

基于HNC理论和依存句法的句子相似度计算

吴佐衍, 王 宇

WU Zuoyan, WANG Yu

大连理工大学 管理科学与工程学院, 辽宁 大连 116024

School of Management Science and Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

WU Zuoyan, WANG Yu. New measure of sentences similarity based on hierarchical network of concepts theory and dependency parsing. *Computer Engineering and Applications*, 2014, 50(3): 97-102.

Abstract: Sentences similarity is one of the several important tasks in natural language processing. A novel measure method based on Hierarchical Network of Concepts(HNC) theory and dependency parsing in natural language processing is put forward to compute the semantic similarity. The meaning of a sentence is made up of the meanings of its individual words and the structural way the words are combined, and semantic information is obtained from HNC theory, and the syntactic information is obtained through a deep parsing process, and it is compared with the current popular methods and subjective judgment of human. Experiment shows that the method has a good performance, which can distinguish the differences between different sentences more accurately.

Key words: Hierarchical Network of Concepts(HNC); dependency parsing; sentence similarity; natural language processing

摘 要: 句子相似度计算是自然语言处理的重要研究内容。运用自然语言处理的概念层次网络(HNC)理论和依存句法理论提出一种句子相似度的计算方法。该方法认为句子的相似度是由词语的语义相似度和句法结构相似度共同决定的, 利用HNC理论词汇层面联想的概念表述体系来计算词语之间的相似度, 利用依存句法理论来获取句子中词语的词语搭配和构成特征, 与现有典型的句子相似度算法和人工判断进行了比较。实验结果表明, 该方法能够较好地反应句子之间的语义差别, 是一种可行有效的方法。

关键词: 概念层次网络; 依存句法; 句子相似度; 自然语言处理

文献标志码: A **中图分类号:** TP391 **doi:** 10.3778/j.issn.1002-8331.1203-0719

1 引言

随着信息资源的爆炸性增长, 如何获取有效信息成为关注焦点, 而信息资源80%是以自然语言形式存在的文本, 如何获取文本中有效的信息是信息处理的目标和难题。自然语言处理是解决这个问题的关键技术, 而句子相似度计算是自然语言处理领域的重要研究内容, 其广泛应用于文本摘要、自动问答系统和机器翻译中^[1]。

目前计算句子相似度的方法可分为两种: 一种是基于词语特征的方法, 典型的方法有向量空间模型^[2], 基于词性和次序的方法^[3], 这类方法将句子看成是词语的线性序列, 利用句子的表层信息来计算相似度; 另一种是基于语义和句法分析的方法, 其又可分为基于语义字

典的方法和语义依存的方法, 前者利用词语的语义相似度得出句子的相似, 例如, 国内外学者使用 WordNet、HowNet 和同义词词林提出相应的算法^[4-6]; 后者的典型方法是利用句法分析获取句子中各词语间语义依存^[7-8]。

基于词语特征的方法只用了句子的表层信息, 不能很好处理不同词汇具有相似或相同语义的句子; 而基于语义的方法运用了词语的语义信息, 一定程度上解决了基于特征方法的不足, 但这种方法的准确性很大程度上依赖于语义字典的完备性, 同时也忽略了词语之间的相互作用关系和句子的深层句法结构, 例如, 相同词语的不同组合来表达不同含义的句子; 基于语义依存的方法运用句子的深层语法信息, 获取句子的组织结构和词语

基金项目: 国家自然科学基金(No.70890083); 教育部人文社科基金(No.09YJA870005)。

作者简介: 吴佐衍(1988—), 男, 硕士研究生, 主要研究方向: 自然语言处理; 王宇(1959—), 男, 博士, 教授, 主要研究方向: 自然语言处理、文本挖掘、信息资源管理。E-mail: wuzy_oye@yahoo.com.cn

收稿日期: 2012-03-30 **修回日期:** 2012-06-28 **文章编号:** 1002-8331(2014)03-0097-06

基于以上分析,为了解决电子商务网站商品评价信息抽取、标题党识别和文本复制检测等问题,提出利用句子词语的语义相似度和句法特征相似度来计算两个句子间的相似度。为了更好地反映词语间语义相似度,本文引入中科院黄曾阳先生的HNC理论,利用该理论的词汇层面概念联想脉络可以非常容易得出词语之间的语义相关性及其量化的数值^[9];并利用依存句法理论来获取句子的深层句法信息。

