

HNC 问答处理系统关键算法研究^{*}

吴晨^{1,2} 张全²

(中国科学院研究生院 北京 100039)¹ (中国科学院声学研究所 北京 100080)²

摘要 HNC 问答处理系统是一种崭新的问答处理模式,与传统的问答处理系统相比,它把对语言的处理深入到语言概念空间中进行,以求尽可能地减少语言中的同义和多义模糊。本文在介绍 HNC 问答处理系统整体构架的同时,设计和实现了问答系统中的 3 个关键算法,服务于问答系统的不同阶段处理需要。实际应用表明,在这些算法指导下的 HNC 问答处理系统具有良好的问题处理能力。

关键词 中文信息处理,问答系统,语言概念空间,语义相关度,HNC 理论

Research on the Key Algorithms of HNC Question Answering System

WU Chen^{1,2} ZHANG Quan²

(Graduate School of the Chinese Academy of Science, Beijing 100039)¹

(Institute of Acoustics, Chinese Academy of Science, Beijing 100080)²

Abstract HNC Question Answering system(HNCQA) is one of the most brand-new question answering processing pattern. Compared with the traditional QA system, HNCQA penetrate into the language concept space to resolve the understanding problem. Processing in this way can reduce the word sense ambiguity as drastically as possible. This paper will introduce the HNCQA architecture on the whole and present some key algorithms which server the HNCQA in some key positions. The experiment, which we have made, indicates that the system based on the algorithms we designed and provided with a middle scale common-sense knowledge database can get good performance.

Keywords Chinese information processing, Question answering system, Concept space of language, Word semantic relativity, Hierarchical network concept theory

1 引言

问答系统是指系统接受用户以自然语言形式描述的提问,并从大量的数据中找出能回答用户提问的正确答案的信息检索系统。经过近十年的发展,问答系统已经成为自然语言处理领域和信息检索领域的一个重要分支和新兴的研究热点。

目前,自然语言问答系统从知识库的构建来看大致可以分为三大类^[1]:一类为限定域问答系统,第二类为开放域问答系统,第三类为基于常用提问集的问答系统。对于这三类问答系统的处理方法多集中在基于语言符号的概率统计模型上。

HNC 问答处理系统知识库构建与传统问答系统相同,但是处理方法发生了较大改变,它将处理深入到语言概念空间来进行,而不是停留在语言空间的语言符号上,这样做的最大好处在于可以减少语言符号带来的固有的语义模糊。

HNC 问答处理系统基于 HNC(概念层次网络)自然语言理解框架(由中科院声学所黄曾阳先生提出),是一种崭新的问答处理模式。本文将 HNC 问答处理系统中的一些关键性问题做一个讨论。本文第 2 节从整体上介绍 HNC 问答系统。第 3 节介绍所涉及的关键性算法。第 4 节给出本算法下的 HNC 问答系统性能测试。最后对文章进行总结并且提出

了今后工作的展望。

2 HNC 问答系统

2.1 问答系统基本原理

问答系统的处理是在语言概念空间完成的,所谓语言概念空间就是语言向抽象语义平面上的一个投影,可以看成是对事物的一种认识。比如苹果这个事物,在语言概念空间中会存在一种认识与其对应,而 HNC 问答系统利用的就是这种认识。为了计算机处理的方便,我们将这些认识用一个概念符号体系来表示。

下面从语言空间到语言概念空间的映射关系来说明问答系统原理。

对于任意一个语言概念空间中概念类 C ,它可以看成唯一对应概念空间中的某一种或者某一类概念节点,而对于语言空间中的词语集合 W ,我们将 W 和 C 之间建立对应关系,那么有,对于每一个属于词语集合 W 的 w ,存在一个带权值的词语到概念类的映射关系,假设词语 w_n 到概念类 C 的权值用 V_n 来表示,所有词语映射的集合就形成了一个映射向量空间 $V(V_1, V_2, V_3, \dots, V_n)$,如果向量空间中的特征项代表我们所要回答用户问题的某一个语义分量的可能性,特征项的值 $V_i (1 \leq i \leq n)$ 就表示形成这个语义分量的可能性值,那么,寻找用户提问中所涉及的特征语义分量就转化为对向量

^{*} 本文承国家 973 项目“自然语言理解的交互引擎研究”(2004CB318104)、中科院声学所知识创新工程“语言知识处理理论及技术”的资助。
吴晨 博士研究生,主要研究方向为自然语言理解处理技术、软件工程。张全 博士,研究员,博士生导师,主要研究方向为 HNC 自然语言理解处理技术,计算语言学。

