



(12) 发明专利

(10) 授权公告号 CN 110706807 B

(45) 授权公告日 2021.02.12

(21) 申请号 201910867117.2

CN 101950340 A,2011.01.19

(22) 申请日 2019.09.12

CN 106502981 A,2017.03.15

US 2017206270 A1,2017.07.20

(65) 同一申请的已公布的文献号

申请公布号 CN 110706807 A

周永梅.《基于本体的自动问答系统》.《中国优秀硕士学位论文全文数据库 信息科技辑》.2013,

(43) 申请公布日 2020.01.17

周永梅.《基于本体的自动问答系统》.《中国优秀硕士学位论文全文数据库 信息科技辑》.2013,

(73) 专利权人 北京四海心通科技有限公司

地址 100080 北京市海淀区海淀中街16号3层5单元308

黄妮.《网络学习平台中的分词与句子相似度算法研究》.《中国优秀硕士学位论文全文数据库 信息科技辑》.2012,

(72) 发明人 郭新龙

(74) 专利代理机构 北京市广友专利事务所有限责任公司 11237

代理人 张仲波

郑志蕴等.《本体语义相似度自适应综合加权算法研究》.《计算机科学》.2016,第43卷(第10期),

(51) Int.Cl.

G16H 50/20 (2018.01)

G06F 16/33 (2019.01)

G06F 16/332 (2019.01)

邹文科.《基于本体技术的语义检索及其语义相似度研究》.《中国优秀硕士学位论文全文数据库 信息科技辑》.2008,

(56) 对比文件

CN 104133812 A,2014.11.05

审查员 罗湘

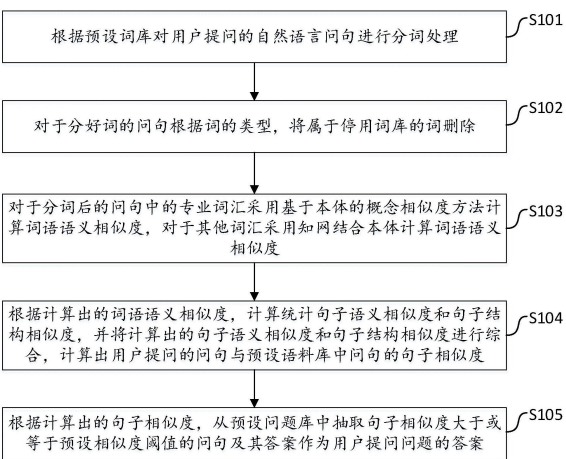
权利要求书3页 说明书13页 附图1页

(54) 发明名称

一种基于本体语义相似度的医学问答方法

(57) 摘要

本发明提供一种基于本体语义相似度的医学问答方法,通过将医学领域用户常见问题与语义分析,语义相似度计算,本体等相关技术结合,提出了一种综合的语义相似度计算方法,并结合浅层语义分析和文本分析技术,提出了基于语义相似度的医学问答方法。本发明中的语义相似度计算方法结合了句子表层的相似度,如句长,词性等,以及基于概念词的语义相似度,并且对于概念词相似度的计算采用了属性和距离综合的方法。使得本发明的医学问答系统可实现五大类问题(方式、时间、程度、数量、实体)处理,达到80%以上的准确率。五类问题基本涵盖基础医学领域的学科问题,较好地实现了系统的功能。



1. 一种基于本体语义相似度的医学问答方法,其特征在于,包括:

根据预设词库对用户提问的自然语言问句进行分词处理;

对于分词后的问句中的专业词汇采用基于本体的概念相似度方法计算词语语义相似度,对于其他词汇采用知网结合本体计算词语语义相似度;

根据计算出的词语语义相似度,计算统计句子语义相似度和句子结构相似度,并将计算出的句子语义相似度和句子结构相似度进行综合,计算出用户提问的问句与预设语料库中问句的句子相似度;

根据计算出的句子相似度,从预设问题库中抽取句子相似度大于或等于预设相似度阈值的问句及其答案作为用户提问问题的答案;所述预设问题库以PA结构为基础动态构建,所述PA结构由谓词和语义组成,所述本体分为事件类本体和论元类本体;所述事件类本体对应着PA结构中的谓词,描述的是动词性概念,所述论元类本体对应着PA结构中的论元,描述的是名词性概念;

其中,所述计算统计句子结构相似度,具体为:

采用基于规则的方法计算句子结构相似度,所述基于规则的方法结合词距,句长,同时考虑词性的作用,综合三个方面确定句子的结构相似度;其中,

词距相似度的公式如下:

$$DisSim(Q_1, Q_2) = 1 - \frac{|SameDis(Q_1) - SameDis(Q_2)|}{Dis(Q_1) + Dis(Q_2)} \quad (1)$$

式中,SameDis(Q_i)表示 Q_1 和 Q_2 中相同关键词在 Q_i 中的距离,若关键词重复出现,则以最大距离为准,Dis(Q_i)表示句子关键词中最左及最右关键词之间的距离,若关键词重复出现,则以最小距离为准;

句长相似度的公式如下:

$$LenSim(Q_1, Q_2) = 1 - \frac{|Len(Q_1) - Len(Q_2)|}{Len(Q_1) + Len(Q_2)} \quad (2)$$

式中,Len(Q_1)和Len(Q_2)分别表示两个句子的长度;

词性相似度的公式如下:

$$WordProSim(Q_1, Q_2) = \frac{SameWordPro(Q_1, Q_2)}{WordPro(Q_1) + WordPro(Q_2) - SameWordPro(Q_1, Q_2)} \quad (3)$$

式中,WordPro(Q_1)和WordPro(Q_2)分别表示句子 Q_1 和 Q_2 中关键词的权值,SameWordPro(Q_1, Q_2)表示句子 Q_1 和 Q_2 中相同关键词的权值;当句子 Q_1 和 Q_2 中关键词全都相同时,分母达到最小值;

结合所述词距相似度、句长相似度,以及词性相似度,将三种相似度进行线性融合,得到句子结构相似度为:

$$SynSim = \lambda_1 LenSim(Q_1, Q_2) + \lambda_2 DisSim(Q_1, Q_2) + \lambda_3 WordProSim(Q_1, Q_2) \quad (4)$$

式中, $\lambda_1, \lambda_2, \lambda_3$ 为权重系数,代表对三种相似度的偏重程度;

所述计算统计句子语义相似度具体为:基于概念语义相似度计算句子语义相似度,所述概念语义相似度计算包括基于距离的概念语义相似度计算、基于属性的概念语义相似度计算和基于主成分分析的概念语义相似度混合计算;其中,所述基于距离的概念语义相似度计算,具体为:

考虑领域本体类的层次关系,利用语义重合度、语义层次深度、语义距离和语义距离多种因素,来计算概念的语义相似度;其中,

语义重合度计算公式为:

$$Sim_{overlap}(a,b) = \frac{ParentNum(a) \cap ParentNum(b)}{Max(ParentNum(a), ParentNum(b))} \quad (5)$$