2.1 依存句法

- (1) 一个句子中只有一个成分是独立的。
- (2) 其他成分直接依存于某一成分。
- (3) 任何一个成分都不能依存于两个或两个以上的成分。

(5)中心成分左右两边的其他成分相互不发生关系。

HNC理论是以语义表达为基础,融语义、语法、语用为一体自然语言处理体系,通过词汇和语句两个联想脉络来“帮助”计算机理解自然语言^[12]。该理论建立词汇层面的概念表述体系,这个体系把概念分为抽象概念和具体概念。抽象概念的外部特征,内涵分别使用五元组和语义网络来表达;对具体概念采取挂靠展开近似表达方法。语义网络是树状的分层结构,树的每个节点代表一个概念,即概念基元,每棵树的节点概念形成一个概念聚类,每个子树节点概念形成一个子类^[9]。词语的语义是通过HNC理论的概念基元、五元组和组合符号来表达,该理论对自然语言概念的符号化表述可以一般化为^[9]: $\Sigma\{\text{类别符号串}\}\{\text{层次符号串}\}\{\text{组合结构符号}\}$

<本体层符号> ::= 500|52|53|9|c|6y (y=0~5)

词语相似度采用文献[13]的思想,两个词语的相似度等于其所有义项之间相似度的最大值,即

$$\text{sim}(w_1, w_2) = \max(\text{sim}(\text{hnc}_{1i}, \text{hnc}_{2j})) \quad (2)$$

概念表示是通过概念基元、概念类别符号和五元组符号按照一定的规则组合而成,因此两个概念表示 cr_1 , cr_2 的相似度为:

$$\text{Sim}(cr_1, cr_2) = \delta_1 \text{Sim}(cp_1, cp_2) + \delta_2 \text{Sim}(cc_1, cc_2) + \delta_3 \text{Sim}(ep_1, ep_2) \quad (3)$$

其中, $\text{Sim}(cp_1, cp_2)$, $\text{Sim}(ep_1, ep_2)$, $\text{Sim}(cc_1, cc_2)$ 分别表示概念基元、概念外部特征和概念类别的相似度,它们的计算方式将在下文 3.1~3.2 节中给出;参数 $\delta_1 + \delta_2 + \delta_3 = 1$, 且 $\delta_1 > \delta_2 > \delta_3$, $\delta_1, \delta_2, \delta_3$ 分别为对应计算的权重。

3.1 概念基元相似度

概念基元是概念层次树中的节点,是语义描述的基本单位,且概念层次树具有良好的树状结构,因此,根据该树的特点,利用语义重合度、语义距离和层次深度来计算概念基元间的相似度,即处于同一概念层次树下的任意两个概念基元 cp_1 , cp_2 的语义相似度为:

$$\text{Sim}(cp_1, cp_2) = \frac{\beta_1 \times \text{dept}(cp_p) + \frac{1}{2} \beta_2}{\text{dist}(cp_1, cp_2) + \beta_1 \times \text{dept}(cp_p) + \beta_2 \max(|\text{hc}(cp_1, cp_2)|, \frac{1}{2})} \quad (4)$$

其中,参数 $\beta_1 > 1 > \beta_2 \geq 0$, 且 β_1, β_2 分别表示两个概念基元的语义重合度和层次深度差对概念基元相似度的影响; $\text{dept}(cp_p)$ 是概念基元的层次深度; $\text{dist}(cp_1, cp_2)$ 为概念基元语义距离; $\text{dist}(cp_1, cp_2) + \beta_1 \times \text{dept}(cp_p)$ 说明了两个概念基元层次深度和;而 $\text{hc}(cp_1, cp_2) = \text{dept}(cp_1) - \text{dept}(cp_2)$ 为它们的层次深度差;分子 $\frac{1}{2} \beta_2$ 是归一化的需要。

