征输入到前馈神经网络中，得到所述语句对中句子的关系概率；

[0012] 根据所述关系概率和解空间检索算法获得排序结果，所述解空间检索算法求解的目标函数是$Score(s,o)=\sum_{i=1}^{k}\sum_{j=i+1}^{k}score(s,o,i,j)$；s为所有目标语句组成的句子集，o为语句对构成的句子顺序，i、j为句子标号，表示第i句话，第j句话；排序结果为多个元素，每个元素包含句子顺序o以及该顺序的打分Score(s,o)；

[0013] 从排序结果中选取打分最高的句子顺序，得到目标答案。

[0014] 进一步地，所述通过对输入的目标数据进行预处理，获得index形式的语句对，包括：

[0015] 根据分词工具对所有的目标数据进行分词，所述目标数据中包括多个目标语句；

[0016] 将多个目标语句两两构建成语句对；

[0017] 通过语句转换功能将所述语句对转换为index形式的序列。

[0018] 进一步地，所述采用基于注意力机制的seq2seq模型，对语句对进行特征提取，包括：

[0019] 将index形式的语句对输入到seq2seq模型中；

[0020] 将两个LSTM模型进行拼接得到所述seq2seq模型；

[0021] 将attention机制应用到seq2seq模型上，通过前一LSTM模块LSTM1的隐藏层输出及后一LSTM模块LSTM2的每个节点的隐藏层输出，计算注意力分配权重，更新中间权重；

[0022] 根据更新得到的中间权重，结合LSTM1隐藏层输出及LSTM2最后节点的隐藏层输出计算得到特征值。

[0023] 进一步地，所述将index形式的语句对输入到seq2seq模型中，包括：

[0024] 将转换得到的index形式的语句对通过embedding层的转换，使其以词向量特征的形式进行表示；

[0025] 将语句对中分词得到的每个词输入到LSTM模块的节点中；

[0026] 对每个节点进行计算得到所述LSTM模块的隐藏层输出。

[0027] 进一步地，在两个LSTM模型中，前一个LSTM模型的隐藏层输出作为后一LSTM隐藏层输入。

[0028] 第二方面提供了一种应用于答案融合系统中的语句排序系统，包括：

[0029] 预处理模块，用于对输入的目标数据进行预处理，得到index形式的语句对，所述语句对为候选答案；

[0030] 关系概率获取模块，用于对语句对进行特征提取，获取语句对中的前后句子的关系概率，包括：

[0031] 特征提取单元，用于判断语句对中的句子是否为前后顺序的关系，采用基于注意力机制的seq2seq模型，对语句对进行特征提取；

[0032] 关系概率获取单元，用于将提取得到的特征输入到前馈神经网络中，得到所述语句对中句子的关系概率；

[0033] 排序结果获取模块，用于根据所述关系概率和解空间检索算法得到答案语句的排序结果；解空间检索算法求解的目标函数是 $Score(s,o)=\sum_{i=1}^{n}\sum_{j=i+1}^{n}score(s,o,i,j)$ ；s为所有目标语句组成的句子集，o为语句对构成的句子顺序，i、j为句子标号，表示第i句话，第j句话；排序结果为多个元素，每个元素包含句子顺序o以及该顺序的打分Score(s,o)；

[0034] 答案获取模块，用于排序结果中选取打分最高的句子顺序，得到目标答案。

[0035] 进一步地，所述预处理模块，包括：

[0036] 分词单元，用于根据分词工具对所有的目标数据进行分词，所述目标数据中包括多个目标语句；

[0037] 语句对构建单元，用于将多个目标语句两两构建成语句对；

[0038] 语句转换单元，用于通过语句转换功能将所述语句对转换为index形式的序列。

[0039] 进一步地，所述特征提取单元，包括：

[0040] 转换语句输入单元，用于将index形式的语句对输入到seq2seq模型中；

[0041] 预设模型拼接单元，用于将两个LSTM模型进行拼接得到所述seq2seq模型；

[0042] 权重获取单元，用于将attention机制应用到seq2seq模型上，通过前一LSTM模块LSTM1的隐藏层输出及后一LSTM模块LSTM2的每个节点的隐藏层输出，计算注意力分配权重，更新中间权重；