式中, $ParentNum(i)$ 表示 i 的父节点个数;分子表示 a 和 b 相同父节点的个数;分母表示 a 和 b 父节点个数较多的那个;

语义层次深度是指概念在结构树中所处层次的深度;两个概念的语义相似度与他们深度之和呈正相关关系,与他们的深度之差呈负相关关系;假设根节点 R 的深度为 1, i 节点的父节点用 $Parent(i)$ 表示,则非根节点 i 在本体结构树中的深度公式为:

$$Depth(i) = Depth(Parent(i)) + 1 \quad (6)$$

本体树 T 的深度为:

$$Depth(T) = Max(Depth(i)) \quad (7)$$

即本体树的深度取树中节点的最大深度;

语义距离对应的相似度公式为:

$$Sim_{Dis}(a,b) = \frac{2Depth(T)}{2Depth(T) - Dis(a,b)} \quad (8)$$

式中, $Dis(a,b)$ 表示节点 a 和 b 在本体结构树中的最短路径边数;

语义密度对应的相似度公式如下:

$$Sim_{Den}(a,b) = \frac{BrotherNum(a) + BrotherNum(b)}{2Max(BrotherNum(a), BrotherNum(b))} \quad (9)$$

式中, $BrotherNum(i)$ 为节点 i 的兄弟节点的个数;

基于公式 (5)、公式 (6)、公式 (7)、公式 (8) 和公式 (9), 得到基于距离的概念语义相似度的计算模型:

$$Sim_1(a,b) = \alpha Sim_{overlap}(a,b) - \beta Sim_{Dis}(a,b) - \gamma Sim_{Den}(a,b) \quad (10)$$

式中, α, β, γ 为权重系数;

所述基于属性的概念语义相似度计算的模型如下:

$$Sim_2(c,d) = \alpha Properties(c \cap d) - \beta Properties(c-d) - \gamma Properties(d-c) \quad (11)$$

其中, $Properties(c \cap d)$ 表示概念 c 和 d 所拥有的公共属性的集合, $Properties(c-d)$ 表示概念 c 拥有而概念 d 不拥有的属性集合, $Properties(d-c)$ 则表示概念 d 拥有而概念 c 不拥有的属性集合;

所述基于主成分分析的概念语义相似度混合计算,具体为:

基于主成分分析法对基于距离因素和基于属性因素计算的数据进行分析,动态计算出各个因素的权重;计算出权重后对多因素的计算结果进行加权线性相加得到最终概念语义相似度;其过程包括:

将基于距离的相似度和基于属性的相似度作为两个维度,通过多个样本的计算得到相似度矩阵作为输入样本矩阵;

对所述输入样本矩阵进行矩阵标准化变换为标准矩阵 Z , 并求出相关系数矩阵 $R1$;

求出相关系数矩阵 $R1$ 的特征方程的 32 个特征根以确定主成分;

解方程组单位特征向量;

转换标准化后的指标变量为主成分；

将两个主成分进行加权并线性求和，即得最终概念语义相似度值，相应的权值为各个主成分的贡献率；

设句子 Q_1 和 Q_2 分别含有 m 个概念 $W_{11}\cdots W_{1m}$ 和 n 个概念 $W_{21}\cdots W_{2n}$ ；

所述基于概念语义相似度计算句子语义相似度，具体为：

从 Q_1 中选出一个概念分别与 Q_2 中的 n 个概念，计算概念语义相似度，直至 Q_1 中概念循环完为止，得到概念语义相似度矩阵：

$$T = \begin{matrix} S(W_{11}, W_{21}) & \cdots & S(W_{11}, W_{2n}) \\ \vdots & \ddots & \vdots \\ S(W_{1m}, W_{21}) & \cdots & S(W_{1m}, W_{2n}) \end{matrix} \quad (12)$$

计算第一个概念集与第二个概念集的概念语义平均最大相似度

$$avgS_{Q_1} = \frac{\sum_{i=1}^m \text{Max}(S(W_{1i}, W_{21}), S(W_{1i}, W_{22}), \dots, S(W_{1i}, W_{2n}))}{m} \quad (13)$$

计算第二个概念集与第一个概念集的概念的平均最大相似度

$$avgS_{Q_2} = \frac{\sum_{j=1}^n \text{Max}(S(W_{11}, W_{2j}), S(W_{12}, W_{2j}), \dots, S(W_{1m}, W_{2j}))}{n} \quad (14)$$

计算公式 (13) 和 (14) 结果，求平均值，得到句子语义相似度

$$\text{SemSim} = \frac{avgS_{Q_1} + avgS_{Q_2}}{2} \quad (15)$$

所述用户提问的问句与预设语料库中间句的句子相似度定义如下：

$$\text{Sim} = (1-\eta) \text{SynSim} + \eta \text{SemSim} \quad (16)$$

其中， η 为权重系数，当 η 为1时，表示只使用语义相似度；当 η 为0时表示只使用句子结构相似度；

所述根据预设词库对用户提问的自然语言问句进行分词处理，具体为：

根据预设词库对用户提问的自然语言问句通过正向最大匹配的方法进行分词处理，并将其中的未登记词借助于预设分词工具添加到预设词库中；

在所述根据计算出的句子相似度，从预设问题库中抽取句子相似度大于或等于预设相似度阈值的问句及其答案后，所述方法还包括：

按照句子相似度大小对抽取的问句及其答案进行排序，选取前五个问句及其答案作为用户提问问题的答案。

2. 如权利要求1所述的基于本体语义相似度的医学问答方法，其特征在于，在所述根据预设词库对用户提问的自然语言问句进行分词处理后，所述方法还包括：对于分好词的问句根据词的类型，将属于停用词库的词删除。

一种基于本体语义相似度的医学问答方法

技术领域

[0001] 本发明涉及问答系统技术领域,特别是指一种基于本体语义相似度的医学问答方法。

背景技术

[0002] 问答系统是融合了自然语言处理和信息检索技术的产品。根据涉及范围的不同,可以分为开放领域和限定领域两个大类。当前领域本体在问答系统中的应用主要体现在对领域概念的分类方面,缺乏对用户问句的语义理解和处理。

[0003] 此外,自然语言处理会涉及到语义相似度的计算,其定义如下:

[0004] Dekang Lira认为任何两个词语的相似度取决于它们的共性(Commonality)和个性(Differences),然后从信息论的角度给出了定义公式:

$$[0005] \quad \text{Sim}(A, B) = \frac{\log_p(\text{Common}(A, B))}{\log_p(\text{description}(A, B))}$$

