



# (12)发明专利申请

(10)申请公布号 CN 106055537 A

(43)申请公布日 2016. 10. 26

(21)申请号 201610349629.6

(22)申请日 2016.05.23

(71)申请人 王立山

地址 510060 广东省广州市东风东路729号  
广东工业大学东风路校区

(72)发明人 王立山

(74)专利代理机构 广州市越秀区海心联合专利  
代理事务所(普通合伙)  
44295

代理人 王洪娟

(51) Int. Cl.

G06F 17/27(2006.01)

G06F 17/30(2006.01)

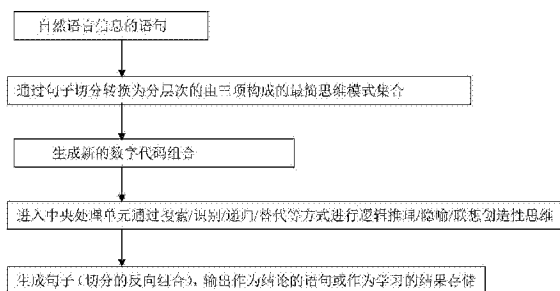
权利要求书2页 说明书35页 附图1页

## (54)发明名称

一种自然语言机器识别方法及系统

## (57)摘要

本发明公开了一种自然语言机器识别方法,包括下述步骤:(1)建立似谓词演算形式词义对应的数据库;(2)输入自然语言信息;(3)对所述的自然语言信息的语句进行逐句切分处理,并根据切分规则转换为一个或一个以上的似谓词演算形式句;(4)将所述的多个似谓词演算形式句转换为机器识别的电信号,然后输入至中央处理单元进行搜索或识别或递归或替代中的至少一种方式进行逻辑推理或隐喻或联想创造性思维的功能处理,并生成新的数字代码组合;(5)将所述的数字代码组合回溯转换为与输入的自然语言信息相应的新的自然语句作为输出,或者作为学习结果进行存储。



1. 一种自然语言机器识别方法,其特征在于包括下述步骤:(1)建立似谓词演算形式词义对应的数据库;(2)输入自然语言信息;(3)对所述的自然语言信息的语句进行逐句切分处理,并根据切分规则转换为一个或一个以上的似谓词演算形式句;(4)将所述的多个似谓词演算形式句转换为机器识别的电信号,然后输入至中央处理单元进行搜索或识别或递归或替代中的至少一种方式进行逻辑推理或隐喻或联想创造性思维的功能处理,并生成新的数字代码组合;(5)将所述的数字代码组合回溯转换为与输入的自然语言信息相应的新的自然语句作为输出,或者作为学习结果进行存储。

2. 根据权利要求1所述的一种自然语言机器识别方法,其特征在於所述的似谓词演算形式的定义为:自然语言句式均是由四种最简思维模式之一或组合构成的句式,每一种最简思维模式句均为包括谓词的最简句式,和现有的谓词演算的表达方式相似,将上述四种最简思维模式定义为似谓词演算形式。

3. 根据权利要求1所述的自然语言机器识别方法,其特征在於所述的数据库至少包括用自然数编码的新代码字词库,所述新代码编码数据库通过手工输入或利用现有的开放代码源的字词库输入字词的方式来建立。

4. 根据权利要求2所述的一种自然语言机器识别方法,其特征在於所述的切分规则为:将输入的自然语言信息的句子切分为一个或一个以上的似谓词演算形式句,把句子每一层的意义完整地保留下来,每一个所述的似谓词演算形式句是由最多三项一组的字词组成的最简句式;所述的自然语言信息对应段落的句子经切分后,转换为多个由三项一组构成的一组数列的集合。

5. 根据权利要求4所述的一种自然语言机器识别方法,其特征在於所述的切分规则采用以下的算法模型来实现:

(1)以句号为一个语句的句义中止标识,以段落为句义群中止标识,以整篇为段义群中止标识;子句以逗号为界,搜索整体句中的最简句式中间项的谓词,依次与字词库进行对比;

(1.1)以此谓词为界确定第一层面的前部项、后部项的句子成份,从而确定句式前部项为第一层面的最简句式第一项,确定句式后部项为第一层面的最简句式第三项;

(1.2)如果原句中间项的谓词有省略,则先补足谓词,然后重复步骤1.1的操作;

(2)按顺序分别对所述的句式前部项和句式后部项为进行第二层面切分,重复进行相同于上述步骤1的切分过程;

(3)限定词的受限主词作为下一层面的最简式的第一项,第二项添加谓词/谓语,第三项则为该限定词;

(4)再在下一层面进行上述切分过程,直至全句完成切分。

6. 根据权利要求5所述的一种自然语言机器识别方法,其特征在於所述的句子中谓词的搜索算法模型为:

(1)在一句子中逐次将字与字词库比对,对应字词库输出各词在该句子中的属性/词性,直到搜索出第一个谓词,继续之后的搜索;如果再未有谓词,则搜索完成;判断词或动词之前者为行为使动者,之后者为表述词或行为受动者,这即找出了最简句式。

(2)如果后边搜索到第二个谓词,继续之后的搜索,如果再未有谓词,则搜索完成;谓词之前者为行为使动者,之后者为表述词或行为受动者,即找出了最简句式的复合结构。

7. 根据权利要求2所述的自然语言机器识别方法,其特征在於所述的自然语言信息转换为似谓词演算形式句后的自动推理、联想过程为:切分后的语句变成了由三项一组的字词构成的最简思维模式的集合,运用搜索匹配识别,进行递归替代计算过程而实现自动推理、联想思维过程。

8. 根据权利要求2或3所述的自然语言机器识别方法,其特征在於所述的由自然语言信息转换的似谓词演算形式进行演算的推理算法模型:

(1)按时间优先原则,取首句第一组数列中的第三项减次句中第一组数列中的第一项,如结果为0,并且首句第一组数列中的第二项减次句第二组数列中的第二项的结果也为0,则将次句第二组的第一项替代首句第一组数列的第三项;新的首句第一组数列构造完成;

(1.1)如未有上述结果,则取首句第一组数列中的第三项减次2句第三组数列中的第一项,如结果为0,并且首句第一组数列中的第二项减次2句第三组数列中的第二项的结果为0,则将次2句第三组的第一项替代首句第一组数列的第三项;新的首句第一组数列构造完成;