[0043] 特征获取单元，根据更新得到的中间权重，结合LSTM1隐藏层输出及LSTM2最后节点的隐藏层输出计算得到特征值。

[0044] 进一步地，所述转换语句输入单元，包括：

[0045] 词向量形式子单元，用于将index形式的语句对通过embedding层的转换，使其以词向量特征的形式进行表示；

[0046] 词语输入节点子单元，用于将语句对中分词得到的每个词输入到LSTM模块的节点中；

[0047] 隐藏层输出子单元，用于对每个节点进行计算得到所述LSTM模块的隐藏层输出。

[0048] 进一步地，在两个LSTM模型中，前一个LSTM模型的隐藏层输出作为后一LSTM模型隐藏层输入。

[0049] 本发明能够实现从不同答案中抽取的句子进行排序，从而使得获取的目标答案更加通顺，利于理解，且具有较强的可读性；从而提升了用户体验。

附图说明

[0050] 为了更清楚地说明本发明实施例中的技术方案，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，下面描述中的附图仅仅是本发明的一些实施例，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。

[0051] 图1是实施例一中应用于问答系统中的语句排序方法流程图；

[0052] 图2是实施例一中通过对输入的目标数据进行预处理，获得index形式的语句对的流程图；

[0053] 图3是实施例一中对语句对进行特征提取，获取语句对中前后句子的关系概率的流程图；

[0054] 图4是实施例一中将index形式的语句对输入到seq2seq模型中的流程图；

[0055] 图5是实施例一中基于注意力机制的seq2seq的模型结构图；

[0056] 图6是实施例一中解空间检索算法求解过程图；

[0057] 图7是实施例二中应用于问答系统中的语句排序结构框图。

具体实施方式

[0058] 为了使本技术领域的人员更好地理解本发明方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分的实施例，而不是全部的实施例。基于本发明中的实施例，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，都应当属于本发明保护的范围。

[0059] 需要说明的是，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列

出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0060] 需要说明的是，本发明利用句子的前后句关系对无序文本实现句子排序，提高提取的答案的可读性。

[0061] 实施例一：

[0062] 本实施例提供了一种应用于答案融合系统的语句排序方法，如图1所示，包括：

[0063] S101.通过对输入的目标数据进行预处理，获得index形式的语句对；所述语句对为候选答案；

[0064] 具体地，所述通过对输入的目标数据进行预处理，获得index形式的语句对，如图2所示，包括：

[0065] S101a.根据分词工具对所有的目标数据进行分词，所述目标数据中包括多个目标语句；

[0066] S101b.将多个目标语句两两构建成语句对；

[0067] S101c.通过语句转换功能将所述语句对转换为index形式的序列。

[0068] 例如，输入所述系统的是a、b、c三句话：

[0069] 其中：a、“专利制度旨在保护技术能够享受到独占性、排他性的权利，权利人之外的任何主体使用专利，都必须通过专利权人的授权许可才能获得使用权。”

[0070] b、“随着法律制度的不断完善，专利的使用呈现出多样化趋势，专利无效、专利撤销、过期专利等一一被列入专利法律范畴。”

[0071] c、“只有充分的认识诸如此类的法律制度，才能充分的利用专利资源，为企业实现更多的经济价值。”

[0072] 利用分词工具将句子进行分词，如将句a进行分词，得到“专利-制度-旨在-保护-技术-能够-享受到-独占性、排他性-的-权利，权利-人-之外-的-任何-主体-使用-专利，都必须-通过-专利权-人-的-授权-许可-才能-获得-使用权。”

[0073] 将输入的目标语句a、b、c构建成语句对(a,a)，(a,b)，(a,c)，(b,a)，(b,b)，(b,c)，(c,a)，(c,b)，(c,c)；

[0074] 进一步通过语句转换功能将a,b,c转换为使用index形式的序列。

[0075] S102.对语句对进行特征提取，获取语句对中前后句子的关系概率，包括：

[0076] 判断语句对中句子是否为前后顺序的关系，利用基于注意力机制的seq2seq模型，对语句对进行特征提取，

[0077] 将提取得到的特征输入到前馈神经网络中，得到所述语句对中句子的关系概率；

[0078] 其中，判断语句对中句子是否为前后顺序的关系，利用基于注意力机制的seq2seq模型，对语句对进行特征提取，将提取得到的特征输入到前馈神经网络中，得到句子前后顺序概率；若概率为1，则是前后关系；若为0，则不是。

[0079] 需要说明的是，基于注意力机制的seq2seq(Sequence to Sequence)模型的是利用了注意力机制以及seq2seq模型。其中，注意力机制是一个资源分配模型，它模拟人脑工作，将更多的资源集中在重要的内容上。而在本方法的基本模型采用了seq2seq模型，并基于seq2seq模型实现了注意力模型，利用注意力机制计算前句与后句之间词与词之间关系。