3.2 概念外部特征相似度

抽象概念需要从动态、静态、属性、值和效应五个侧面加以表达,这就是抽象概念的五元组特性,简记为: (v, g, u, z, r) , 它们是抽象概念外在多元性表现的基元^[8]。例如,“思考和思维”就是从五元组的 vg, g 侧面对同一概念内涵的表达。两个概念表示的五元组集合分别为 ep_1 和 ep_2 , 则两个概念表示外部特征的相似度为:

$$\text{Sim}(ep_1, ep_2) = \begin{cases} 1, & ep_1 \cap ep_2 \neq \emptyset \\ 0, & ep_1 \cap ep_2 = \emptyset \end{cases} \quad (5)$$

其中,五元组集合 $ep_i \subseteq \{u, g, v, z, r\}$ 。

3.3 概念类别相似度

HNC 理论将概念分为抽象概念、具体概念和两可概念。两可概念兼具有抽象概念和具体概念的特征。对抽象概念设置了五种类别基元:基元概念、基本概念、语言逻辑概念、“语法”概念和综合类概念,分别用符号 \emptyset, j, l, f, s 表示;对具体概念设置两个概念基元:人和物,

用符号 p 和 w 表示;以及兼具有抽象概念和具体概念双重特性的物性概念 x ^[8]。概念类别是通过语义网络符号和五元组的组合来表示,其相似度为:

$$\text{Sim}(cc_1, cc_2) = \begin{cases} \frac{1}{2}, & cc_1, cc_2 \text{ 都为抽象概念或具体概念} \\ 1, & cc_1, cc_2 \text{ 都为具体概念人、事或物} \\ 0, & \text{其他} \end{cases} \quad (6)$$

其中, cc_i 表示概念类别,具体概念的网络符号中 p 是人, jw 和 w 是物,其他五元组的 g, r 是事。

4 句子相似度计算

本文提出基于 HNC 和语法分析的句子相似度计算的主要思想:句子的相似度不仅受句子的词语语义相似度的影响,而且受词语搭配和构成特征的影响,即两个句子的相似度是有句子间的词语相似度和句法相似度共同决定。进行相似度计算时首先利用依存句法理论获取句子的句法特征和词语对句子语义表达的重要性,然后利用第 3 章的方法来计算词语之间的相似度。

为了便于下文计算句子间的相似度,给出以下定义。

定义 1(句子相似度) 指两个句子表达语义的相似程度,任意两个句子 S_1, S_2 的相似度 $\text{Sim}(S_1, S_2) \in [0, 1]$, 0 表示不相似,1 表示完全相似,数值越大表示两个句子越相似。

定义 2(重要词语) 指语义依存树的根节点,即全局核心词,以及与核心词直接相连的名词、动词、形容词和副词的集合,用集合 H 表示。

定义 3(次要词语) 句子中除重要词语外词语,分为实词集合和虚词集合,前者是指句子中除重要词语外的名词、动词、形容词和副词的集合,用 A 表示;后者为除了 H, A 之外的剩余的词语集合,用 E 表示。

为了获取句子中的重要词语、次要词语和它们的依存关系,使用哈尔滨工业大学社会计算与信息检索研究中心研制的依存句法分析器,该分析器能够达到 86% 的正确率^[7]。例如,“经理召开一个市场会议”的依存分析如图 1 所示,其中箭头表示依存关系,词语下方的字母表示词性;其语义依存树为图 2 所示,其中,“召开”为核心词,除核心词外的重要词语有“经理”和“会议”,且前者语法特征为主语,后者为宾语;其余词语为次要词语。

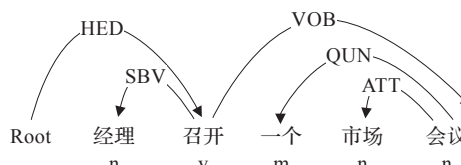


图 1 例句 1 的依存分析

一个完整的汉语句子是由句子的主要成分和修饰成分所构成,而人们往往从关键成分就可以了解一个句子

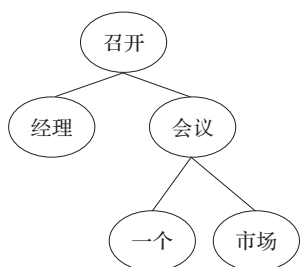


图2 例句1的语义依存树

的大概意思^[7]。基于这个思想,在计算句子相似度时,只考虑重要词语的句法特征,而将次要词语看成是线性的。

通过以上分析,任意两个句子 S_1, S_2 的相似度为:

$$Sim(S_1, S_2) = \alpha_1 Sim(H_1, H_2) + \alpha_2 Sim(A_1, A_2) + \alpha_3 Sim(E_1, E_2) \quad (7)$$