(2)取第一组数列中的第三项减去已成功完成步骤1运算的次组数列中的第一项,如结果为0,并且第一组数列中的第二项减去已成功完成步骤1运算的次组数列中的第二项的结果为0,则将已成功完成步骤1运算的次组数列中的第一项替代第一组数列的第三项;新的第一组数列构造完成;

(3)继续上述过程直至无法完成上述过程,中止;输出新的第一组数列,这即是推理结果;

(4)如果所选的一组数列无法完成上述过程,则选择次一组数列进行上述过程;

(5)有推理结果或无推理结果,输出结论。

9. 一种自然语言机器识别系统,其特征在於包括人机接口模块、句子切分模块、中央处理单元、句式合成模块及数据库模块,所述的句子切分模块和句式合成模块分别与所述中央处理单元的输入端和输出端通过电信号相连,所述的数据库模块至少包括字词库管理模块。

10. 根据权利要求9所述的一种自然语言机器识别系统,其特征在於所述的数据库模块为多库协同模块,还包括知识库管理模块、情景库管理模块、多元语义网络库管理模块及隐喻网络库管理系统。

## 一种自然语言机器识别方法及系统

### 技术领域

[0001] 本发明涉及自然语言文字处理研究领域,特别涉及一种自然语言识别方法及系统。

### 背景技术

[0002] 随着人工智能的快速发展,各种人工智能产品在社会各方面获得了愈加广泛的应用。但是,人与机器直接使用人类自然语言交流,是实现人工智能的终极目标。目前,由于方法论和技术路线的欠缺,人工智能尚远远做不到这一点。

### 发明内容

[0003] 本发明要解决的技术问题是针对现有人工智能的上述不足,提出对于现有人工智能具有颠覆性的方法论,提供由人类思维通过自然语言文字完全转为机器思维载体的方法,从而实现人机思维的一致性,实现机器像人一样思维和与人互动。

[0004] 本发明的技术方案是这样的:一种自然语言机器识别方法,包括下述步骤:(1)建立似谓词演算形式词义对应的数据库;(2)输入自然语言信息;(3)对所述的自然语言信息的语句进行逐句切分处理,并根据切分规则转换为一个或一个以上的似谓词演算形式句;(4)将所述的多个似谓词演算形式句转换为机器识别的电信号,然后输入至中央处理单元进行搜索或识别或递归或替代中的至少一种方式进行逻辑推理或隐喻或联想创造性思维的功能处理,并生成新的数字代码组合;(5)将所述的数字代码组合回溯转换为与输入的自然语言信息相应的新的自然语句作为输出,或者作为学习结果进行存储。

[0005] 本发明所述的似谓词演算形式的定义为:自然语言句式均是由四种最简思维模式之一或组合构成的句式,每一种最简思维模式句均为包括谓词(谓语)的最简句式,和现有的谓词演算的表达方式相似,将上述四种最简思维模式定义为似谓词演算形式。

[0006] 本发明所述的数据库至少包括用自然数编码的新代码字词库,所述新代码编码数据库通过手工输入或利用现有的开放代码源的字词库输入字词的方式来建立。

[0007] 本发明所述的切分规则为:将输入的自然语言信息的句子切分为一个或一个以上的似谓词演算形式句,把句子每一层的意义完整地保留下来,每一个所述的似谓词演算形式句是由最多三项一组的字词组成的最简句式;所述的自然语言信息对应段落的句子经切分后,转换为多个由三项一组构成的一组数列的集合。

[0008] 本发明所述的切分规则采用以下的算法模型来实现:

[0009] (1)以句号为一个语句的句义中止标识,以段落为句义群中止标识,以整篇为段义群中止标识;子句以逗号为界,搜索整体句中的最简句式中间项的谓词,依次与字词库进行比时;

[0010] (1.1)以此谓词为界确定第一层面的前部项、后部项的句子成份,从而确定句式前部项为第一层面的最简句式第一项,确定句式后部项为第一层面的最简句式第三项;

[0011] (1.2)如果原句中间项的谓词有省略,则先补足谓词,然后重复步骤1.1的操作;

[0012] (2)按顺序分别对所述的句式前部项和句式后部项为进行第二层面切分,重复进行相同于上述步骤1的切分过程;

[0013] (3)限定词的受限主词作为下一层面的最简式的第一项,第二项添加谓词/谓语,第三项则为该限定词;

[0014] (4)再在下一层面进行上述切分过程,直至全句完成切分。

[0015] 本发明所述的句子中谓词的搜索算法模型为:

[0016] (1)在一句子中逐次将字与字词库比对,对应字词库输出各词在该句子中的属性/词性,直到搜索出第一个谓词,继续之后的搜索;如果再未有谓词,则搜索完成;判断词或动词之前者为行为使动者,之后者为表述词或行为受动者,这即找出了最简句式。

[0017] (2)如果后边搜索到第二个谓词,继续之后的搜索,如果再未有谓词,则搜索完成;谓词之前者为行为使动者,之后者为表述词或行为受动者,即找出了最简句式的复合结构。

[0018] 本发明所述的自然语言信息转换为似谓词演算形式句后的自动推理、联想过程为:切分后的语句变成了由三项一组的字词构成的最简思维模式的集合,运用搜索匹配识别,进行递归替代计算过程而实现自动推理、联想思维过程。

[0019] 本发明所述的由自然语言信息转换的似谓词演算形式进行演算的推理算法模型:

[0020] (1)按时间优先原则,取首句第一组数列中的第三项减次句中第一组数列中的第一项,如结果为0,并且首句第一组数列中的第二项减次句第二组数列中的第二项的结果也为0,则将次句第二组的第一项替代首句第一组数列的第三项;新的首句第一组数列构造完成;

[0021] (1.1)如未有上述结果,则取首句第一组数列中的第三项减次2句第三组数列中的第一项,如结果为0,并且首句第一组数列中的第二项减次2句第三组数列中的第二项的结果为0,则将次2句第三组的第一项替代首句第一组数列的第三项;新的首句第一组数列构造完成;

[0022] (2)取第一组数列中的第三项减去已成功完成步骤1运算的次组数列中的第一项,如结果为0,并且第一组数列中的第二项减去已成功完成步骤1运算的次组数列中的第二项的结果为0,则将已成功完成步骤1运算的次组数列中的第一项替代第一组数列的第三项;新的第一组数列构造完成;

[0023] (3)继续上述过程直至无法完成上述过程,中止;输出新的第一组数列,这即是推理结果;