[0080] 其中，所述采用基于注意力机制的seq2seq模型，对语句对进行特征提取，如图3所

示，包括：

[0081] S102a.将index形式的语句对输入到seq2seq模型中；

[0082] 具体地，所述将index形式的语句对输入到seq2seq模型中，如图4所示，包括：

[0083] S1021a.将转换得到的index形式的语句对通过embedding层的转换，使其以词向量特征的形式进行表示；

[0084] 其中，词向量特征是利用gensim工具，通过大规模自然语言文本训练得到的。用词向量表示句子，是将句子进行分词，并将句子中每个词用词向量进行特征表示，从而用词向量表示整个句子。

[0085] S1022a.将语句对中分词得到的每个词输入到LSTM模块的节点中；

[0086] S1023a.对每个节点进行计算得到所述LSTM模块的隐藏层输出。

[0087] S102b.将两个LSTM模型进行拼接得到所述seq2seq模型；

[0088] 其中，LSTM(Long-Short Term Memory)为长短期记忆模型模型；在两个LSTM模型中，前一个LSTM模型的隐藏层输出作为后一LSTM隐藏层输入。具体地，seq2seq模型是由两个LSTM模型拼接而成，第一个LSTM模型L1的输入是由词向量表示的句子a。第二个LSTM模型L2的输入包括L1的最后一个节点的输出及由词向量表示的第二个句子b。

[0089] 还有，LSTM模型是一种用于处理时序数据的模型，是深度学习模型中的一种，对于句子级别的特征提取具有良好的效果。LSTM由一系列cell组成，图5中的c1，c2等代表一个代表LSTM的cell。每个cell有两个输入，一是输入一个词，如“今天”，“天气”等，二是上一cell输出。每个LSTM cell的输出包括隐藏层输出h以及当前cell的状态c。LSTM内部包含三个门输入门、遗忘门、输出门以及细胞状态。LSTM计算公式如下所示：

[0090] $$a_j^t = \sum_{i=1}^{I} w_{ij} x_i^t + \sum_{h=1}^{H} w_{hj} b_h^{t-1} + \sum_{c=1}^{C} w_{cj} s_c^{t-1} \quad (1)$$

[0091] $$b_j^t = f(a_j^t) \quad (2)$$

[0092] $$a_\phi^t = \sum_{i=1}^{I} w_{i\phi} x_i^t + \sum_{h=1}^{H} w_{h\phi} b_h^{t-1} + \sum_{c=1}^{C} w_{c\phi} s_c^{t-1} \quad (3)$$

[0093] $$b_\phi^t = f(a_\phi^t) \quad (4)$$

[0094] $$a_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} b_h^{t-1} \quad (5)$$

[0095] $$s_c^t = b_\phi^t s_c^{t-1} + b_j^t g(a_c^t) \quad (6)$$

[0096] $$a_\omega^t = \sum_{i=1}^{I} w_{i\omega} x_i^t + \sum_{h=1}^{H} w_{h\omega} b_h^{t-1} + \sum_{c=1}^{C} w_{c\omega} s_c^{t-1} \quad (7)$$

[0097] $$b_\omega^t = f(a_\omega^t) \quad (8)$$

[0098] 其中，式(1)(2)为输入门计算公式，式(3)(4)为遗忘门计算公式，式(5)(6)为细胞状态计算公式，式(7)(8)为输出门计算公式；x为输入，b为隐藏层输出，s为细胞状态，w为中间权值。

[0099] S102c.将attention机制应用到seq2seq模型上，通过前一LSTM模块LSTM1的隐藏层输出及后一LSTM模块LSTM2的每个节点的隐藏层输出，计算注意力分配权重，更新中间权重；

[0100] 其中，如图5所示，为基于注意力机制的seq2seq的模型结构图，其中，以句A和句B为例，句A为“今天天气适合出行”，句B为“那么去哪里玩”，进行处理示意图的表示。

[0101] S102d.根据更新得到的中间权重，结合LSTM1隐藏层输出及LSTM2最后节点的隐藏层输出计算得到特征值。

[0102] seq2seq模型中attention模型的计算公式如下所示：

[0103] $M_t = \tanh(W^Y Y + (W^h h_t + W^h r_{t-1}) \times e_L)$ (1)

[0104] $a_t = \mathrm{sotfmax}(w^T M_t)$ (2)

[0105] $r_t = Ya_t^T + \tanh(W^r r_{t-1})$ (3)

[0106] $h^* = \tanh(W^p r_N + W^x h_N)$ (4)