其中, $Sim(H_1, H_2)$, $Sim(A_1, A_2)$, $Sim(E_1, E_2)$ 分别表示重要词语、次要词语中实词和次要词语中虚词的相似度,它们的相似度计算方法将在下文中给出;参数 $0 \leq \alpha_1 \leq \alpha_2 \leq \alpha_3 \leq 1$, 且 $\alpha_1 + \alpha_2 + \alpha_3 = 1$ 。

由于一个句子中并不一定总是存在次要词语,例如,句子“菜牛咬菜农”就没有次要词语。因此,当两个句子的词语集合 A_i 或 E_i ($i=1, 2$) 都为空时,则需对参数 α_j ($j=1, 2, 3$) 进行调整,调整方法如下:

(1) 当集合 A_i 都为空集时,则 $\alpha_2 = 0$, $\alpha_t = \frac{\alpha_t}{\alpha_1 + \alpha_3}$, $t=1, 3$ 。

(2) 当集合 E_i 都为空集时,则 $\alpha_3 = 0$, $\alpha_t = \frac{\alpha_t}{\alpha_1 + \alpha_2}$, $t=1, 2$ 。

(3) 当集合 A_i 和 E_i 都为空集时,则 $\alpha_1 = 1$, $\alpha_t = 0$, $t=1, 2$ 。

4.1 $Sim(H_1, H_2)$ 的定义

两个句子间的相似度很大程度由其重要词语间的相似度决定,而重要词语的相似度是由词语的相似度和词语所担任的语法功能的相似度共同决定的。因此,两个句子 S_1, S_2 的重要词语 H_1 和 H_2 的相似度为:

$$Sim(H_1, H_2) = \frac{1}{p} \sum_{i=1}^p Sim(h_{1i}, h_{2i}) \quad (8)$$

其中, $p = count(H_1 \cup H_2 - H_1 \cap H_2)$, 表示两个句子重要词语在句子中担任不同语法功能的个数,函数 $count()$ 用于计算重要词语担任语法功能个数; h_{1i}, h_{2i} ($0 \leq i \leq p$) 是具有相同语法功能的重要词语;当 $h_{1i} = \emptyset$ 或 $h_{2i} = \emptyset$ 时,令 $Sim(h_{1i}, h_{2i}) = \varepsilon$, ε 为很小的正数,表示其对句子相似度的贡献很小。

4.2 $Sim(A_1, A_2)$ 的定义

句子中的次要词语对句子的语义表达起辅助作用,

考虑到目前句法分析器的准确性和汉语表达形式的多样性,忽略次要词语集合的语法信息,将其看做线性关系。参考文献[14]中计算次要义原相似度的思想来计算次要词语中实词集合 A 的相似度,即采用完备二部图的构造方法,将两个集合分别作为二部图中的两个顶点的集合,建立连接关系,计算过程如下:

(1) 设 a_{1i} ($i=0, 1, \dots, m$), a_{2j} ($j=0, 1, \dots, n$), 分别为集合 A_1, A_2 中的词语,其中: m, n 为集合中词语的个数; a_{1i}, a_{2i} 分别作为二部图中一个顶点。

(2) 利用公式(3)计算词语 a_{1i} 和 a_{2i} 之间的相似度 $Sim(a_{1i}, a_{2j})$, 并将它作为两个顶点边的权重,获得一个关于两个集合中元素相似度的二部图,如图3所示。

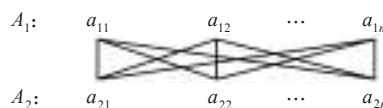


图3 元素相似度二部图

(3) 令 $l = \max(Sim(a_{1i}, a_{2j}))$, 将 l 加入到集合 L 中, 集合 $CS = \{Sim(a_{1i}, a_{2j})\}$; 删除集合中 CS 中与 a_{1i} 和 a_{2i} 有关系的边, 并从集合 A_1, A_2 中删除 a_{1i} 和 a_{2i} 。