[0006] 其中,分子表示描述A,B共性所需要的信息量;分母表示完整地描述A,B所需要的信息量。刘群、李素建以基于实例的机器翻译为背景,认为语义相似度就是两个词语在不同的上下文中可以互相替换使用而不改变文本的句法语义结构的程度。两个词语,如果在不同的上下文中可以互相替换且不改变文本的句法语义结构的可能性越大,二者的相似度就越高,否则相似度就越低。对于两个词语W1、W2,如果记其相似度为Sim(W1,W2),其词语距离为Dis(W1,W2),根据刘群、李素建的公式:

$$[0007] \quad \text{Sim}(W_1, W_2) = \frac{\alpha}{\alpha + \text{Dis}(W_1, W_2)}$$

[0008] 其中是 α 一个可调节的参数。 α 的含义是当相似度为0.5时的词语距离值。

[0009] 词语距离和词语相似度是一对词语的相同关系特征的不同表现形式,如果两个概念之间的语义距离越近,就认为它们越相似,因此二者之间可以给出一个简单对应关系:

$$[0010] \quad \text{Sim}(W_1, W_2) = \frac{k}{\text{Dis}(W_1, W_2)}$$

[0011] 其中,Dis(W1,W2)为数中W1、W2所代表的结点在数中的距离,k为比例系数。

[0012] 一般地说,相似度一般被定义为一个0到1之间的实数。特别地,当两个词语完全一样时,它们的相似度为1;当两个词语是完全不同的概念时,它们的相似度接近于0。

[0013] 目前,关于语义相似度的计算主要有以下方式:

[0014] 1、基于句法结构的相似度算法

[0015] 基于树的语义相似度计算的算法大体上分为两种:一是基于距离的语义相似性测度;二是基于信息内容的语义相似性测度。一般是利用一部语义词典(如Wordnet,HowNet),语义词典都是将所有的词组织在一棵或几棵树状的层次结构中。在一棵树状图中,任何两个结点之间有且只有一条路径。于是,这条路径的长度就可以作为这两个词语概念间语义距离的一种度量;而且随着概念所处结点越深,其所包含的语义信息越丰富,越能准确地决

定概念的性质,它们对语义相似度起着决定作用。

[0016] 2、基于本体的语义相似度算法

[0017] 本体概念语义相似度计算是自然语言处理中的重点,也是人工智能领域中的难点之一。本体概念语义相似度计算的理论基础来自离散数学中的图论和树的匹配理论,但这种思想没有考虑概念的语义信息。一般在不同的知识领域中,概念的相似度被赋予不同的含义。例如,在信息整合中,相似度通常指多个文本之间匹配的相关程度。在信息检索领域,相似度用来反映用户查询内容与被检索文本在语义上的相关度。相似度越高,表示被检索文本与用户的需求越符合。同时,由于词语语义多元化的特征,即相同的词语在不同的语境中可能具有不同的语义,因此,在本体语义相似度计算中充分考虑概念的语义信息成为近年来本体概念相似度研究的热点。

[0018] 而单一地采用上述现有的语义相似度的计算方法进行语义计算的准确率不够理想,不能满足医学领域问答系统的应用需求。

发明内容

[0019] 本发明要解决的技术问题是提供一种基于本体语义相似度的医学问答方法,通过提出一种用于语义相似度计算的综合计算模型,实现对现有计算方法的整合,有效提高语义相似度计算的准确率;并在此基础上结合浅层语义分析和文本分析技术,实现基于本体语义相似度的医学问答方案。

[0020] 为解决上述技术问题,本发明提供一种基于本体语义相似度的医学问答方法,所述基于本体语义相似度的医学问答方法包括:

[0021] 根据预设词库对用户提问的自然语言问句进行分词处理;

[0022] 对于分词后的问句中的专业词汇采用基于本体的概念相似度方法计算词语语义相似度,对于其他词汇采用知网结合本体计算词语语义相似度;

[0023] 根据计算出的词语语义相似度,计算统计句子语义相似度和句子结构相似度,并将计算出的句子语义相似度和句子结构相似度进行综合,计算出用户提问的问句与预设语料库中间句的句子相似度;

[0024] 根据计算出的句子相似度,从预设问题库中抽取句子相似度大于或等于预设相似度阈值的问句及其答案作为用户提问问题的答案。

[0025] 进一步地,所述根据预设词库对用户提问的自然语言问句进行分词处理,具体为:根据预设词库对用户提问的自然语言问句通过正向最大匹配的方法进行分词处理,并将其中的未登记词借助于预设分词工具添加到预设词库中。

[0026] 进一步地,在根据预设词库对用户提问的自然语言问句进行分词处理后,所述方法还包括:对于分好词的问句根据词的类型,将属于停用词库的词删除。

[0027] 进一步地,在所述根据计算出的句子相似度,从预设问题库中抽取句子相似度大于或等于预设相似度阈值的问句及其答案后,所述方法还包括:

[0028] 按照句子相似度大小对抽取的问句及其答案进行排序,选取前五个问句及其答案作为用户提问问题的答案。

[0029] 进一步地,所述计算统计句子结构相似度,具体为:

[0030] 采用基于规则的方法计算句子结构相似度,所述基于规则的方法结合词距,句长,

同时考虑词性的作用,综合三个方面确定句子的结构相似度;其中,

[0031] 词距相似度的公式如下:

$$[0032] \quad DisSim(Q_1, Q_2) = 1 - \frac{|SameDis(Q_1) - SameDis(Q_2)|}{Dis(Q_1) + Dis(Q_2)} \quad (1)$$

[0033] 式中, SameDis(Q_i) 表示Q₁和Q₂中相同关键词在Q_i中的距离,若关键词重复出现,则以最大距离为准, Dis(Q_i) 表示句子关键词中最左及最右关键词之间的距离,若关键词重复出现,则以最小距离为准;

[0034] 句长相似度的公式如下:

$$[0035] \quad LenSim(Q_1, Q_2) = 1 - \frac{|Len(Q_1) - Len(Q_2)|}{Len(Q_1) + Len(Q_2)} \quad (2)$$

[0036] 式中, Len(Q₁) 和 Len(Q₂) 分别表示两个句子的长度;

[0037] 词性相似度的公式如下:

$$[0038] \quad WordProSim(Q_1, Q_2) = \frac{SameWordPro(Q_1, Q_2)}{WordPro(Q_1) + WordPro(Q_2) - SameWordPro(Q_1, Q_2)} \quad (3)$$

[0039] 式中, WordPro(Q₁) 和 WordPro(Q₂) 分别表示句子Q₁和Q₂中关键词的权值, SameWordPro(Q₁, Q₂) 表示句子Q₁和Q₂中相同关键词的权值;当句子Q₁和Q₂中关键词全都相同时,分母达到最小值;