[0107] LSTM2阶段每个cell节点利用attention计算公式(1)(2)(3)反复更新r，在最终节点计算h^*，最终得特征M。

[0108] 需要说明的是，所述关系概率涉及到前句和后句之间的词与词之间的关系。

[0109] S103.根据所述关系概率和解空间检索算法获得排序结果；

[0110] 所述解空间检索算法求解的目标函数是$Score(s,o) = \sum_{i=1}^{s} \sum_{j=i+1}^{s} score(s,o,i,j)$；s为所有目标语句组成的句子集，o为语句对构成的句子顺序，i、j为句子标号，表示第i句话，第j句话；排序结果为多个元素，每个元素包含句子顺序o以及该顺序的打分Score(s,o)；

[0111] 其中，解空间检索算法为最优排序求解算法，具体地，所述解空间检索算法包括beam search算法，根据前后句关系模块得到的语句对中前后句子的关系概率，检索句子排序的解空间，从而得到概率最大的句子排序，得到最后的排序结果。进一步地，Beam Search算法是一种启发式搜索算法，通常用在图的解空间比较大的情况下，为了减少搜索所占用的空间以及时间，在每一步深度扩展的时候，减掉一些质量比较差的节点，从而减少空间消耗。

[0112] 最优排序求解模型求解过程如图6所示，其中open表记录候选排序组合并记录概率和，如[(1,0,3),1.228]，(1,0,3)为已经考虑到的句子组合，1.228为(1,0)，(0,3)两个句子对概率的概率和。

[0113] open表大小为100，即存储了100个如[(1,0,3),1.228]一样的记录。beam表存储的记录形式与open相同，beam表记录根据open得到的记录候选，是临时记录。

[0114] S104.从排序结果中选取打分最高的句子顺序，得到目标答案。

[0115] 需要说明的是，在本方法中利用计算得到的句子对中前后句的关系，求解最佳排序；通过基于注意力机制的seq2seq模型求解句子对间的概率，此概率代表句子对是前后句关系的可能性。通过Beam Search算法求解句子排列的组合，求解的目标函数是$Score(s,o) = \sum_{i=1}^{s} \sum_{j=i+1}^{s} score(s,o,i,j)$，即求解前后句关系概率和最大的句子组合o，这样才属于目标答案。

[0116] 实施例二：

[0117] 本实施例提供了一种应用于答案融合系统中的语句排序系统，如图7所示，包括：

[0118] 预处理模块110，用于对输入的目标数据进行预处理，得到index形式的语句对，所述语句对为候选答案；

[0119] 进一步地，所述预处理模块110，包括：

[0120] 分词单元111，用于根据分词工具对所有的目标数据进行分词，所述目标数据中包括多个目标语句；

[0121] 语句对构建单元112，用于将多个目标语句两两构建成语句对；

[0122] 语句转换单元113，用于通过语句转换功能将所述语句对转换为index形式的序列。

[0123] 关系概率获取模块120，用于对语句对进行特征提取，获取语句对中的前后句子的关系概率，包括：

[0124] 特征提取单元，用于判断语句对中的句子是否为前后顺序的关系，采用基于注意力机制的seq2seq模型，对语句对进行特征提取；

[0125] 关系概率获取单元，用于将提取得到的特征输入到前馈神经网络中，得到所述语句对中句子的关系概率；

[0126] 进一步地，所述特征提取单元121，包括：

[0127] 转换语句输入单元1211，用于将index形式的语句对输入到seq2seq模型中；

[0128] 具体地，所述转换语句输入单元1211，包括：

[0129] 词向量形式子单元1211a，用于将index形式的语句对通过embedding层的转换，使其以词向量特征的形式进行表示；

[0130] 词语输入节点子单元1211b，用于将语句对中分词得到的每个词输入到LSTM模块的节点中；

[0131] 隐藏层输出子单元1211c，用于对每个节点进行计算得到所述LSTM模块的隐藏层输出。

[0132] 进一步地，在两个LSTM模型中，前一个LSTM模型的隐藏层输出作为后一LSTM模型隐藏层输入。

[0133] 预设模型拼接单元1212，用于将两个LSTM模型进行拼接得到所述seq2seq模型；

[0134] 权重获取单元1213，用于将attention机制(注意力机制)应用到seq2seq模型上，通过前一LSTM模块LSTM1的隐藏层输出及后一LSTM模块LSTM2的每个节点的隐藏层输出，计算注意力分配权重，更新中间权重；