(4) 重复步骤(2), 直到集合 CS 为空。

(5) 若集合 A_1, A_2 有一个为非空时, 将这个集合中的元素与空元素 ϕ 对应, 设 $Sim(a, \phi) = \theta$, θ 为很小的正数, 并将其加入到集合 L 中。

(6) 将集合 L 中所有词语相似度的平均值作为集合 A_1, A_2 的相似度, 即

$$Sim(A_1, A_2) = \frac{\sum_{i,j} Sim(a_{1i}, a_{2j})}{\max(m, n)} \quad (9)$$

其中, $Sim(a_{1i}, a_{2j}) \in L$ 。

4.3 $Sim(E_1, E_2)$ 的定义

次要词语中的虚词集合 E 对句子语义信息的完整表达所起的作用很小, 为了提高算法的效率, 这部分的相似度计算采用简化的空间向量模型, 其计算过程如下:

(1) 设集合 E_1, E_2 中的词语分别为: $E_1 = \{e_{11}, e_{12}, \dots, e_{1m}\}$, $E_2 = \{e_{21}, e_{22}, \dots, e_{2n}\}$, 其中: m, n 为集合中词语的个数。

(2) 设集合 $E = E_1 \cup E_2 = \{e_1, e_2, \dots, e_p\}$, 其中: p 为集合 E_1 和 E_2 中出现不同词语的个数。

(3) 构造两个向量 $V_1 = \{v_{11}, \dots, v_{1k}, \dots, v_{1p}\}$ 和 $V_2 = \{v_{21}, \dots, v_{2i}, \dots, v_{2p}\}$, 其中 v_{ij} 的取值是根据集合 E 中的第 j 个词语是否在集合 E_1, E_2 中出现, 即

$$v_{ij} = \begin{cases} 1, & e_{ij} \in E_i \\ 0, & \text{其他} \end{cases}$$

(4)用向量夹角余弦作为集合 E_1, E_2 的相似度,即

$$Sim(E_1, E_2) = \frac{V_1 \times V_2}{\sqrt{(V_1)^2 \times (V_2)^2}} \tag{10}$$

5 实验与结果分析

为了便于验证与分析,选择几个典型的例子对几种经典的算法进行了比较,其中:方法1是文献[3]提出的基于词形和词序的方法;方法2是文献[7]提出的基于语义依存的方法;方法3是文献[5]提出基于知网的方法;方法4为本文的方法;方法5为人工判断的相似度。实验参数设置如表1所示,实验结果如表2所示。

表1 实验参数设置

α_1	α_2	α_3	θ	ε
0.70	0.20	0.10	0.05	0.05

观察表2,可以发现在表的前半部分不同方法之间的计算结果差异较大,体现了它们之间的特性;后半部分的实验结果吻合度较高,体现了它们的共性,即都具有合理性。下面对表前半部分不同方法之间的结果差异进行详细的分析。

方法1基于词形词序的方法由于不考虑词语的语义差别,计算效果较差。例如,表2中的第5,8~10组的实验;方法2基于语义依存的方法由于只计算句子中有效搭配对的相似度,忽略了其余部分词语的语义信息,而且在句子较短时,其有效搭配对数量很少,甚至没有,因而计算结果不够准确。例如,表2中的第3,4,13,17组实验;方法3和方法4都能较好地反映句子中词语的语义差别,例如,表2中的第5,8~10组实验;但是该方法没有考虑句子的深层语法信息,不能很好地处理由相同

词语不同次序表达不同概念句子间的相似度,因而计算结果偏高。例如,表2中的第1~3组实验。方法4不仅考虑了词语的语义信息,而且考虑了重要词语担任的语法信息,能够有效解决前面几个方法面对的难题,实验结果较为合理。