[0024] (4)如果所选的一组数列无法完成上述过程,则选择次一组数列进行上述过程;

[0025] (5)有推理结果或无推理结果,输出结论。

[0026] 本发明还提供一种自然语言机器识别系统,包括人机接口模块、句子切分模块、中央处理单元、句式合成模块及数据库模块,所述的句子切分模块和句式合成模块分别与所述中央处理单元的输入端和输出端通过电信号相连,所述的数据库模块至少包括字词库管理模块。

[0027] 本发明所述的数据库模块为多库协同模块,还包括知识库管理模块、情景库管理模块、多元语义网络库管理模块及隐喻网络库管理系统。

[0028] 本发明与现有技术相比,具有下述优点:

[0029] 自然语言语句的基础是基本思维表达模式,与谓词演算形式相似。本发明可以实

现自然语言与谓词演算形式的完全及自动转换,从而为实现机器直接使用人类自然语言的人工智能奠定基础,实现与人的思维使用完全相同的载体——语言文字系统,智能工作的基础机制是相同的。由于转换保留着全部完整句义信息,因而完全实现了机器对人类自然语言直接使用的技术路径,这体现在本发明所提出并论述的自然语言似谓词形式转换后的基本思维表达方式及推理、隐喻、联想等表述之中。

[0030] 按照本发明的方法实现的人工智能,通过使用和人一样的思维载体——自然语言信息(如文字),使机器能够和人一样对话、推理思考、解决问题,所以,最接近“人脑”。自然语言符号系统直接作为人机交流的工具,其效率是最高的。

[0031] 本发明的自然语言机器识别方法及系统,应用前景广泛:例如可以应用本发明提炼文章的摘要或主题,可以应用到互联网智能搜领域索,可以实现自动阅读和理解文章及书籍,分类审查以及翻译等工作。本发明可以让机器通过自动学习并自动扩充知识库;还可以通过阅读自然语言文字来使用其它的人工智能专业软件,以及使机器人使用自然语言进行人机交流。

## 附图说明

[0032] 图1是本发明工作流程示意图;

[0033] 图2是本发明的工作原理方框图。

## 具体实施方式

[0034] 下面接合具体实施方式对本发明的技术方案做进一步的详细说明,但不构成对本发明的任何限制。

[0035] 本发明的具体实施方式如图1、图2所示,一种自然语言机器识别方法,包括下述步骤:(1)建立似谓词演算形式词义对应的数据库;(2)输入自然语言信息;(3)对所述的自然语言信息的语句进行逐句切分处理,并根据切分规则转换为一个或一个以上的似谓词演算形式句;(4)将所述的多个似谓词演算形式句转换为机器识别的电信号,然后输入至中央处理单元进行搜索或识别或递归或替代中的至少一种方式进行逻辑推理或隐喻或联想创造性思维的功能处理,并生成新的数字代码组合;(5)将所述的数字代码组合回溯转换为与输入的自然语言信息相应的新的自然语句作为输出,或者作为学习结果进行存储。

[0036] 下面将本发明的方法详细叙述如下:

[0037] (一)首先,将人类思维归纳为四种简单模式及其少量模式的组合。

[0038] 人类思维对各种事物的描述可以归纳为四种情形,这就是“什么是什么”、“什么做(引起)什么”、“什么在哪里”、“什么在何状态”四种模式以及它们复合之表达。上述这种四种情形被本发明称为“思维的最基本模式”。

[0039] 与之相反的是被动性方面,这就是“什么(被认为)是什么”、“什么(被认为)在何状态”、“什么(被认为)在哪里”、和“什么被做(引起)什么,这是由作用是相互的自然原理所决定的,即发生作用的必是作用者与被作用者参与的过程,按照这个作用与反作用同时发生的原理,参与作用过程的二者同时又分别是作用者与被作用者,只是表述的对象因空间与时间特质(由时间与空间限定唯一性原理决定)不同而已。

[0040] 本发明所述的似谓词演算形式的定义为:自然语言句式均是由四种最简思维模式

之一或组合构成的句式,每一种最简思维模式句均为包括谓词(谓语)的最简句式,和现有的谓词演算的表达方式相似,将上述四种最简思维模式定义为似谓词演算形式。

[0041] 似谓词演算形式句是由最多三项一组的字词构成的最简句式,对应于四种思维的最基本模式,包括以下四种类型:

[0042] (1)“WHAT BE WHAT”模式——“什么是什么”模式;

[0043] (2)“WHAT DO WHAT”模式——“什么做(引起)什么”模式;

[0044] (3)“WHAT BE WHERE”模式——“什么在哪里”模式;

[0045] (4)“WHAT BE STATE”模式——“什么在何状态”模式。

[0046] (1)WBW模式(“WHAT BE WHAT”模式);

[0047]  $=(\text{determiner}[x], \text{noun}(\text{WHAT1}), \text{IS}, \text{determiner}[z] \text{noun}(\text{WHAT2}))$ 。

[0048] (2)WDW模式(“WHAT DO WHAT”模式);

[0049]  $=(\text{determiner}[x], \text{noun}(\text{WHAT}), \text{determiner}[y] \text{verb}(\text{DO}), \text{determiner}[z] \text{noun})$ 。

[0050] (3)WBWH模式(“WHAT BE WHERE”模式);

[0051]  $=(\text{determiner}[x], \text{noun}(\text{WHAT}), \text{IS}, \text{determiner}[z] \text{noun}(\text{Where}))$ 。

[0052] (4)WBST模式(“WHAT BE STATE”模式);

[0053]  $=(\text{determiner}[x], \text{noun}(\text{WHAT}), \text{IS}, \text{determiner}[z] \text{adjective}(\text{State}))$ 。

[0054] 当determiner均为空集时,句子为最简式。

[0055] 下面对WDW模式加以举例说明。

[0056] 假设我们要描述一幕人打狗的情景,造句“人打狗”作为核心句义,这是最简式。句子中三个词都是抽象概念,一般需要分别加以限定,才更具有描述性,从而成为日常的描述。即“人”是“怎样的人?”、“如何打?”、“狗”是“怎样的狗?”、“事件发生在哪儿?”等等。下面例句即是:

[0057] 早晨,一个污头垢面的人在大街上狠狠地打一条黄皮狗。

[0058] (二)其次,人类自然语言的句式归结为上述四种简单模式及其少量模式的组合。

[0059] 自然语言是思维的载体,句式是人类思维的组织结构形式。句式由不同层次的一个或几个相同思维模式(即思维的最基本模式)构成。

[0060] (三)句式切分

[0061] 建立以概念生成系统为基础的字词数据库,该数据库至少包括用自然数编码的新代码字词库,所述新代码编码数据库通过手工输入或利用现有的开放代码源的字词库输入字词的方式来建立。句式切分时具有自动搜索功能并依据思维的最基本模式将句子分层次逐项分割,自然语言文字的句子经分层次切分后,句子变为三项一组的字词构成的最简思维模式的集合。

[0062] 一段相对完全的语句以标点为界,逗号之间的语句可以认为是合析(与 $\wedge$ )或者(或 $\vee$ )关系,也许是因果关系。中文通常将限定成分放在主题词——名词或代词主语及动词谓语,它们是WHAT和DO的基本核心。在句子后部动词之前的名词,特别是前面有“的”的名词及其前面的句子成分通常是WHAT,其后的动词为DO,再后边的成分为另一个WHAT。另一种情形恰好相反,当句首为名词或代词,紧接着动词BE,则可直接套用基本思维模式对句子切分。

[0063] 具体切分规则为:将输入的自然语言信息的句子切分为一个或一个以上的似谓词

演算形式句,把句子每一层的意义完整地保留下来,每一个所述的似谓词演算形式句是由最多三项一组的字词组成的最简句式;所述的自然语言信息对应段落的句子经切分后,转换为多个由三项一组构成的一组数列的集合。

[0064] 所述的切分规则采用以下的算法模型来实现,此算法模型适用于大部分的规整句式:

[0065] 1.以句号为一个语句的句义中止标识。以段落为句义群中止标识,以整篇为段义群中止标识。子句以逗号为界,搜索整体句中的最简式中间项“BE,DO”,依次与字词库进行比时;

[0066] 1.1 以此“BE”或“DO”为界确实第一层面的前后两个“WHAT”句子成份,从而确定前WHAT1为第一层面的最简式第一项,确定后WHAT2为第一层面的最简式第三项;

[0067] 1.2 如果原句中间项“BE,DO”有省略,则先补足谓词,然后重复步骤1.1的操作。

[0068] 2.按序分别对WHAT1和WHAT2进行第二层面切分,重复进行相同于上述步骤1的切分过程;

[0069] 3.限定词的受限主词作为下一层面的最简式的第一项,第二项添加“BE”,第三项则为该限定词。限定词指“……的,……地,摹状词,数量词等起限定作用的词,包括主名词之前的起限定作用的名词。

[0070] 4.再在下一层面进行上述切分过程,直至全句完成切分。

[0071] 自然语言以逗号作为子句的分界标识,各子句间有某种逻辑关系或为并列表述。以条件词引出的子句,逗号位置替换为结果子句,下文例句中用标识“→”来表示两子句的因果关系。

[0072] 句子中BE或DO的谓词搜索算法模型:

[0073] 1.在一句子中逐次将字与字词库比对,对应字词库输出各词在该句子中的属性/词性,直到搜索出第一个谓词(判断词或动词),继续之后的搜索,如果再未有谓词(判断词或动词),则搜索完成。判断词或动词之前者为行为使动者(主语),之后者为表述词或行为受动者。这即找出了最简句式。

[0074] 如果后边搜索到第二个谓词(判断词或动词),继续之后的搜索,如果再未有谓词(判断词或动词),则搜索完成。谓词(判断词或动词)之前者为行为使动者(主语),之后者为表述词或行为受动者。句子一般最长是这种结构,即找出了最简句式的复合结构。

[0075] 1.1.如果第一个基本句子模式中主语数项少(即数值个数少),则基本可判定此基本句子模式为整句的主结构,第二个基本句子模式是次结构,第一个基本句子模式中的主语即是整句主语。

[0076] 1.1.1.计算第一个基本句子模式中主语数项数量与谓语后的数项数量并比较,如果主语数项数量小,则输出此基本句子模式为整句的主结构的判定。如非,则进入下一步。

[0077] 1.2 如果第二个基本句子模式中的表述词或行为受者数项比第一个基本句子模式中主语数项数量少(即数值个数少),则基本可判定此基本句子模式为整句的主结构,第一个基本句子模式是整句主语。

[0078] 1.2.1.计算第二个基本句子模式中的表述词或行为受者数项数量并与第一个基本句子模式中主语数项数量比较。该数项数量小则输出此基本句子模式为整句的主结构的判定。如非,则输出第一个基本句子模式中的主语即是整句主语的判定。



[0079] 下面是一种较简单的句型,但不同句型的算法类似。

[0080] 2. 找出整句的主结构后,将作为数列形式的句子主结构中的谓词(代码)前的数列各项作为W1,移出作为另一数列,将此谓词(代码)后的数列各项作为W2,移出作为另一数列。

[0081] 2.1. 对W1进行步骤1的搜索判定,找出名词中心词,将名词中心词的代码数与其他词性的词的代码数重新排序,依次为:名词中心词,判定词“是”、“在”或动词,表词或受词;如果W1仅是三个数项,则是最简句式(最简基本句型),切分中止。

[0082] 2.1.1. 复制W1中心词。

[0083] 2.1.2. 如果W1无判定词“是”或“在”,仅是二个数字项,则补缺。将“是”或“在”的代码数引入,并将三代码数重新排序为:名词中心词,判定词“是”、“在”,表词,重新排序后的W1设名为RW1。

[0084] 2.2. 对W2进行2.1. 搜索判定

[0085] 如果句子主结构中谓词(代码)是判定词“是”或“在”,则W2的中心词是表词;如果是动词,则为名词。复制W2中心词。重新排序后的W2设名为RW2。

[0086] 2.3. 将W1中心词、W2中心词,句子主结构中谓词排序成W1中心词、句子主结构的谓词,W2中心词。

[0087] 2.4. 将原句数列重新排序为三项数列的集合:

[0088] (W1中心词,句子主结构的谓词,W2中心词) $\wedge$ (RW1) $\wedge$ (RW2)

[0089] 句子中判断词组的搜索算法模型:

[0090] 1. 第一个数加第二个数之和(非算术加法之和)与词库比对,识别成功,则由这两数所代码(定义)的两个字词是一个组合词组。则转入步骤2;