空间中特征项的值(词语语义与概念类之间语义相关度)的计算和认定,这也是解决问题的关键所在。

在问答系统中,用户需要得到回复的是系统针对确定的某一个或者某几个语义分量上的回答,我们需要对各分量的确定设立筛选机制,这种机制通过语义相关度计算和阈值的设定来实现,语义分量权重由 HNC 概念基元符号体系语义相关度计算规则给出。系统根据词语与概念类之间的语义相关度为 V_i ,结合语义相关度阈值 α ,就可得出映射向量 V 向二值逻辑平面的投影后得出的筛选认定向量 $X(X_1, X_2, X_3, \dots, X_n)$, X 具体含义是认定问题求解是否收入该多维度语义表达分量。

为了问题求解的方便,我们将常识知识库中的知识,根据预先的设定划分成了与 C 等同的多个维度,每个维度对应一个语义分量,这些语义分量是我们通过对大量用户提问类型进行研究后得来的,具有一般意义,而信息的萃取则是在 HNC 交互平台下,通过手工干预校正完成的,准确率达到 99% 以上。

最后,应答文本生成主要是分情况考虑了几类文本的生成,文本生成充分利用了 HNC 句类以及语义块知识,实现了语义从概念层面向语言表达层面的反映射。

2.2 问答系统框图

HNC 问答系统框图如图 1 所示,系统输入是一个带有询问性质的句子,输出是针对该问句的应答文本。

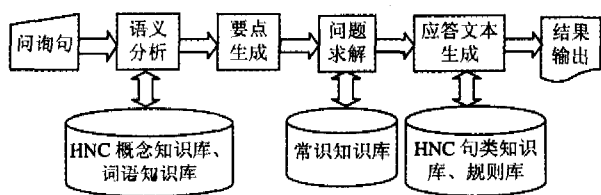


图 1 HNC 问答处理系统流程图

系统首先获取用户以自然语言形式提交的问句,然后对问句进行语义分析,获取概念空间中的句子语义结构、各组成成分之间的语义关系。语义分析的基础在于两类知识库:概念知识库和词语知识库。概念知识库又包含概念节点知识库、句类知识库和概念关联知识库。概念知识库与语言的表现形式无关,是在语言概念层面形成的一套知识表示模型。词语知识库是针对一种具体语言的、与语种相关的知识库,包括常用的词语的语义、概念类别、句类代码等语义知识,它描述的是语言空间中的具体词语与语言概念空间中的语义符号间的投影关系。它是语义分析的基础,相对稳定。

要点生成获取投影在概念空间语义平面上的句子要点,句子要点体现为问句所询问的语义分量,语义分量可以理解为传统问答系统的某个预期答案类型。在 HNC 问答系统中,允许一个问句提出对某个主体事物多个语义分量的询问,而不仅仅局限于事件的某一方面。

问题求解根据句子要点所对应的语义分量对提问进行求解操作,求解主要依靠经过预处理的结构化常识知识库中所包含的知识。

应答文本生成依据 HNC 概念知识库中的句类知识库和文本生成规则库对所获取的答案进行自然语言生成处理。

本文将对后三个阶段的处理算法做一个重点的研究讨论。

3 HNC 问答系统关键算法

本节将结合图 1 所示框图对系统所涉及的关键性算法进行研究。

3.1 语义分析

系统采用的语义分析技术是 HNC 句类分析技术^[2],句类分析可以给出巨量的语言概念层面上的语义结构以及各组成成分之间的语义关系。

例句:中国科学院声学研究所的地址 || 是 || 什么? + % 邮编 || 是 || 多少?