[0040] 结合所述词距相似度、句长相似度,以及词性相似度,将三种相似度进行线性融合,得到句子结构相似度为:

$$[0041] \quad SynSim = \lambda_1 LenSim(Q_1, Q_2) + \lambda_2 DisSim(Q_1, Q_2) + \lambda_3 WordProSim(Q_1, Q_2) \quad (4)$$

[0042] 式中, $\lambda_1, \lambda_2, \lambda_3$ 为权重系数,代表对三种相似度的偏重程度。

[0043] 进一步地,所述计算统计句子语义相似度具体为:

[0044] 基于概念语义相似度计算句子语义相似度,所述概念语义相似度计算包括基于距离的概念语义相似度计算、基于属性的概念语义相似度计算和基于主成分分析的概念语义相似度混合计算。

[0045] 进一步地,所述基于距离的概念语义相似度计算,具体为:

[0046] 考虑领域本体类的层次关系,利用语义重合度、语义层次深度、语义距离和语义距离多种因素,来计算概念的语义相似度;其中,

[0047] 语义重合度计算公式为:

$$[0048] \quad Sim_{overlap}(a, b) = \frac{ParentNum(a) \cap ParentNum(b)}{Max(ParentNum(a), ParentNum(b))} \quad (5)$$

[0049] 式中, ParentNum(i) 表示i的父节点个数;分子表示a和b相同父节点的个数;分母表示a和b父节点个数较多的那个;

[0050] 语义层次深度是指概念在结构树中所处层次的深度;两个概念的语义相似度与他们深度之和呈正相关关系,与他们的深度之差呈负相关关系;假设根节点R的深度为1, i节点的父节点用Parent(i)表示,则非根节点i在本体结构树中的深度公式为:

$$[0051] \quad Depth(i) = Depth(Parent(i)) + 1 \quad (6)$$

[0052] 本体树T的深度为:

$$[0053] \quad Depth(T) = Max(Depth(i)) \quad (7)$$

[0054] 即本体树的深度取树中节点的最大深度；

[0055] 语义距离对应的相似度公式为：

$$[0056] \quad Sim_{Dis}(a, b) = \frac{2Depth(T)}{2Depth(T) - Dis(a, b)} \quad (8)$$

[0057] 式中, $Dis(a, b)$ 表示节点a和b在本体结构树中的最短路径边数；

[0058] 语义密度对应的相似度公式如下：

$$[0059] \quad Sim_{Den}(a, b) = \frac{BrotherNum(a) + BrotherNum(b)}{2Max(BrotherNum(a), BrotherNum(b))} \quad (9)$$

[0060] 式中, $BrotherNum(i)$ 为节点i的兄弟节点的个数；

[0061] 基于公式 (5)、公式 (6)、公式 (7)、公式 (8) 和公式 (9)，得到基于距离的概念语义相似度的计算模型：

$$[0062] \quad Sim_1(a, b) = \alpha Sim_{Overlap}(a, b) - \beta Sim_{Dis}(a, b) - \gamma Sim_{Den}(a, b) \quad (10)$$

[0063] 式中, α, β, γ 为权重系数。

[0064] 进一步地, 所述基于属性的概念语义相似度计算的模型如下：

$$[0065] \quad Sim_2(c, d) = \alpha Properties(c \cap d) - \beta Properties(c - d) - \gamma Properties(d - c) \quad (11)$$

[0066] 其中, $Properties(c \cap d)$ 表示概念c和d所拥有的公共属性的集合, $Properties(c - d)$ 表示概念c拥有而概念d不拥有的属性集合, $Properties(d - c)$ 则表示概念d拥有而概念c不拥有的属性集合。

[0067] 进一步地, 所述基于主成分分析的概念语义相似度混合计算, 具体为：

[0068] 基于主成分分析法对基于距离因素和基于属性因素计算的数据进行分析, 动态计算出各个因素的权重；计算出权重后对多因素的计算结果进行加权线性相加得到最终概念语义相似度；其过程包括：

[0069] 将基于距离的相似度和基于属性的相似度作为两个维度, 通过多个样本的计算得到相似度矩阵作为输入样本矩阵；

[0070] 对所述输入样本矩阵进行矩阵标准化变换为标准矩阵Z, 并求出相关系数矩阵R1；

[0071] 求出相关系数矩阵R1的特征方程的32个特征根以确定主成分；

[0072] 解方程组单位特征向量；

[0073] 转换标准化后的指标变量为主成分；

[0074] 将两个主成分进行加权并线性求和, 即得最终概念语义相似度值, 相应的权值为各个主成分的贡献率。

[0075] 进一步地, 设句子 Q_1 和 Q_2 分别含有m个概念 $W_{11} \dots W_{1m}$ 和n个概念 $W_{21} \dots W_{2n}$ ；

[0076] 所述基于概念语义相似度计算句子语义相似度, 具体为：

[0077] 从 Q_1 中选出一个概念分别与 Q_2 中的n个概念, 计算概念语义相似度, 直至 Q_1 中概念循环完为止, 得到概念语义相似度矩阵：

$$[0078] \quad T = \begin{matrix} S(W_{11}, W_{21}) & \dots & S(W_{11}, W_{2n}) \\ \vdots & \ddots & \vdots \\ S(W_{1m}, W_{21}) & \dots & S(W_{1m}, W_{2n}) \end{matrix} \quad (12)$$

[0079] 计算第一个概念集与第二个概念集的概念语义平均最大相似度

$$[0080] \quad avgS_{Q_1} = \frac{\sum_{i=1}^n Max(S(W_{1i}, W_{21}), S(W_{1i}, W_{22}), \dots, S(W_{1i}, W_{2n}))}{n} \quad (13)$$

[0081] 计算第二个概念集与第一个概念集的概念的平均最大相似度

$$[0082] \quad avgS_{Q_2} = \frac{\sum_{j=1}^n Max(S(W_{11}, W_{2j}), S(W_{12}, W_{2j}), \dots, S(W_{1m}, W_{2j}))}{n} \quad (14)$$

[0083] 计算公式 (13) 和 (14) 结果, 求平均值, 得到句子语义相似度

$$[0084] \quad SemSim = \frac{avgS_{Q_1} + avgS_{Q_2}}{2} \quad (15)$$

[0085] 所述用户提问的问句与预设语料库中间句的句子相似度定义如下:

$$[0086] \quad Sim = (1-\eta) SynSim + \eta SemSim \quad (16)$$

[0087] 其中, η 为权重系数, 当 η 为 1 时, 表示只使用语义相似度; 当 η 为 0 时表示只使用句子结构相似度。

[0088] 本发明的上述技术方案的有益效果如下:

[0089] 本发明通过将医学领域用户常见问题与语义分析, 语义相似度计算, 本体等相关技术结合, 提出了一种综合的语义相似度计算方法, 实现了对现有计算方法的整合, 有效提高了语义相似度计算的准确率; 并在此基础上结合浅层语义分析和文本分析技术, 实现了基于本体语义相似度的医学问答方法, 该方法可实现五大类问题 (方式、时间、程度、数量、实体) 处理, 达到 80% 以上准确率。五类问题基本涵盖基础医学领域的学科问题, 较好地实现了系统的功能。