[0091] 反之为单词,判别中止。由此字词(第二个数)重新开始步骤1的过程。

[0092] 2. 加上第三个数再与词库比对,如果依然被识别,则此三个数之和对应的三个字词是一个三字组合词,则转入步骤3;

[0093] 反之判别中止。由此字词(第三个数)重新开始步骤1的过程;

[0094] 3. 再与第四个数相加,如果依然被识别,则此四个数之和对应的四个字词是一个四字组合词,则转入步骤4;

[0095] 反之判别中止,由此字词(第四个数)重新开始步骤1的过程;

[0096] 4. N数再与次邻数n+1相加,如果依然被识别,则此n+1数之和对应的n+1个字词是一个n+1字组合词。

[0097] 反之判别中止,由此字词(第n+1个数)重新开始步骤1的过程。)

[0098] 5. 上述过程一直进行到识别失败为止。后面的数再进行上述相同的过程,直到遇到判定词“是”、“在”及唯一词性的动词为止。

[0099] 6. 由上述方式找到主语词组及谓语。输出结果。

[0100] 重要注释:上述算法是基于最大多数的语句情形,对于语句中省略谓词的和谓词属于较少数的复合词组的,则在字词库特别设定的基础上,对上述语句切分算法的相关部分加以微调即可。

[0101] 我们以下面具有代表性的例句切分,来说明将自然语句直接转换为似谓词演算的形式的基本方法:

- [0102] 玛丽在过生日时收到两只风筝,所以她把一个退还给商店。
- [0103] = (玛丽在过生日时,收到,两只风筝) → (所以)(她把一个退还,给,商店)
- [0104] = ((玛丽,在,过生日时),收到,(风筝,是,两只)) → (所以)(她,退还,一个(风筝)),把…给,商店)
- [0105] = ((玛丽,在,(时,是,(玛丽,过,生日)),收到,(风筝,是,两只)) → (所以)(她,退还,(风筝,是,一只)),把…给,商店)
- [0106] (她,退还,一个(风筝)),把…给,商店)
- [0107] = (玛丽,退还,(风筝,是,一只)) ∧ (玛丽,把…给,商店)
- [0108] 即
- [0109] ((玛丽,在,(时,是,(玛丽,过,生日)),收到,(风筝,是,两只))
- [0110] → (玛丽,退还,(风筝,是,一只)) ∧ (玛丽,把风筝给,商店)
- [0111] 可以逆方向从底部的最简式抽取始,将不同层面的结构简化为最简式抽取出来。具体操作如下:
- [0112] ((玛丽,在,(时,是,(玛丽,过,生日)),收到,(风筝,是,两只))
- [0113] → (玛丽,退还,(风筝,是,一只)) ∧ (玛丽,把风筝给,商店)
- [0114] (1)抽取最底部的最简式:(玛丽,过,生日) ∧ (风筝,是,两只)
- [0115] → (风筝,是,一只)
- [0116] (2)将上一层面的结构简化为最简式抽取出来
- [0117] (时,是,生日)
- [0118] (3)将再上一层面的结构简化为最简式抽取出来
- [0119] (玛丽,在,时)
- [0120] (4)将最上一层面的结构简化为最简式抽取出来
- [0121] (玛丽,收到,风筝)
- [0122] → (玛丽,退还,风筝) ∧ (玛丽,把风筝给,商店)
- [0123] 将以上四个层面合并,有
- [0124] (玛丽,在,时) ∧ (时,是,生日) ∧ (玛丽,过,生日) ∧ (玛丽,收到,风筝)
- [0125] ∧ (风筝,是,两只)
- [0126] → (风筝,是,一只) ∧ (玛丽,退还,风筝) ∧ (玛丽,把风筝给,商店)
- [0127] 注:
- [0128] ((玛丽,在,(时,是,(玛丽,过,生日)),收到,(风筝,是,两只))中,“玛丽、收到、风筝”是整句的核心词,其他可取消,则有
- [0129] ((玛丽,收到,风筝)
- [0130] 另外,(时,是,(玛丽,过,生日)中核心句义是(时,是,生日)
- [0131] (她,退还,(风筝,是,一只)) ∧ (她,把…给,商店)中核心句义是
- [0132] (她,退还,风筝) ∧ (她,把风筝给,商店)
- [0133] “把…退还给…”是个复合行为表述,即“退还”和“把…给…”的两种连续行为的表述。
- [0134] 上段中没有与下段中完全相同的或可递归的最简式,故不能合一。由之判断二者不直接相干,即不能构成二者的直接因果关系。换言之,此上下文因果关系为假。由上面两

项相同的最简式

[0135] (风筝,是,两只)→(风筝,是,一只)

[0136] 可以判断这正是原因所在,进入知识库查询。

[0137] 可连接指令,由(玛丽过生日→收到两只风筝)发出疑问句。

[0138] 玛丽过生日→玛丽应该收到一只风筝?

[0139] 注:由陈述句改疑问句有简单的固定模式,如

[0140] “时间到了,我去上班”可改为疑问句“时间到了,我应该去上班?”

[0141] 如果在切分语句时注意到动词隐含的主动性还是被动性,从而利用它在句中因果关系的位置,该语句也可以这样切分:

[0142] 玛丽在过生日时收到两只风筝,所以她把一个退还给商店。

[0143] =(玛丽在过生日时→(…,收到,两只风筝))→(她把一个退还,给,商店)

[0144] =(玛丽,在(时,是,过生日)→(玛丽,收到,两只风筝))→((玛丽,把…退还,一个(风筝)),给,商店)

[0145] =(玛丽,在(时,是,(玛丽,过,生日))→(玛丽,收到,(风筝,是,两只)))→((玛丽,把…退还,(风筝,是,一只)),给,商店)

[0146] =(玛丽,在(时,是,(玛丽,过,生日))→(玛丽,收到,(风筝,是,两只)))

[0147] →(她,把…退还,(风筝,是,一只))^(她,把…给,商店)

[0148] 简化后得到

[0149] (玛丽,过,生日)→(玛丽,收到,风筝)^(风筝,是,两只)

[0150] →(她,退还,风筝)^(她,把风筝给,商店)^(风筝,是,一只)

[0151] 注:收到=→,

[0152] 因为……原因,才收到……。所以“收到……”是因果关系中的结果部分。

[0153] 自然语言句式的构成及其切分详述:

[0154] 基本思维模式为上述的四种最简句式,自然语言句式就由基本思维模式(最简句式)构成。

[0155] WBW模式(“WHAT BE WHAT”模式)

[0156] determiner[x]noun(WHAT1),IS,determiner[z]noun(WHAT2).