这是两个是否判断句的迭代,句类分析给出它的语义结构是:

第一句:jDJ=DB(判断对象)“中国科学院声学研究所(fpj2 * 101/fpea6)的地址(jz214)”+jD(判断词)“是(jlv111)”+DC(判断内容)“什么(f429)”。

jDJ 是句类,DB、jD、DC 为语义块名称。fpj2 * 101/fpea6 是中国科学院声学研究所的语义符号表示,jz214 表示“地址”是特定位置或特定物的效应,jlv111 表示“是”字的语义是一种基本逻辑概念的肯定,f429 表示“什么”是语习概念中的询问类概念。

第二句:jDJ=DB(判断对象)“中国科学院声学研究所(fpj2 * 101/fpea6)的邮编(ga25\5 + j3/j00)”+jD(判断词)“是(jlv111)”+DC(判断内容)“多少(f429)”。

ga25\5 + j3/j00 表示“邮编”是邮政意思下地区的数字。

3.2 要点生成

要点生成的主要任务是将句子的要点信息进行提取,形成一个以要点为维度的向量空间结构 $Y(Y_1, Y_2, \dots, Y_m)$,供问题求解模块调用。

算法 1 要点向量空间结构 Y 的生成算法:

(1)根据句子的语义结构,获取包含询问内容的整个语义块结构。

(2)从语义块结构中获取具有询问内容特征的关键要素。

(3)获取这些关键要素的语义符号表示。

(4)同一关键要素对应的不同语义符号为一组,假设共有 n 个关键要素,形成 n 组语义符号序列,并且这 n 组语义符号序列中分别有 $m_1, m_2, m_3, \dots, m_{n-1}, m_n$ 个语义符号。

(5)以组为基本单位,进行组与组之间元素的笛卡尔乘积运算,形成 n 列, $m_1 * m_2 * m_3 * m_{n-1} * m_n$ 行的矩阵。

(6)这个矩阵对应以行号为下标,列数为维度的要点向量数组 y_1, y_2, \dots, y_k ,其中 $k = m_1 * m_2 * m_3 * m_{n-1} * m_n$ 。这个向量数组的组合就是我们所求的以要点为维度的向量空间结构 $Y, Y = \{y_1, y_2, \dots, y_k\}^T$ 。

如例句中:根据句子的语义块可以提取出两项关键要素,依据是两个是否判断句的 DB 语义块的 C(内容)团块,第一个句子主语义块的 C 团块是“地址”,第二个是“邮编”,分别对应语义符号 jz214(特定位置或特定物的值表达)和 ga25\5 + j3/j00(邮政上地区的数字),根据这两个要素,我们就得到一个二维的要点向量 $y(jz214, ga25\5 + j3/j00)$ 。

说明:向量的维数随着要点的增加或者减少会改变。随着 C 团块语义的不同,有时得到的是一个向量数组,比如句子中主语义块的 C 团块内容如果是“地方”不是“地址”,那么向量的第一个特征值就会有三种情况,对应:(1)gc400-0; pj2-0(关系的基本构成方面,相对中央的地方)(2)wj2; j214(地域、位置)(3)j40-0(局部)。此时, Y 就是一个向量数组,包含

y1(gc400-0; pj2-0, ga25\5 + j3/j00); y2(wj2; j214, ga25\5 + j3/j00); y3(j40-0, ga25\5 + j3/j00)。

3.3 问题求解

问题求解就是把句子要点分析获取的结果与预先设定好的多维度语义表示框架下的语义分量进行比对, 确定问句中包含的关键信息, 并以这些信息为依据, 根据常识知识库中的知识形成一个包含答案信息的数据结构, 提供应答文本生成时用。

算法中按概念类别将语义分量划分成了 15 类, 包括日期、时间、地点、人物、事件、组织、事物数量、比重以及与应用密切结合的电话、邮编、E-mail、金额等等。这 15 类概念形成一个维数为 15 的特征向量 $Z(Z_1, Z_2, Z_3, \dots, Z_{15})$, 规定 Z_i 和 $Z_j (j < i)$ 的位置不可以随意调换, 特征向量对应的特征值 Z_i 是该语义分量的一个 HNC 符号。比如, 地点类概念对应 HNC 符号为 j214(意为第一参照系: 特定位置或特定物); 时间类概念对应 j1(意为时间的概念整体); 事物数量类概念对应 j308(意为基本数); 邮编概念比较狭义, 表示符号为 ga25\5 + j3/j00, 将这些 HNC 符号依次填入特征向量 Z 中就形成了一个具体的向量 z , 接下来的操作就是把要点向量(组) Y 中的所有特征项的值与语义分量概念特征向量 z 里的每一个特征值依据 HNC 语义相关度进行语义距离计算, 获得一个 0 到 1 空间内的相关度值, 构成语义分量权重向量 $V(V_1, V_2, V_3, \dots, V_n)$ 。对于 $Y(Y_1, Y_2, \dots, Y_m)$ 和 $Z(Z_1, Z_2, Z_3, \dots, Z_n)$ 形成的 V 有: 对于任意 $V_i, V_i = \text{MAX}(R(Y_j, Z_i))$, 其中 $(0 < j \leq m)$, R 表示 Y_j 和 Z_i 之间的语义相关度值。如果 Y 是一个向量组, 我们取向量组中每一向量 y_1, y_2, \dots 依次与 Z 计算后所得到的最大的那个 V_i 作为最后的结果。