对比表2的第1和第2组实验结果,方法4的结果比方法2小,这也更符合实际,这是由于方法4不但考虑句子的句法结构,而且在计算相似度时考虑了重要词语的语法特征。

对比表2的第7和第10组实验数据,可以发现方法3的实验数据偏大,而方法4的结果较为理想。例如,在第7组实验中,方法3计算出“猪”和“菜牛”的相似度为1;而在方法4中这两个词的相似度小于1,所以两个句子的相似度小于1,这也是选择HNC计算词语相似度的原因之一。

最后,也发现第4,5组实验的结果并不够理想,原因之一是分词时已经不正确,例如,实验的第5组,五百/克/西红柿,符号“/”表示句子切词分隔符,使五百克和一斤不能进行等价换算,同时也是中文分词的难点之一。但方法还是能够更好地贴近实际。

为了更好地比较本文方法的合理性,请大连理工大学信息管理系的8位同学独立给出每个句子间的相似度,去掉明显的离散点取平均值,结果为表2中的方法5。通过对比,可以发现方法4和方法5的实验结果具有较高的一致性,这在一定程度上说明了方法4的有效性和合理性。

6 结束语

本文通过使用HNC理论和依存句法理论来计算句

表2 实验结果

组	句子1	句子2	方法1	方法2	方法3	方法4	方法5
1	菜牛咬菜农	菜农咬菜牛	0.900 000 0	1.00	1.000 000	0.359 722 0	0.63
2	他的菜牛咬了一个老菜农	一个老菜农咬了他的菜牛	0.971 429 0	1.00	1.000 000	0.684 722 0	0.62
3	菜农有四条腿	菜农养的菜牛有四条腿	0.792 308 0	0.75	0.830 688	0.567 567 0	0.59
4	历史是一种财富	历史是财富	0.775 000 0	1.00	0.700 000	0.875 000 0	0.88
5	一斤番茄	五百克西红柿	0	0	0.700 000	0.875 000 0	0.95
6	题目是正确的	题目不是错误的	0.700 000 0	1.00	0.868 954	0.897 429 0	0.87
7	中国位于亚洲	美国位于美洲	0.400 000 0	0.50	0.933 333	0.723 197 0	0.65
8	河边有一头菜牛	河边有一块石头	0.640 000 0	0.75	0.679 608	0.401 118 0	0.48
9	河边有一头菜牛	河边有一群动物	0.640 000 0	0.75	0.743 580	0.599 948 0	0.58
10	河边有一头菜牛	河边有一头猪	0.640 000 0	0.75	1.000 000	0.805 556 0	0.56
11	河边有一头菜牛	河边的动物一定是菜牛	0.327 273 0	0.25	0.505 556	0.238 333 0	0.53
12	历史是一种财富	文明是历史的轨迹	0.360 000 0	0.50	0.534 254	0.420 833 0	0.38
13	他是一个历史教师	他是一个精力旺盛的教师	0.700 000 0	1.00	0.805 767	0.815 503 0	0.62
14	他有一本张爱玲的书	召开会议	0	0	0.131 720	0.119 374 9	0.08
15	河边有一头菜牛	菜刀在哪里	0	0	0.093 068	0.111 806 0	0.12
16	历史是一种财富	历史是一种演变过程	0.754 546 0	0.75	0.800 601	0.758 333 0	0.40
17	召开会议	经理召开一个关于市场计划的会议	0.460 000 0	0.50	0.529 570	0.488 333 0	0.70
18	事故发生后,伤员被及时送到就近医院救治	晚上7时左右,所有伤员被送往了医院	0.384 210 5	0.67	0.486 132	0.553 036 0	0.61

子的相似度,前者能够很好地反映词语之间的语义差别;后者能够得到句子的深层语法信息和词语对句子语义的贡献度,即重要词语和次要词语,并在计算重要词语相似度时结合它们的语法特征;通过这两个理论结合起来有效提高句子相似度计算的准确性。同时实验表明,该方法特别适用于短句间的相似度计算,在今后的研究工作中,考虑将其应用于Web的博客或电子商务网站商品评价信息抽取中。

致谢 本文实验中的方法2和方法4使用了哈工大社会计算与信息检索研究中心的句法分析器;方法3中的词语相似度计算使用了中国科学院计算技术研究所刘群的基于知网的词汇语义相似度计算软件包;方法4中的词语相似度计算使用了中科院声学所的HNC词语知识库。在此表示衷心感谢。