[0157] WDW模式(“WHAT DO WHAT”模式)

[0158] determiner[x]noun(WHAT1),determiner[y]verb(DO),determiner[z]noun(WHAT2).

[0159] WBWH模式(“WHAT BE WHERE”模式,WHERE指时空位置)

[0160] determiner[x]noun(WHAT),IS,determiner[z]noun(Where).

[0161] WBST模式(“WHAT BE STATE”模式)

[0162] determiner[x]noun(WHAT),IS,determiner[z]adjective(State).

[0163] 当determiner均为空集时,句子为最简式。

[0164] 一般的句子(这里指的句子是以标点符号为界的表述部分)不会太长,在整句层面以单一模式构成主体框架的句子为多,最多不会超过三个基本模式。下面以部分句式加以说明。

[0165] 1.WDW模式(“WHAT DO WHAT”模式)

[0166] determiner[x]noun(WHAT1),determiner[y]verb(DO),determiner[z]noun(WHAT2).

[0167] 当determiner均为空集时,句子为最简式。

[0168] 1.1 adj[x]noun(WHAT1),determiner[y]verb(DO),determiner[z]noun(WHAT2).

[0169] determiner[x]=adjective[x],quantifier[x]———[x]是与名词匹配的表  
达某种表现、性质、关系、数量状态的形容词类属,包括数量词等。

[0170] determiner[x]=adj[X1,X2,⋯]

[0171] =(WHAT1,IS,X1∧X2,⋯)

[0172] 如果x中有归属性的类形容词,如“名词+的”,则也可变换为

[0173] =(WHAT1,has,X1)∧(WHAT1,IS,X2,⋯)

[0174] 例句:他的漂亮的房子就位于湖岸。

[0175] 他的漂亮的房子就位于湖岸

[0176] =(他的漂亮的房子,就位于,湖岸)

[0177] =((他,有,房子)∧(房子,IS,漂亮的),位于,湖岸)

[0178] 得到

[0179] (他,有,房子)∧(房子,IS,漂亮的)∧(房子,位于,湖岸)

[0180] 1.2 determiner[x]noun(WHAT1),adverb[y]verb(DO),determiner[z]noun(WHAT2).

[0181] determiner[y]=adverb[y]———[y]是与动词匹配的表达行为某种程度、特性、状态的副词类属

[0182] determiner[y]=adverb[Y1,Y2,⋯]

[0183] =(DO,IS,Y1∧Y2,⋯)

[0184] 例句:他流利地回答了外宾三个问题。

[0185] 他流利地回答了外宾三个问题

[0186] =(他,流利地回答了,外宾三个问题)

[0187] =(他,(回答,IS,流利的),(外宾,has,三个问题))

[0188] =(他,(回答,IS,流利的),(外宾,has,(问题,IS,三个)))

[0189] 底层:(问题,IS,三个)

[0190] 第二层:(外宾,has,问题)∧(回答,IS,流利的)

[0191] 顶层:(他,回答,外宾)∧(他,回答,问题)

[0192] =(他,回答,外宾∧问题)

[0193] 合并:(外宾,has,问题)∧(问题,IS,三个)→(他,回答,外宾∧问题)

[0194] 注:“回答”决定了(他,回答,外宾∧问题)在全句因果关系中的“果”之特性。所以

[0195] “回答”=←

[0196] 利用Prolog语言表达,即

[0197] reply(he,guest,question):-has(guest,question),three(question).

[0198] 1.3 determiner[x]noun(WHAT1),determiner[y]verb(DO),adjective[z]noun(WHAT2).

[0199] determiner[z]=adjective[z],quantifier[z]———[z]是与名词匹配的表

达某种表现、性质、关系、数量状态的形容词类属,包括数量词等。。

[0200] 以上句式是完全表示,从中可以演变出许许多多句型来。例如,有下面句型

[0201] adj[x]noun(WHAT1),adverb[y]verb(DO),adj[z]noun(WHAT2).

[0202] 伟大的人民一定创造出伟大的成绩。

[0203] 伟大的人民一定创造出伟大的成绩

[0204] =(伟大的人民,一定创造出,伟大的成绩)

[0205] =((人民,IS,伟大的),(创造出,IS,一定的),(成绩,IS,伟大的))

[0206] 或=((人民,IS,伟大的)→(创造出,(成绩,IS,伟大的)))

[0207] 由上式,得出

[0208] 底层:(人民,IS,伟大的)→(创造出,IS,一定的)^(成绩,IS,伟大的)

[0209] 顶层:(人民,创造出,成绩)

[0210] 合并:(人民,IS,伟大的)→(人民,创造出,成绩)^(创造出,IS,一定的)^(成绩,IS,伟大的)

[0211] 2.(WDW)DW模式(“WHAT DO WHAT”模式中前部项(主语位)是另一个WDW模式)

[0212] (determiner[x1]noun(WHAT3),determiner[y1]verb(DO1),determiner[z1]noun(WHAT4)),determiner[y]verb(DO),determiner[z]noun(WHAT2).

[0213] 基本模式中WHAT1=determiner[x1]noun(WHAT3),determiner[y1]verb(DO1),determiner[z1]noun(WHAT4)

[0214] 当determiner[x]noun(WHAT1)=WHAT DO WHAT,如:

[0215] 他打开抽屉拿出词典来。

[0216] 他打开抽屉拿出词典来

[0217] =(他打开抽屉,拿出…来,词典)

[0218] =((他,打开,抽屉),拿出…来,词典)

[0219] 底层:(他,打开,抽屉)

[0220] 顶层:(他,拿出…来,词典)

[0221] 他打开抽屉拿出词典来

[0222] 或(他,打开,抽屉)^(他,拿出…来,词典),

[0223] opens(he,drawer)^(takes out(he,dictionary))

[0224] 这是两个WHAT DO WHAT构成的语句,(他,打开,抽屉)所表述的时间特性在先,从时序上可知

[0225] (他,打开,抽屉)→(他,拿出…来,词典)

[0226] opens(he,drawer)→takes out(he,dictionary)

[0227] 3.WD(WDW)模式(“WHAT DO WHAT”模式中后部项(宾语位)是另一个WDW模式)

[0228] determiner[x]noun(WHAT1),determiner[y]verb(DO),(determiner[x1]noun(WHAT3),determiner[y1]verb(DO1),determiner[z1]noun(WHAT4)).