[0135] 特征获取单元1214，根据更新得到的中间权重，结合LSTM1隐藏层输出及LSTM2最后节点的隐藏层输出计算得到特征值。

[0136] 排序结果获取模块，用于根据所述关系概率和解空间检索算法得到答案语句的排序结果；解空间检索算法求解的目标函数是 $Score(s,o)=\sum_{i=1}^{n}\sum_{j=i+1}^{n}score(s,o,i,j)$ ；s为所有目标语句组成的句子集，o为语句对构成的句子顺序，i、j为句子标号，表示第i句话，第j句话；排序结果为多个元素，每个元素包含句子顺序o以及该顺序的打分Score(s,o)；

[0137] 答案获取模块，用于排序结果中选取打分最高的句子顺序，得到目标答案。

[0138] 本发明基于注意力机制及seq2seq模型获取到语句对中前后句子的关系概率，其中所述关系概率涉及到前句和后句之间的词与词之间的关系；结合所述关系概率得到目标语句中最佳的排序结果，

[0139] 本发明能够实现从不同答案中抽取的句子进行排序，从而使得获取的目标答案更加通顺，利于理解，且具有较强的可读性；从而提升了用户体验。

[0140] 在本发明的上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其他实施例的相关描述。

[0141] 本发明中的技术方案中的各个模块均可通过计算机终端或其它设备实现。所述计

算机终端包括处理器和存储器。所述存储器用于存储本发明中的程序指令/模块，所述处理器通过运行存储在存储器内的程序指令/模块，实现本发明相应功能。

[0142] 本发明中的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在存储介质中，包括若干指令用以使得一台或多台计算机设备(可为个人计算机、服务器或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。

[0143] 本发明中所述模块/单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。可以根据实际的需要选择其中的部分或者全部模块/单元来达到实现本发明方案的目的。

[0144] 另外，在本发明各个实施例中的各模块/单元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。

[0145] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

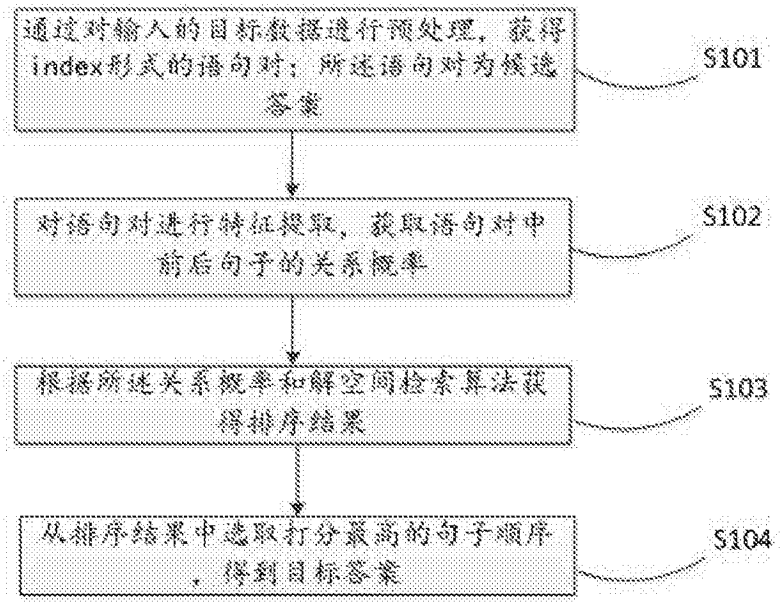

通过对输入的目标数据进行预处理，获得index形式的语句对；所述语句对为候选答案
S101
对语句对进行特征提取，获取语句对中前后句子的关系概率
S102
根据所述关系概率和解空间检索算法获得排序结果
S103
从排序结果中选取打分最高的句子顺序，得到目标答案
S104

图1

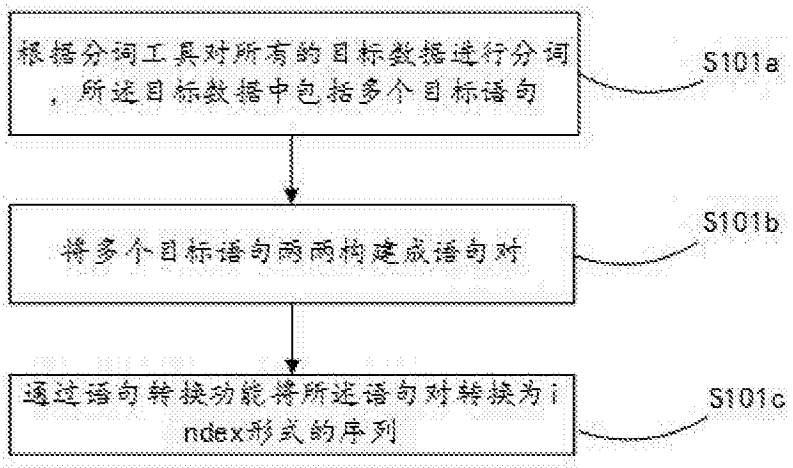

根据分词工具对所有的目标数据进行分词，所述目标数据中包括多个目标语句
S101a
将多个目标语句两两构建成语句对
S101b
通过语句转换功能将所述语句对转换为index形式的序列
S101c