参考文献:

- [1] Lee Ming Che. A novel sentence similarity measure for semantic-basic expert systems[J]. Expert System with Applications, 2011, 38: 6392-6399.
- [2] Dagan I, Lee L, Pereira F. Similarity-based models of word cooccurrence probabilities[J]. Machine Learning, 1999, 34.
- [3] 吕学强,任飞亮,黄志丹,等. 句子相似模型和最相似句子查找算法[J]. 东北大学学报:自然科学版, 2003, 24(6): 531-534.
- [4] Lewis W D. Measuring conceptual distance using WordNet: the design of a metric for measuring semantic similarity[J]. Coyote Papers, 2001, 12.
- [5] 孔胜,王宇. 基于句子相似度的文本主题句提取算法研究[J]. 情报学报, 2011, 30(6).
- [6] 李素建. 基于语义计算的语句相关度研究[J]. 计算机工程与应用, 2002, 38(7): 75-77.
- [7] 李彬,刘挺,秦兵,等. 基于语义依存的汉语句子相似度计算[J]. 计算机应用研究, 2003(12): 15-17.
- [8] Achananuparp P, Hu X, Zhou X, et al. Utilizing sentence similarity and question type similarity to response to similar questions in knowledge-sharing community[C]//Proceedings of QAWeb 2008 Workshop, 2008.
- [9] 黄曾阳. HNC(概念层次网络)理论—计算机理解语言研究的新思路[M]. 北京:清华大学出版社, 1998: 11-43.
- [10] 刘海涛. 依存语法和机器翻译[J]. 语言文字应用, 1997(3): 89-93.
- [11] 郭艳华,周昌乐. 一种汉语语句依存关系网协动生成方法研究[J]. 杭州电子工业学院学报, 2000, 20(4): 24-32.
- [12] 晋耀红. HNC(概念层次网络)语言理解技术及其应用[M]. 北京:科学出版社, 2005: 50-55.
- [13] 刘群,李素建. 基于《知网》的词汇语义相似度计算[C]//第三届汉语词汇语义学研讨会, 台北, 2002: 59-76.
- [14] 葛斌,李芳芳,郭丝路,等. 基于知网的词汇语义相似度计算方法研究[J]. 计算机应用研究, 2010, 27(9): 3329-3333.

(上接35页)

柔性工作流引擎模型。利用工作流开源产品资源,在其基础上进行功能适应改造,增加工作流对跨异构组织的流程运行支持,以及流程实例的动态变更能力支持。本文模型在实际应用已经过了验证,结果证明该模型的可行性。如何支持处理较为复杂的流程定义变更,以及和流程实例变更二者叠加变更所带来的数据变更,是下一步的研究重点内容。

参考文献:

- [1] 刘慧敏,王刊良,杨姝. 跨组织工作流建模及分析研究述评[J]. 管理学报, 2010, 7(3): 468-474.
- [2] 孙华林,赵正文. 基于Web Services的面向服务架构(SOA)的探索与研究[J]. 信息技术, 2007(1): 50-53.
- [3] 付燕宁,刘磊,张家晨. 构造语义Web服务工作流的模型[J]. 吉林大学学报:工学版, 2011, 37(3): 600-605.
- [4] 韦伯,帕拉斯泰迪斯,鲁滨逊. REST 实战(中文版)[M]. 李琨,译. 南京:东南大学出版社, 2011: 9-149.
- [5] 侯志松,余周,冯启高. 工作流管理系统开发实录——基于JBoss jBPM实现[M]. 北京:中国铁道出版社, 2010: 16-23, 109-155.
- [6] 李昂. REST 架构工作流中间件设计与实现[J]. 计算机工程与设计, 2012, 33(9): 3455-3459.
- [7] 李竞杰,王维平,杨峰. 柔性工作流理论方法综述[J]. 计算机集成制造系统, 2010, 16(8): 1569-1577.
- [8] 林永毅,罗锟. 基于实例的动态定义执行工作流引擎研究[J]. 计算机工程与科学, 2007, 29(6): 132-134.