V_i 产生后, 我们根据预先设定的权值 a , 将 V_i 投影到筛选认定向量 $X(X_1, X_2, X_3, \dots, X_n)$ 中。

假设我们设立的语义相关度阈值为 a , 通过语义相关度计算得到的词语语义与概念类之间的语义相关度为 V_i , 对于筛选认定向量 $X(X_1, X_2, X_3, \dots, X_n)$ 中 X_i 的值可以通过公式 $X_i = \begin{cases} 1 & V_i \geq a \\ 0 & V_i < a \end{cases} (1 \leq i \leq n)$ 得来。实质上, 筛选认定向量是映射向量 V 向二值逻辑平面的一个投影。它所表示的具体含义是是否收入该多维度语义表达分量。

假设要点向量空间结构为 Y_m , 其结构如算法 1 中所述。特征向量为 $Z(Z_1, Z_2, Z_3, \dots, Z_p)$ 。筛选认定向量为 $X(X_1, X_2, X_3, \dots, X_p)$ 。筛选认定向量的生成算法为:

算法 2 筛选认定向量的生成算法:

(1) 将 Y 中的每一个元素与 Z 中的每一个元素进行语义距离计算(参看算法 3), 形成一个行数为 $k * n$ 、列数为 p 的权重矩阵 $W_{(k * n) * p}$ 。

(2) 以权重矩阵的列为单位, 取出每列中权重值的最大值, 组成权重向量 V_p 。

(3) 将 V_p 中的每一个元素与预先设定的权值 a 进行比较。

(4) 计算出筛选认定向量为 $X(X_1, X_2, X_3, \dots, X_p)$ 。计算公式为 $X_i = \begin{cases} 1 & V_i \geq a \\ 0 & V_i < a \end{cases} (1 \leq i \leq p)$ 。

算法 3 语义距离计算算法:

在概念空间中的语义距离计算, 实质为字符串的比较运算以及规则修正。

语义距离计算公式为: $rsem(c_n, c_m) = \lambda nrel(c_n, c_m) + (1 -$

$\lambda) rel(c_n, c_m)$ 。

其中, $nrel(c_n, c_m)$ 为语义不相关计算函数:

$$nrel(c_n, c_m) = 1 - \frac{1}{\max(len(c_n), len(c_m)) \sum_{i=1}^{\max(len(c_n), len(c_m))} \frac{1}{2^{(|m[i] - m[lj])}}}$$

$rel(c_n, c_m)$ 为语义相关计算函数:

$$rel(c_n, c_m) = \frac{1}{\max(len(c_n), len(c_m)) \sum_{i=1}^{\max(len(c_n), len(c_m))} \frac{i / \max(len(c_n), len(c_m))}{2^{(|m[i] - m[lj])}}$$

c_n 和 c_m 为词语概念空间中的语义符号。

$len()$ 函数表示获取字符串的长度。

$Cr[i]$ 表示取字符串 Cr 的 i 个字符, 在设计字符时已将字符取值范围限制在 16 进制的 0 至 F 之间。字符不存在时取数值 0。

λ 为相关度平衡系数, 通常取经验值 0.6。

如例句, 通过算法 2 的计算我们可以得到这样的结果: $V = (0, 0, 0.91, 0.01, 0, 0, \dots, 0.99, 0.2, 0, 0, 0.1)$, 其中 V_1 表示日期类概念, V_2 表示地点类概念, V_3 表示人物类概念, V_4 表示事件, V_{10} 邮编类概念。HNC 语义相关度经验阈值 $a = 0.7$, 投影得到 $X = (0, 0, 1, 0, 0, 0, \dots, 1, 0, 0, 0, 0)$, 于是, 认定了 X_3 (地点) 和 X_{10} (邮编) 两个问题语义分量。

最后, 根据认定的问题语义分量从结构化的常识知识库中获取对象“中国科学院声学研究所”的相应语义分量知识: 地址信息为“北京市海淀区北四环西路 21 号”邮编信息为“100080”, 形成一个“对象-问题-解答”数据结构。