图2

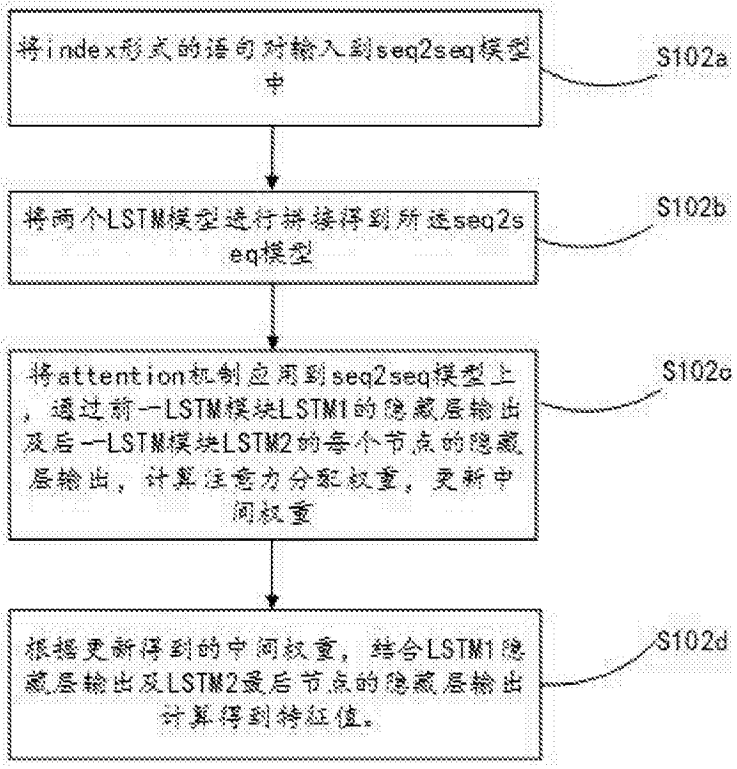

将index形式的语句对输入到seq2seq模型中
S102a
将两个LSTM模型进行拼接得到所述seq2seq模型
S102b
将attention机制应用到seq2seq模型上，通过前一LSTM模块LSTM1的隐藏层输出及后一LSTM模块LSTM2的每个节点的隐藏层输出，计算注意力分配权重，更新中间权重
S102c
根据更新得到的中间权重，结合LSTM1隐藏层输出及LSTM2最后节点的隐藏层输出计算得到特征值。
S102d

图3

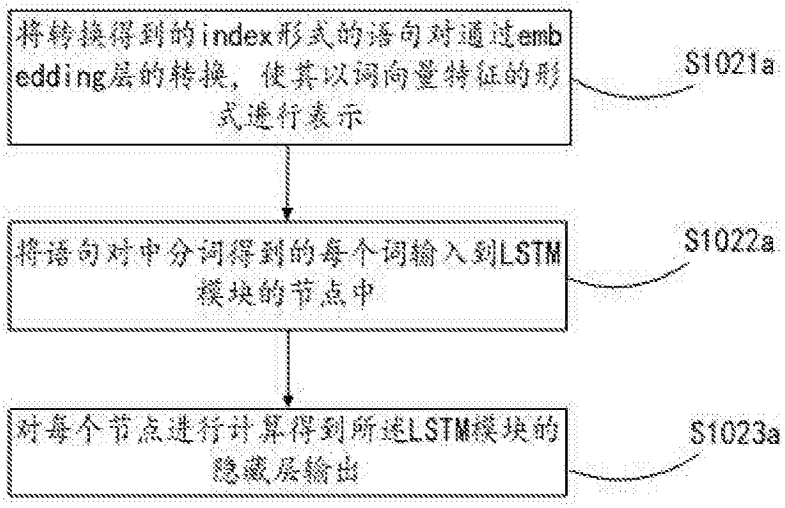

将转换得到的index形式的语句对通过embedding层的转换，使其以词向量特征的形式进行表示
S1021a
将语句对中分词得到的每个词输入到LSTM模块的节点中
S1022a
对每个节点进行计算得到所述LSTM模块的隐藏层输出
S1023a