[0229] 下面列举一个符合上式变形之一的复杂语句:

[0230] 我站在高耸得像一株亭亭玉立的玉兰花似的灯柱下。

[0231] 语句切分如下:

[0232] 我站在高耸得像一株亭亭玉立的玉兰花似的灯柱下

- [0233] =(我,站入在…下,(高耸得像一株亭亭玉立的玉兰花似的灯柱))
- [0234] =(我,站入在…下,(灯柱,IS,高耸得像一株亭亭玉立的玉兰花似的))
- [0235] =(我,站入在…下,(灯柱,IS,(灯柱,高耸得,像一株亭亭玉立的玉兰花似的)))
- [0236] =(我,站入在…下,(灯柱,IS,(… ,高耸得,(… ,像…似的,一株亭亭玉立的玉兰花))))
- [0237] =(我,站入在…下,(灯柱,IS,(… ,高耸得,(… ,像…似的,(玉兰花,IS,一株亭亭玉立的))))))
- [0238] 底层:(玉兰花,IS,一株亭亭玉立的)
- [0239] 第四层:(灯柱,像…似的,玉兰花)
- [0240] 第三层:(灯柱,IS,高耸的)
- [0241] 第二层:(灯柱,IS,高耸的)
- [0242] 顶层:(我,站入在…下,灯柱)
- [0243] 合并:(我,站入在…下,灯柱)^(灯柱,IS,高耸的)^(灯柱,像…似的,玉兰花)^(玉兰花,IS,一株亭亭玉立的)
- [0244] 利用Prolog语言表达,即
- [0245] stand\_beneath(i,lamp standard),tower(lamp standard),like(lamp standard,magnolia),
- [0246] is a(magnolia,slim,graceful).
- [0247] 4.(WBWHERE)DW模式(“WHAT DO WHAT”模式),前部项(主语位)是另一个WBW模式。
- [0248] (determiner[x1]noun(WHAT3),IS,adjective[z1]noun(WHAT4)),determiner[y]verb(DO),determiner[z]noun(WHAT2).
- [0249] 基本模式中WHAT1=determiner[x1]noun(WHAT3),IS,determiner[z1]noun(WHAT4)
- [0250] 例句:他在取得重大科研成果之后就走上了经商的道路。
- [0251] 他在取得重大科研成果之后就走上了经商的道路
- [0252] =((他在取得重大科研成果之后),就走上了,经商的道路)
- [0253] =((他,在…之后,取得重大科研成果)→走上了,(道路,IS,经商的))
- [0254] =((他,在…之后,(取得,重大科研成果))→走上了,(道路,IS,经商的))
- [0255] =((他,在…之后,(取得,(科研成果,IS,重大)))→走上了,(道路,IS,经商的))
- [0256] =((他,在…之后,(取得,((成果,IS,科研),IS,重大)))→走上了,(道路,IS,经商的))
- [0257] (走上了,(道路,IS,经商的))
- [0258] =(他,走上了,道路)^(道路,IS,经商的))
- [0259] 他(X)
- [0260] 道路(X,Y)
- [0261] 道路(X,Y,Z)
- [0262] is(Y,Z)
- [0263] 道路(Y)
- [0264] 经商的(Z)

[0265] 用Prolog表示即

[0266]  $he(X), go\_on(X, Y, Z), is(Y, Z), road(Y), mercantile(Z).$

[0267] 5.(WBWHERE)D((WBWHERE)DW)模式,“WHAT DO WHAT”模式中前部项(主语位)是WBWHERE模式,后部项(宾语位)则是自相似的(WBWHERE)DW模式。

[0268]  $(determiner[x1]noun(WHAT3), IS, determiner[z1]noun(WHAT4)), verb(DO), (determiner[x2]noun(WHAT5), IS, determiner[z2]noun(WHAT6)), determiner[y2]verb(DO2), determiner[z2]noun(WHAT7)).$

[0269] 基本模式中WHAT1=determiner[x]noun(WHAT3), IS, determiner[z1]noun(WHAT4)

[0270] WHAT2=((WISWHERE)DW)模式

[0271] 例句:他在大街上邂逅了那位日后成为他妻子的漂亮女孩。

[0272] 他在大街上邂逅了那位日后成为他妻子的漂亮女孩

[0273] =((他在大街上)邂逅了,那位日后成为他妻子的漂亮女孩)

[0274] =((他,在…上,大街)邂逅了,(女孩,IS,那位日后成为他妻子的^漂亮的))

[0275] =((他,在…上,大街)邂逅了,(女孩,IS,((那位,在,日后),成为,他的妻子)^漂亮的))

[0276] 得到

[0277] (他,在…上,大街),(他,邂逅了,女孩),(女孩,IS,漂亮的),(女孩,在,日后)→(女孩,成为,他的妻子)。

[0278] 注:在此句中,“那位日后成为他妻子的漂亮女孩”是另一个相同的(WISWHERE)DW模式,所以切分过程也相同。

[0279] (四)语句切分后的计算

[0280] 切分后的自然语言语句转换为由三项字词构成的最简思维模式的集合,运用搜索匹配识别,进行递归替代等计算过程而实现自动推理、联想等思维过程。

[0281] 将自然语言语句切分后的似谓词演算形式的完全自动转换为谓词演算形式,从而利用现有的人工智编程软件,以实现特定的功能。

[0282] 由句子进行似谓词演算形式切分后与谓词演算形式的完全对应,将二者通过机器自动转换,进而参照规划算法及描述,用Prolog编程语言对转换为谓词演算形式后的句式加以表达。

[0283] 【实例】将下面的句子转换为谓词演算形式:

[0284] ①任何通过历史考试和中奖的人是快乐的。

[0285]

$$\forall x (\text{pass}(x, \text{history}) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x))$$

S1=任何通过历史考试和中奖的人是快乐的。

= (任何通过历史考试和中奖的人, 是, 快乐的)

= (任何(通过历史考试和中奖的人), 是, 快乐的)

= ( $\forall$ 人((人, 通过, 历史考试)  $\wedge$  (人, 中, 奖)), 是, 快乐的)

= ( $\forall$ 人(通过(人, 历史考试)  $\wedge$  中(人, 奖)), 是, 快乐的(人))