3.4 应答文本生成

应答文本生成通过分类解决三类文本的生成来完成, 这三类文本分别针对三种可能出现的处理结果, 第一种情况是系统无法理解用户提问或者用户提问模糊不清时做出的回答; 第二种情况是系统无法找到能够回答用户提问时做出的回答; 第三种情况是系统可以正确回答用户提问时做出的回答。前两种情况的应答文本生成较为固定, 前者提醒用户确认输入或者提供更为详细的询问信息, 后者告知用户所询问信息的解答没有被系统录入。而第三种情况文本生成相对复杂, 系统统一采用 HNC 句类知识中的是否判断句(jDJ)作为语句生成基本语义结构, 这一类句子常见的表现形式是传统语言学中所指的“是”字句。当然, 这里面存在许多相似和不同处, 不再过多讨论。

根据问题求解输出的“对象-问题-解答”数据结构, 生成最终应答文本。例句所生成的最终应答文本是: 中国科学院声学研究所的地址是北京市海淀区北四环西路 21 号, 邮编是 100080。

4 算法应用与测试结果

该算法已经应用于一个商用的问答系统, 并且与中国移动通信网相结合, 形成了一个基于短信的自然语言问答系统, 主要面向大型集团用户, 供集团内部信息查询使用。系统在原有处理能力的基础上增加了(1)误输入的同音字处理;(2)不同称谓同一对象识别;(3)模糊对象查询等功能。

我们对应用系统进行了系统的测试, 测试模仿 TREC 的 QA Track 建立, 比如:

英文提问: Who wrote “East is east, west is west and

(下转第 217 页)

步骤 3: TA 接到协商信息表明自己没有参加协商的意愿;协商结束。

步骤 4: TA 与有意协商的 CSA 进入协商状态,双方交换初始 offer;交换初始 offer 可以时 CSA 与 TA 同时交换,也可以是其中一方向另一方提供 offer(CSA 根据 TA 的排课期望提供接近的空教室,相邻时间段等资源信息)。

步骤 5: TA 收到 CSA 的 offer 后做出应答,如果同意修改排课偏好,以 Accept 作为应答,协商结束。

步骤 6: 如果参与协商的一方希望变更协商问题(增加或减少一个排课偏好),则可以 Change 作为应答;TA 可根据教师排课期望的偏好值依次增加或减少。

步骤 7: 如果参与协商的一方拒绝对方的 offer;则以 Reject 作为应答,或参与协商的一方退出,或到达协商限定时间,协商结束。

协商结束后没有达成排课要求的 TA,可提出与其他 TA 进行等长时段交换的要求来达成一致。这样确保每一个教师最后都能排课成功。

6 系统的试验

研究采用 Java 语言撰写,以 JADE(Java Agent Development Environment)的多 Agent 系统平台实作排课系统。本研究以某一大学一个学期的课程排列进行测试,结果如下:对 40 个班级进行了 120 个课程的排课试验,参加排课的 36 个教师第一次满足教师期望,成功安排教学资源的为 28,第一次排课消耗的时间为 7 分钟,在接下来的 5 次协商中,最长的消耗时间也仅为 25 分钟。与同类排课系统相比,大大提高了工作效率。在试验测试过程中,第一次排课教师满意度达到 86%,通过地点和时间的协商,90%以上的教师满意度都有所提高。这样的测试结果基本满足了学校排课工作的要求,同时在很大的程度上也满足了教师的排课期望,提高了排课质量。

结论 本研究是以高校排课为背景研究多 Agent 之间协

商问题。本文提出的系统构架在解决排课问题的基础上满足了教师个人排课的偏好,提高了教师对课表的满意度。本文利用多 Agent 之间的协商来代替教师之间的实际协调,降低了行政单位与教师之间以及教师与教师之间排课协调的时间。研究提出的以教师偏好来协商排课能在不违反公平正义的原则下满足多数教师的个人排课期望,避免教师因排课不公产生不满情绪而影响教学质量。本研究还有许多需要改进的地方。比如对于权值较低先进的 TA 与后进权值更高的 TA 之间怎样协调,使得更好地满足各个 TA 的排课期望;对于 TA 之间等长的时间段交换的协商都需要进一步的研究。

本文已经提出了一个以多 Agent 协商为设计理念的排课系统,也已经初步做出了一些试验结果,后续研究的重点将是实际学校课程作为原始资料来测试系统,验证系统的稳定性以及运作结果是否能达成相关的设计目标。