图4

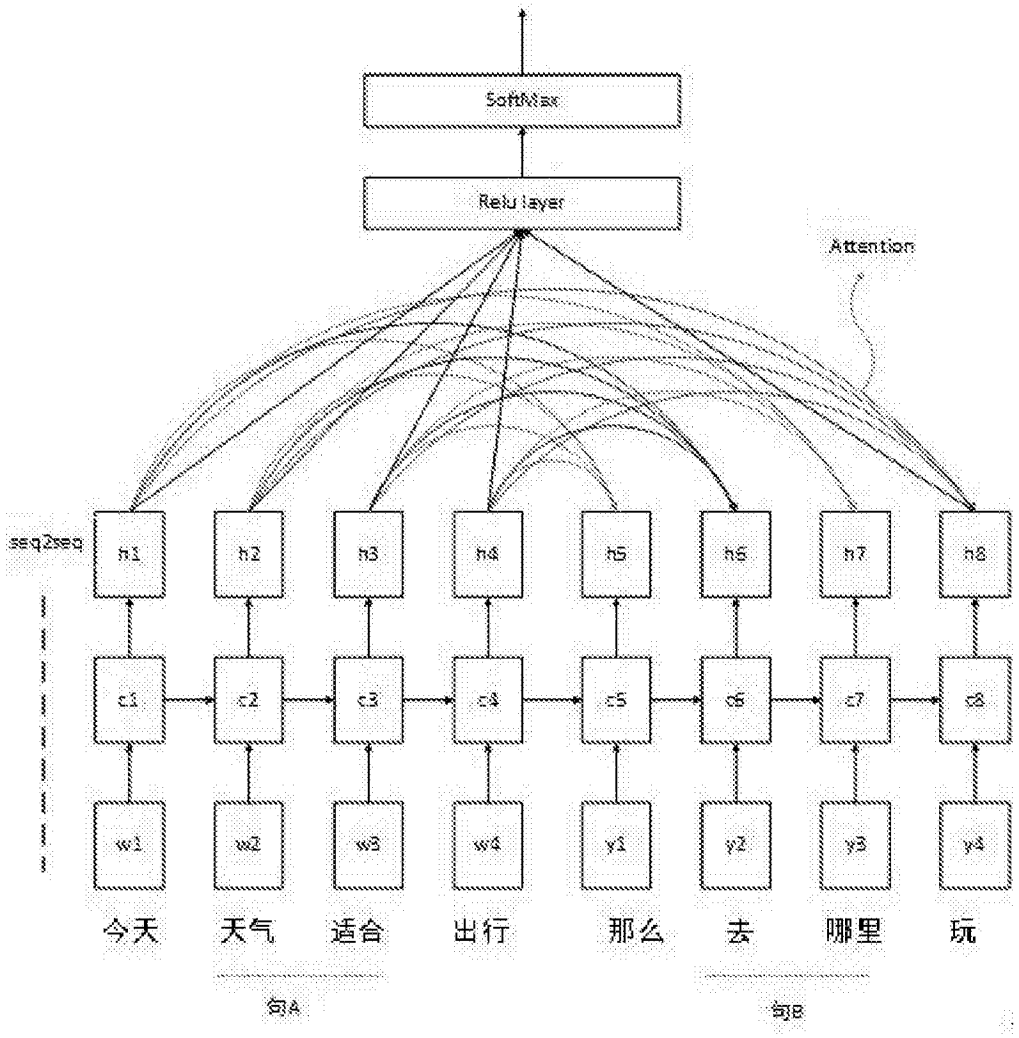

SoftMax
Relu layer
Attention
seq2seq
h1
h2
h3
h4
h5
h6
h7
h8
c1
c2
c3
c4
c5
c6
c7
c8
w1
w2
w3
w4
y1
y2
y3
y4
今天
天气
适合
出行
那么
去
哪里
玩
句A
句B