附图说明

[0090] 图1为本发明的基于本体语义相似度的医学问答方法的流程图;

[0091] 图2为语义角色标注效果示意图;

[0092] 图3为候选答案收集模块流程图。

具体实施方式

[0093] 为使本发明要解决的技术问题、技术方案和优点更加清楚, 下面将结合附图及具体实施例进行详细描述。

[0094] 本实施例提供一种基于本体语义相似度的医学问答方法, 本实施例包括:

[0095] 1、系统知识库的构建

[0096] 本实施例中所用到的本体是使用 RDF/RDFS 语言来表示的。而知识库构建是以 PA 结构为基础动态构建的。PA 结构由谓词和语义组成, 根据 PA 结构的这种特点, 本实施例将本体分为事件类本体和论元类本体。其中, 事件类本体对应着 PA 结构中的谓词, 它们描述的是动词性概念。比如改善、缓解……药品、症状这样的概念。事件类本体中有一个通用的事件类, 其他的事件都是他的子类。论元类本体对应着 PA 结构中的论元, 他们描述的是一种名词性概念, 比如药品、症状这样的概念描述。根据 PA 结构的语义的特点, 论元类本体又可分为一些子类: 方式、时间、数量、人称代词、意义、程度等等, 这些可以根据背景本体所涉及到的领域来进行设计。

[0097] 首先是定义的通用事件类。

[0098] <事件, rdf:type, rdfs:Class>

- [0099] 谓词P以及其6个核心语义角色。这些语义都为事件类的属性。
- [0100] <P,rdf:type,rdf:property>
- [0101] <A0,rdf:type,rdf:property>
- [0102]
- [0103] <A5,rdf:type,rdf:property>
- [0104] 接下来是15个附加语义。
- [0105] <ADV,rdf:type,rdf:property>
- [0106]
- [0107] <PSE,rdf:type,rdf:property>
- [0108] 定义语义的约束关系,这里以MNR语义为例。MNR表示的是“方式”语义,因此本实施例规定MNR属性的值域为“方式”类。
- [0109] <MNR,rdf:domain,事件>
- [0110] <MNR,rdf:range,Manner>
- [0111] 这样,根据上面定义的背景知识,就可以构建“通过安神补脑液,患者改善了失眠”了。
- [0112] <安神补脑液,rdf:type,Manner>
- [0113] <事件1,rdf:type,事件>
- [0114] <事件1,P,改善>
- [0115] <事件1,A0,患者>
- [0116] <事件1,A1,失眠>
- [0117] <事件1,Manner,安神补脑液>
- [0118] 这种方式直接,能够涵盖所有的PA结构。对于上文所提到的会见事件来说,如果知识库中有特定的“改善事件”类,它是通用事件的子类,相对于上文定义的通用事件,本实施例将其定义为特定事件,它根据特定的某一个事件进行了一些专门的定义。
- [0119] 例如,“改善事件”包含“改善方”和“被改善方”两个属性,其定义域为“改善事件”类,值域为“药品症状”类。相关的定义如下。
- [0120] <药品症状,rdf:type,rdfs:Class>
- [0121] <改善事件,rdfs:subClassOf,事件>
- [0122] <改善方,rdf:type,rdf:property>
- [0123] <改善方,rdf:range,药品症状>
- [0124] <改善方,rdf:domain,改善事件>
- [0125] <被改善方,rdf:type,rdf:property>
- [0126] <被改善方,rdf:range,事物的发展>
- [0127] <被改善方,rdf:domain,改善事件>
- [0128] 相比通用的事件类表现方式,这种方式更能体现特定事件的特点,所以本发明采用的是该方式。
- [0129] 2、问句预处理
- [0130] 问题分类模块定义问题结构、类型、枚举方式。系统对问题进行分词和词性标注处理后,该模块把问题结构分为:热词、问词、映射词。

[0131] 2.1分词和词性标注

[0132] 以“通过安神补脑液,患者改善了失眠”为例,词法分析后得到的结构如图2所示,可以看到句子被分成了一个一个的词,每个词的词性也识别了出来。

[0133] 2.2语义角色标注

[0134] 语义角色标注(Semantic Role Labeling,SRL)是一种基于PA结构的浅层语义分析方法,它采用“谓语-动词-角色”的结构形式,标注句法成分为给定谓语动词的语义角色,每个语义角色被赋予一定的语义含义。如施事、受事、时间和地点等。对于上面的“改善”例子,语义角色标注的结果如图2所示。

[0135] 其中谓词“改善”。“患者”是它的施事(一般用A0表示),而“失眠”则是它的受事(一般用A1表示),安神补脑液是它的方式(一般用MNR表示)

[0136] PA结构中的核心的语义角色为A0-5六种,A0通常表示动作的施事,A1通常表示动作的影响等,A2-5根据谓语动词不同会有不同的语义含义。其余的15个语义角色为附加语义角色,如MNR表示方式,TMP表示时间等。

[0137] 3、基于本体语义相似度的医学问答

[0138] 本实施例的基于本体语义相似度的医学问答流程如图1所示,包括:

[0139] S101,根据预设词库对用户提问的自然语言问句进行分词处理;

[0140] S102,对于分好词的问句根据词的类型,将属于停用词库的词删除

[0141] S103,对于分词后的问句中的专业词汇采用基于本体的概念相似度方法计算词语语义相似度,对于其他词汇采用知网结合本体计算词语语义相似度;

[0142] S104,根据计算出的词语语义相似度,计算统计句子语义相似度和句子结构相似度,并将计算出的句子语义相似度和句子结构相似度进行综合,计算出用户提问的问句与预设语料库中间句的句子相似度;

[0143] S105,根据计算出的句子相似度,从预设问题库中抽取句子相似度大于或等于预设相似度阈值的问句及其答案作为用户提问问题的答案。

[0144] 进一步地,上述S101具体为:根据专业词库、常用词库和同义词库对用户输入的自然语言问句通过正向最大匹配的方法进行分词处理,并将其中的未登记词借助于中科院的分词工具添加到预设词库中。

[0145] 上述S105具体为根据计算出的句子相似度,从预设问题库中抽取句子相似度较高(对相似度大于或等于预设相似度阈值的问句及其答案进行排序,选取前五个)的问句及其答案作为用户提问问题的答案。

[0146] 下面,对基于本体的句子相似度计算过程进行阐述,首先需要说明的是,句子信息的完整表达,需要依赖于组成句子的词语信息,词语的语义信息和句子的结构信息。为此,进行句子相似度计算时从句子的句法结构、组成句词语的语义两个方面进行研究,最大程度的将一个句子所包含的信息挖取出来,用于计算两个句子的相似度计算中。以这种思想为指导,本发明提出了综合特征的句子相似度计算方法。该计算方法主要从组成句子的词序,词距,句长,语义信息等方面进行相似度研究,以下是具体介绍:

[0147] 3.1、句子结构相似度计算

[0148] 句子的结构相似度计算采用基于规则的方法,该方法结合词距,句长,同时考虑词性的作用,综合三个方面确定句子的结构相似度;其中,

[0149] 词距相似度的公式如下：

$$[0150] \quad DisSim(Q_1, Q_2) = 1 - \frac{|SameDis(Q_1) - SameDis(Q_2)|}{Dis(Q_1) + Dis(Q_2)} \quad (1)$$

[0151] 式中, SameDis (Q_i) 表示Q₁和Q₂中相同关键词在Q_i中的距离, 若关键词重复出现, 则以最大距离为准, Dis (Q_i) 表示句子关键词中最左及最右关键词之间的距离, 若关键词重复出现, 则以最小距离为准;

[0152] 句长相似度的公式如下：

$$[0153] \quad LenSim(Q_1, Q_2) = 1 - \frac{|Len(Q_1) - Len(Q_2)|}{Len(Q_1) + Len(Q_2)} \quad (2)$$

[0154] 式中, Len (Q₁) 和Len (Q₂) 分别表示两个句子的长度; 句长在一定程度上反应了两个句子的相似程度, 句子长度相差越小, 相似的程度越大。

[0155] 分词完成后得到关键词集, 从词汇属性方面, 可能存在医学领域专业词、方言词和普通常用词, 比如, “安神补脑液可以改善失眠的状况”, 很显然, 医学领域专业词“安神补脑液”比普通常用词“状况”、承载了更多的信息量, 因此, 医学领域专业词将被赋予更大的权值。从词性方面, 这些词当中可能包含有名词、动词、形容词、数词等多种词形, 而经过大量的实践发现, 名词和动词占整个句子的信息量比重最大, 即句子所表达的中心信息是围绕名词和动词来展开的, 且名词比动词更为重要。所以, 本实施例从词性和词汇属性两方面, 对分词后的关键词进行加权, 对不同的词性赋予对应的权值, 增加准确度。

[0156] 词性相似度的公式如下：

$$[0157] \quad WordProSim(Q_1, Q_2) = \frac{SameWordPro(Q_1, Q_2)}{WordPro(Q_1) + WordPro(Q_2) - SameWordPro(Q_1, Q_2)} \quad (3)$$

[0158] 式中, WordPro (Q₁) 和WordPro (Q₂) 分别表示句子Q₁和Q₂中关键词的权值, SameWordPro (Q₁, Q₂) 表示句子Q₁和Q₂中相同关键词的权值; 当句子Q₁和Q₂中关键词全都相同时, 分母达到最小值;

[0159] 结合上述常见的三种基于规则的表层相似度, 即句长相似度、距离相似度, 以及对关键词词性进行考虑的词性相似度, 见公式 (1) (2) (3), 将三种相似度进行线性融合, 得到句子结构相似度为:

$$[0160] \quad SynSim(Q_1, Q_2) = \lambda_1 LenSim(Q_1, Q_2) + \lambda_2 DisSim(Q_1, Q_2) + \lambda_3 WordProSim(Q_1, Q_2) \quad (4)$$

[0161] 式中, $\lambda_1, \lambda_2, \lambda_3$ 为权重系数, 代表对三种相似度的偏重程度, 其值可由本领域技术人员根据经验评估或依据实际需求预设。

[0162] 3.2、概念语义相似度计算

[0163] 目前, 基于本体的语义相似度计算方法研究已经形成了丰富的研究成果, 针对本体语义相似度计算中两个主要影响因素: 概念属性和概念距离, 分析传统计算方法中的不足, 研究者一般将基于本体的语义相似度计算方法划分为基于距离的语义相似度计算、基于属性的语义相似度计和混合语义相似度计算。

[0164] 3.2.1 基于距离的概念语义相似度计算:

[0165] 基于距离的概念语义相似度计算模型考虑领域本体类的层次关系, 利用语义重合度、语义层次深度、语义距离和语义距离多种因素, 来计算概念的语义相似度; 其中,

[0166] 语义重合是指在领域本体树形结构中, 两个节点的父节点中为公共节点的个数;

也可以映射为概念级别:包含相同上位概念的数量。语义重合度是度量两个概念的相同程度。在计算中,实例i的父节点数是从实例i进行向上溯源直到根节点所遍历的节点数。计算公式为:

$$[0167] \quad Sim_{overlap}(a, b) = \frac{ParentNum(a) \cap ParentNum(b)}{Max(ParentNum(a), ParentNum(b))} \quad (5)$$

[0168] 式中, ParentNum(i) 表示i的父节点个数;分子表示a和b相同父节点的个数;分母表示a和b父节点个数较多的那个;

[0169] 语义层次深度是指概念在结构树中所处层次的深度;两个概念的语义相似度与他们深度之和呈正相关关系,与他们的深度之差呈负相关关系;假设根节点R的深度为1,i节点的父节点用Parent(i)表示,则非根节点i在本体结构树中的深度公式为:

$$[0170] \quad Depth(i) = Depth(Parent(i)) + 1 \quad (6)$$

[0171] 本体树T的深度为:

$$[0172] \quad Depth(T) = Max(Depth(i)) \quad (7)$$

[0173] 即本体树的深度取树中节点的最大深度;

[0174] 语义距离是指在本体结构树中连接两个概念节点的通路中最短路径边数Dis(a, b)来表示节点a和b在本体结构树中的最短路径边数。Dis(a, b)越大,两个概念的相似度越低;反之,相似度越高。语义距离对应的相似度公式为:

$$[0175] \quad Sim_{Dis}(a, b) = \frac{2Depth(T)}{2Depth(T) - Dis(a, b)} \quad (8)$$

[0176] 式中, Dis(a, b)表示节点a和b在本体结构树中的最短路径边数;

[0177] 语义密度是指概念的所有同位节点,或叫兄弟节点的个数。在本体树中,局部节点密度越大,说明该节点对应的概念的细化程度越大,因此,对应的相似度就越高。用表示节点i他的兄弟节点的个数。对应的相似度公式如下:

$$[0178] \quad Sim_{Den}(a, b) = \frac{BrotherNum(a) + BrotherNum(b)}{2Max(BrotherNum(a), BrotherNum(b))} \quad (9)$$

[0179] 式中, BrotherNum(i)为节点i的兄弟节点的个数;

[0180] 综合考虑以上四个影响因素分别对应的语义相似度算法,得到基于距离的概念语义相似度的计算模型:

$$[0181] \quad Sim_1(a, b) = \alpha Sim_{overlap}(a, b) - \beta Sim_{Dis}(a, b) - \gamma Sim_{Den}(a, b) \quad (10)$$