参考文献

- 1 Münch L, Stehli M, Schulz R. An Agent-Based Architecture for Solving Dynamic Resource Allocation Problems in Manufacturing. In: A. Verbraeck, W. Krug, eds. Proceeding 14th European Simulation, (c)SCS Europe BVBA, 2002
- 2 Aronsson M, Kreuger P, Sjöland T. Heterogeneous Scheduling and Rostering (Extended Abstract). <http://www.sics.se/%7Ealf/cotic/2000a.ps>, 2000
- 3 Fatima S S, Wooldridge M. Adaptive Task and Resource Allocation in Multi-Agent Systems. AGENTS'01, Montreal, Quebec, Canada, 2001
- 4 Wang LiChun, Chen ShiFu. A Multi-Agent Multi-Issue Negotiation Model [J]. Journal of Software, 2002, 13(8):1637~07
- 5 Kraus S, Wilkenfeld J, Zlotkin G. Multi-agent Negotiation under Time Constraints. Artificial Intelligence, 1995, 75:297~345
- 6 Kraus S, Schechter O. Strategic-negotiation for Sharing a Resource between Two Agents, Computational Intelligence, 2003, 19(1):9~41
- 7 唐勇,唐雪飞,王玲.基于遗传算法的排课系统[J].计算机应用,2002(10)

(上接第 202 页)

never the twain shall meet"? 参见 TREC 2002(http://trec.nist.gov/data/qa/2002_qadata/main_task_QAdata/t11_500-numbered.txt).

相应的中文提问改为:谁提出了《相对论》?

英文提问:What is the name of CEO of Apricot Computer?

相应的中文提问改为:中科院的现任院长叫什么名字?

在这基础上,我们并且增加了对随机限定域下的提问的测试。

目前系统常识库中存在有 27887 条知识,对于 35 次预先设立的主题询问以及 50 次随机产生的限定域下的针对地址、联系方式、人物、数量等方面信息的询问或者错误的、无结果的、过于模糊的询问输入,系统一次做出正确回答的次数达到了 32 和 46 次,CWS(Confidence Weighted Score)=91.76%,高于在 TREC 会议中,一般基于结构数据库问答系统 70%左右的准确率。考虑机器自动生成结构化数据库的正确率 42.32%,系统总体表现性能为 $91.76\% * 42.32\% = 38.83\%$,仍然高于 TREC 会议的平均水平 30%。同时,我们将所实现的系统与之前完成的基于关键词的结构数据库 QA 系统的性能相比,基于内容分析的中文问答系统 CWS 提高了 25.3 个百分点(基于关键词的问答系统 CWS 为 66.45%)。这主要是由于采用了语义分析,减少了大量的同义和多义现象的原因。

结论 本文设计了 HNC 问答处理系统中的一些关键性算法,并且进行了系统实现。这些算法服务于 HNC 问答处理系统问题理解及问题求解阶段的任务,重点考虑了语言概念空间中的语义相关度计算算法。该算法可以满足 90%的语义相关度计算,其他 10%的特例情况则通过特殊处理规则对其进行修正,达到了较好效果。在这些算法指导下的商用系统取得了令人满意的结果。今后的工作方向和重点将放在结构化常识知识库生成方面,目前的知识库内容萃取是依靠人工干预校正完成的,机器自动处理正确率较低,需要进一步提高。

参考文献

- 1 吴友政,等.问答式检索技术及评测研究综述[J].中文信息学报,2005,19(3):1~11
- 2 黄曾阳. HNC(概念层次网络)理论[M].北京:清华大学出版社,1998
- 3 黄曾阳.语言概念空间的基本定理和数学物理表示式[M].北京:海洋出版社,2004
- 4 Kim S-M, Baek D-H, Kim S-B, Rim H-C. Question Answering Considering Semantic Categories and Co-Occurrence Density [R]. In: Proc. of the night Text Retrieval Conf. (TREC-9)
- 5 Cooper R J, Ruger S M. A simple Question Answering System [R]. In: Proc. of the night Text Retrieval Conf. (TREC-9)
- 6 Hermjakob U. Parsing and Question Classification for Question Answering [R]. In: Proc. of the workshop on Open-Domain Question Answering at ACL-2001
- 7 郑实福,刘挺,秦兵,李生.自动问答综述[J].中文信息学报,2002,16(6):46~52