图5

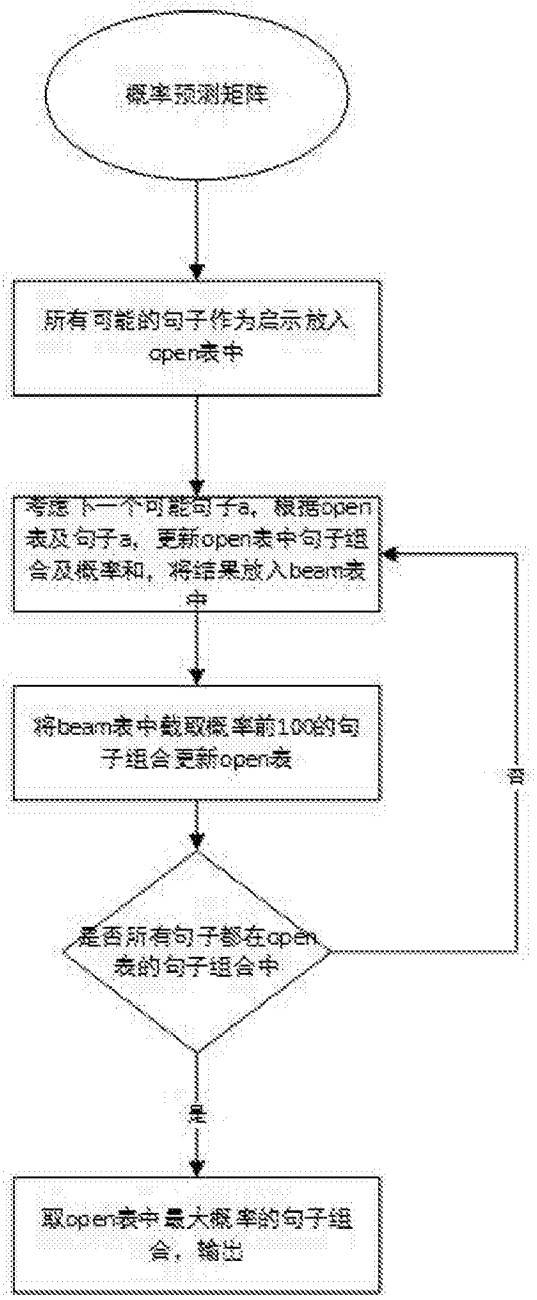

概率预测矩阵
所有可能的句子作为启示放入open表中
考虑下一个可能句子a，根据open表及句子a，更新open表中句子组合及概率和，将结果放入beam表中
将beam表中截取概率前100的句子组合更新open表
是否所有句子都在open表的句子组合中
否
是
取open表中最大概率的句子组合，输出

图6

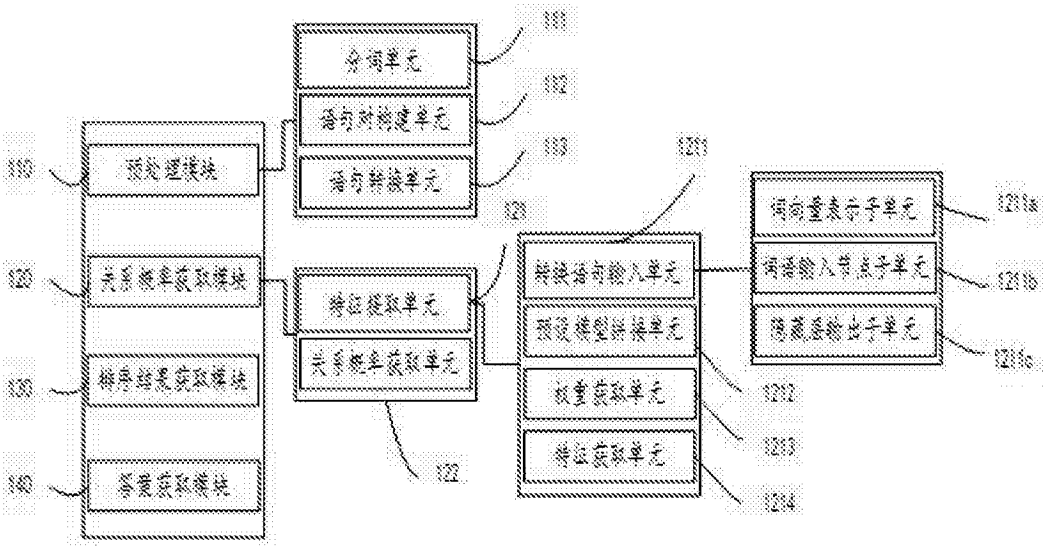

预处理模块
关系概率获取模块
排序结果获取模块
答案获取模块
分词单元
语句树构建单元
语句转换单元
特征获取单元
关系概率获取单元
特征输入单元
预设模型获取单元
权重获取单元
特征获取单元
词向量表示子单元
词语输入节点子单元
隐藏层输出子单元

图7