[0182] 式中, α, β, γ 为权重系数,其值可由本领域技术人员根据经验评估或依据实际需求预设。

[0183] 3.2.2基于属性的概念语义相似度计算模型

[0184] 在比较现实世界中的两个事物时,通常是比较事物他们的属性,如果他们的属性相同的个数越多,也就说明这两个事物越相似。正是这种思想成为了基于属性的语义相似度计算模型的基本原理:本体概念的特征通过概念属性来表明,基于属性的计算方法利用统计概念所具有的公共属性的个数来完成概念的语义相似度计算。概念的相似度与概念拥有的公共属性个数成正比。Tversky提出了最经典的基于属性语义相似度计算方法,计算模型如下:

$$[0185] \quad Sim_2(c, d) = \alpha Properties(c \cap d) - \beta Properties(c - d) - \gamma Properties(d - c) \quad (11)$$

[0186] 其中, $\text{Properties}(c \cap d)$ 表示概念 c 和 d 所拥有的公共属性的集合, $\text{Properties}(c - d)$ 表示概念 c 拥有而概念 d 不拥有的属性集合, $\text{Properties}(d - c)$ 则表示概念 d 拥有而概念 c 不拥有的属性集合。基于属性的计算方法是一种模拟人类认知和辨识现实生活中各类事物的过程, 且需事物属性的详细信息。

[0187] 3.2.3 基于主成分分析的概念语义相似度混合计算

[0188] 很多研究者提出了一更加有效的混合式算法, 综合多因素的计算结果判定各个因素权并进行线性相加得到一个综合的相似度结果在一定程度上提高了计算结果的准确性。但综合加权过程中的权值是由本领域的专家确定的, 存在主观性、不准确性而且对不同的本体不具有适应性。

[0189] 针对权值精确判定和综合计算进行研究, 本发明采用一种利用数学中的主成分分析法 (PCA) 并进行相应改进, 对通过各个因素计算的数据进行分析, 计算各个因素权重, 是一种客观赋权方法。计算出权值后对多因素计算结果进行加权线性相加得到最终相似度。

[0190] PCA 是一种多元统计方法, 通过损失较少信息将多个指标转化成几个综合指标。通常将转化生成的综合指标定义为主成分, 相比于原始变量, 主成分更加优越, 因为每个主成分都是原始变量的线性组合, 并且各主成分间互不相关。在 PCA 中个主成分的权重不是人为确定。而是根据其贡献率来分配的, 这样更好的实现了对多因素分析中人为确定权值的不足, 保证了结果的客观性、合理性和准确性。

[0191] 本发明采用主成分分析法动态计算权值主要是利用主成分分析法计算出各因素的贡献率作为权值。原始主成分分析法是按累计贡献率大于设定阈值来确定主成分。本文提出两个因素: 距离、属性都要作为主成分, 可以忽略这个环节, 提高算法的效率。基于 PCA 的动态权值计算方法主要思想如下:

[0192] (1) 将基于距离的相似度和基于属性的相似度作为两个维度, 通过多个样本的计算得到相似度矩阵作为输入样本矩阵;

[0193] (2) 对所述输入样本矩阵进行矩阵标准化变换为标准矩阵 Z , 并求出相关系数矩阵 $R1$;

[0194] (3) 求出相关系数矩阵 $R1$ 的特征方程的 32 个特征根以确定主成分;

[0195] (4) 解方程组单位特征向量;

[0196] (5) 转换标准化后的指标变量为主成分;

[0197] (6) 将两个主成分进行加权并线性求和, 即得最终概念语义相似度值, 相应的权值为各个主成分的贡献率。

[0198] 算法如下:

Begin

Input 节点概念 c, d , 公共节点概念 e , 根节点概念 r

Compute $Sim_{overlap}(c, d)$

$Sim_{Dis}(c, d)$

$Sim_{Den}(c, d)$

$Sim_1(c, d)$

[0199]

define α, β, γ

$Sim_2(c, d)$

PCA ($Sim_1(c, d), Sim_2(c, d)$) /*用主成分分析法 PCA 计算三个因素的动态权值*/

$Sim_{total}(c, d)$

End

[0200] 3.3、问句的语义相似度计算

[0201] 3.3.1基于概念语义相似度的句子语义相似度计算

[0202] 设句子 Q_1 和 Q_2 分别含有 m 个概念 $W_{11} \dots W_{1m}$ 和 n 个概念 $W_{21} \dots W_{2n}$;

[0203] (1) 从 Q_1 中选出一个概念分别与 Q_2 中的 n 个概念, 计算概念语义相似度, 直至 Q_1 中概念循环完为止, 得到概念语义相似度矩阵:

$$[0204] \quad T = \begin{matrix} S(W_{11}, W_{21}) & \dots & S(W_{11}, W_{2n}) \\ \vdots & \ddots & \vdots \\ S(W_{1m}, W_{21}) & \dots & S(W_{1m}, W_{2n}) \end{matrix} \quad (12)$$

[0205] (2) 计算第一个概念集与第二个概念集的概念语义平均最大相似度

$$[0206] \quad avgS_{Q_1} = \frac{\sum_{i=1}^n \max(S(W_{11}, W_{21}), S(W_{11}, W_{22}), \dots, S(W_{11}, W_{2n}))}{n} \quad (13)$$

[0207] (3) 计算第二个概念集与第一个概念集的概念的平均最大相似度

$$[0208] \quad avgS_{Q_2} = \frac{\sum_{j=1}^n \max(S(W_{11}, W_{2j}), S(W_{12}, W_{2j}), \dots, S(W_{1m}, W_{2j}))}{n} \quad (14)$$

[0209] (4) 计算公式 (13) 和 (14) 结果, 求平均值, 得到句子语义相似度

$$[0210] \quad SemSim = \frac{avgS_{Q_1} + avgS_{Q_2}}{2} \quad (15)$$

[0211] 3.3.2问句语义相似度计算模型

[0212] 针对医学领域特点, 综合考虑多重信息, 包括句子结构相似度和句子语义相似度, 将相似度算法进行融合, 最后定义两个句子的相似度为:

$$[0213] \quad Sim = (1-\eta) SynSim + \eta SemSim \quad (16)$$

[0214] 其中, η 为权重系数, 当 η 为1时, 表示只使用语义相似度; 当 η 为0时表示只使用句子结构相似度, 其值可由本领域技术人员根据经验评估或依据实际需求预设。

[0215] 4、候选答案提取模块实现

[0216] 根据分词模块处理得出的关键字组合, 利用网络检索相关文本信息。系统通过证据评分模型评判证据的问题支持度, 然后按照从高分到低分顺序传递给候选答案提取模

块。

[0217] 答案提取模块主要评价候选答案的可信度,通过抽取函数提取证据中的有效关键词,再利用候选答案评分模型计算每个候选答案的分值,只留下排名前三的候选答案再传递给用户展示界面。

[0218] 提取函数提取的关键字即可能的答案,一些提取函数寻找的目标样例(p 是文本, ψ 是中心词, f 是提取函数)如表1所示:

[0219] 表1提取函数目标示例表

	Extraction function	Example of criteria
[0220]	<i>definition</i> (ρ, φ)	Hypernyms of φ
	<i>specialisation</i> (ρ, φ)	Hyponyms of φ
	<i>cardinality</i> (ρ, φ)	Pattern: NUMBER φ
	<i>measure</i> (ρ, φ)	Pattern: NUMBER UNIT φ
	<i>attribute</i> (ρ, φ)	Various patterns
	<i>person</i> (ρ)	<PERSON> entities
	<i>time</i> (ρ)	<TIME> entities
	<i>location</i> (ρ)	<LOCATION> entities
	<i>reason</i> (ρ)	Not implemented for TREC
	<i>manner</i> (ρ)	Not implemented for TREC
	<i>object</i> (ρ)	Any noun phrase

[0221] 提取多个候选答案后,由于中文语言处理技术不够完善,关键词里夹杂无关文本,致使候选答案质量参差不齐。因此需要对所有候选答案进行评分。本系统共使用七种候选答案评分模型。系统根据问题分类及证据的分词结果选择一种评分模型,并随机赋予它一个值(一般是1),或者使用综合评分模型,进行简单线性加权得出最终分值。最后按从高到低的顺序得出其置信度。该模块的工作流程如图3所示。

[0222] 本实施例通过将医学领域用户常见问题与语义分析,语义相似度计算,本体等相关技术结合,提出了一种综合的语义相似度计算方法,实现了对现有计算方法的整合,有效提高了语义相似度计算的准确率;并在此基础上结合浅层语义分析和文本分析技术,实现了基于本体语义相似度的医学问答方法,该方法可实现五大类问题(方式、时间、程度、数量、实体)处理,达到80%以上准确率。五类问题基本涵盖基础医学领域的学科问题,较好地实现了系统功能。

[0223] 此外,需要说明的是,本领域内的技术人员应明白,本发明实施例的实施例可提供为方法、装置、或计算机程序产品。因此,本发明实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0224] 本发明实施例是参照根据本发明实施例的方法、终端设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生

用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0225] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上,使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0226] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例做出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明实施例范围的所有变更和修改。

[0227] 还需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排除性的包含,从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

[0228] 以上所述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,在不脱离本发明所述原理的前提下,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。

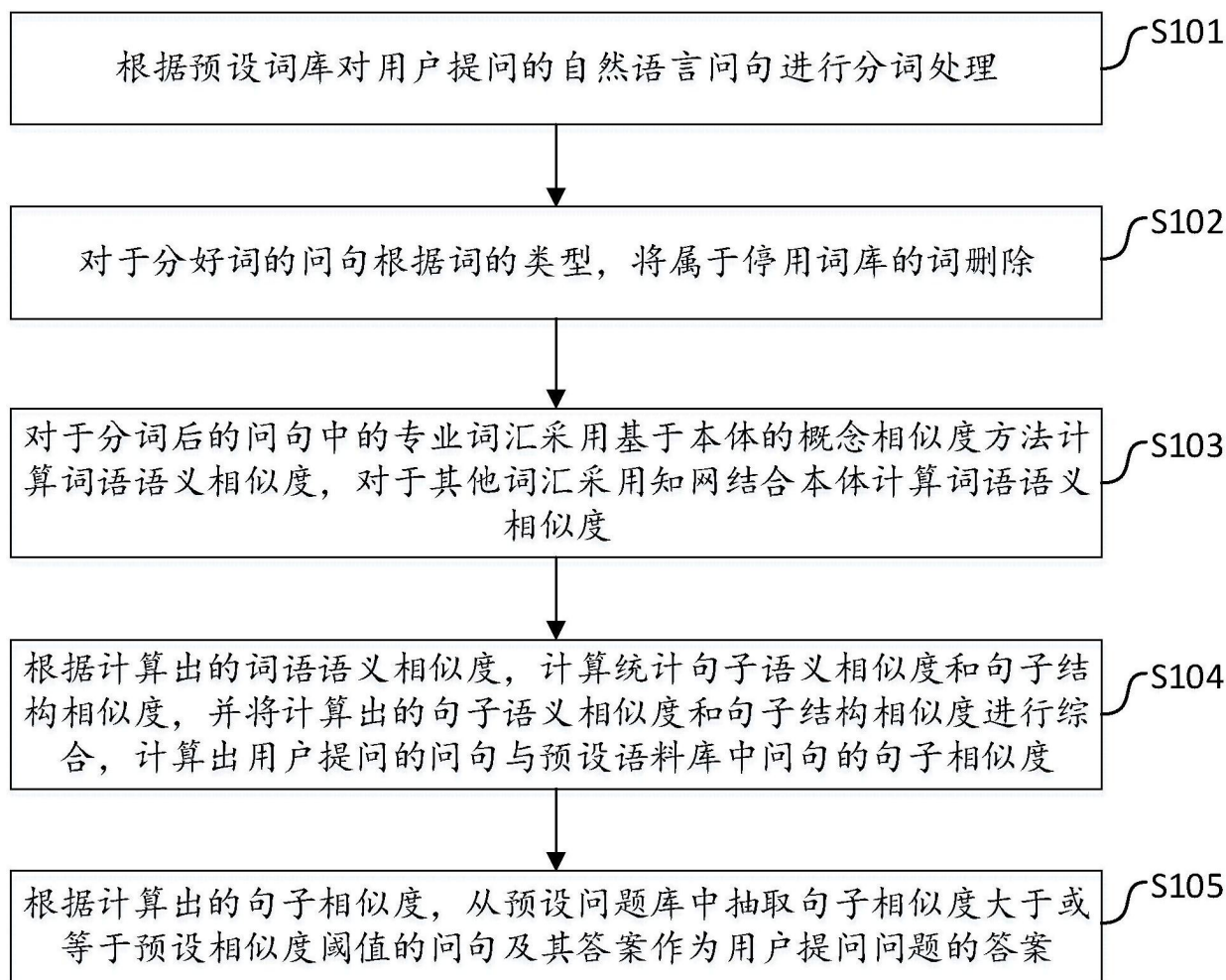


图1



图2

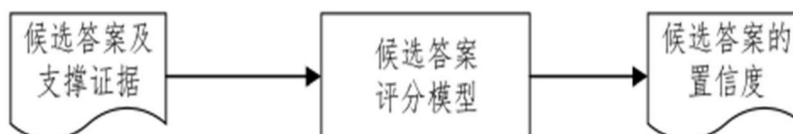


图3