= ( $\forall$ 人(通过(人, (考试, 是, 历史))  $\wedge$  中(人, 奖)), 是, 快乐的

(人))

= ( $\forall$ 人(通过(人, 历史)  $\wedge$  中(人, 奖)), 是, 快乐的(人))

= ( $\forall x$ (通过(x, 历史)  $\wedge$  中(x, 奖)), 是, 快乐的(x))

=  $\forall x$ (pass(x, history)  $\wedge$  win(x, lottery)  $\rightarrow$  happy(x))

[0286] 从此句可得到最基础的核心句义,即IF(人,通过,历史考试) $\wedge$ (人,中,奖)THEN(人,是,快乐的),由三项组成的表述式,与谓词演算直接转换。

[0287] sentence(人,是,快乐的):-WDW(人,通过,历史考试),WDW(人,中,奖)。

[0288] WDW(人,通过,历史考试):-

[0289] WDW(人,中,奖):-

[0290] 回应这句话,可以从上面三个基本句子引起。

[0291] 注:(考试,是,历史)表示“考试”与“历史”等价,为了简洁,可取消“考试”一词。并将“人”一词替换为变量x。

[0292] 由上面

[0293]

$$S1 = \forall x (\text{pass}(x, \text{history exam}) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x))$$

$$= \forall x (\text{pass}(x, \text{history\_exam}) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x))$$

[0294] 比较上式 $\forall x$ (pass(x, history)  $\wedge$  win(x, lottery)  $\rightarrow$  happy(x))

[0295] 利用规划算法及描述,用Prolog编程语言将句式转换为谓词形式:

[0296] 句式开始状态为

start=任何通过历史考试和中奖的人是快乐的

= ( $\forall$ 人(通过(人, (考试, 是, 历史))  $\wedge$  中(人, 奖)), 是, 快乐的)

[0297] = [ $\forall$  nounphrase(determiner [x], noun), BE, determiner(x)]

goal= [ $\forall x$ (pass(x, history)  $\wedge$  win(x, lottery)  $\rightarrow$  happy(x))]

[0298] 转换步骤如下:

[0299] 1、transform(add(whole\_curves),out\_of\_whole\_curves(universal quantification全称量化 $\forall x$ ),(add(curves1),put\_in\_curves1(noun),put\_in\_curves1(comma),add\_in\_curves1(BE),add\_in\_curves1(comma),put\_in\_curves1(determiner [x]),add(comma),change(BE,implication $\rightarrow$ ),add(comma),determiner(x))

[0300] 2、transform(add(whole\_curves),out\_of\_whole\_curves(universal



quantification全称量化 $\forall x$ ),(add(curves1)(add(curves2),put\_in\_curves2(noun),  
add\_in\_curves2(comma),put\_in\_curves2(D01),add\_in\_curves2(comma),(put\_in\_  
curves2(determiner X1[x])),add(conjunction $\wedge$ ),add(curves3),put\_in\_curves3  
(noun),add\_in\_curves3(comma),(put\_in\_curves3(D02),add\_in\_curves3(comma),(put\_  
in\_curves3(determiner X2[x])),add\_in\_whole\_curves(comma),(change(BE,  
implication $\rightarrow$ ),determiner X3(x))

[0301] 3、transform(add(whole curves),out\_of\_curves(universal quantification  
全称量化 $\forall x$ ),(add(curves1),(add(curves2),out\_of\_curves2(D01),put\_in\_curves2  
(noun),add\_in\_curves2(comma),(put\_in\_curves2(determiner(X1)[x])),add  
(conjunction $\wedge$ ),add(curves3),out\_of\_curves3(D02),put\_in\_curves3(noun),add\_in\_  
curves3(comma),(put\_in\_curves3(determiner(X2)[x])),add\_in\_whole\_curves  
(comma),(change(BE,implication $\rightarrow$ ),determiner(X3)(x))

[0302] 4、transform(add(whole curves),out\_of\_curves(universal quantification  
全称量化 $\forall x$ ),(curves(curves1),(curves(curves2),out\_of\_curves2(D01),change\_to  
(x),in\_curves2(comma),(determiner(X1),conjunction( $\wedge$ ),curves(curves3),out\_of\_  
curves3(D02),change\_to(x),in\_curves3(comma),(determiner(X2))),(implication  
( $\rightarrow$ ),determiner(X3)(x))

[0303] 自然语言信息转换为似谓词演算形式句后的自动推理、联想过程为:切分后的语  
句变成了由三项一组的字词构成的最简思维模式的集合,运用搜索匹配识别,进行递归替  
代计算过程而实现自动推理、联想思维过程。段落的句子经切分后,段落变为多个由三项构  
成的一组数列的集合。由最简式三项构成的一组数列分放在寄存器地址。

[0304] 自然语言似谓词演算形式的自动推理

[0305] 【实例】作为“事”的描述的主语包含两种情形,一种是行为事件,另一种是某种关  
系表述。关于作为“事”的描述的主语,在句中则以独立的子句(即另一个基本思维模式)存  
在为多见。例如

[0306] 短语:在年青时,我做了一个重大决定,这就是去林场工作,在那里我遇到了一位  
美丽的姑娘,她给我带来一生的幸福。如果一个决定给你带来幸福,这个决定就是明智的。

[0307] 得出结论:我去林场工作是明智的决定。

[0308] 在年青时,我做了一个重大决定,这就是去林场工作,在那里我遇到了一位美丽的  
姑娘,她给我带来一生的幸福。如果一个决定给你带来幸福,这个决定这个就是明智的。

[0309] = (我,在 $\dots$ 时,年青) $\wedge$ (我,做了,一个重大决定) $\wedge$ (决定,是,去林场工作) $\rightarrow$   
(我,在,那里(林场)) $\wedge$ (我,遇到了,一位美丽的姑娘) $\rightarrow$ (她,给我带来,一生的幸福)。if  
(一个决定,给你带来,幸福), $\rightarrow$ then(这个决定,是,明智的)。

[0310] 一系列的行为动作构成因果关系链条,这隐含在上面语句分解后的行为动作次序  
中。

[0311] 注:去 $\dots$ (那里),“去”后边的词即是“那里”。

[0312] 上式=(我,在 $\dots$ 时,年青) $\wedge$ (我,做了,(决定,IS,一个 $\wedge$ 重大) $\wedge$ (决定,是,(我,  
去 $\dots$ 工作,林场)) $\rightarrow$ (我,在,林场) $\wedge$ (我,遇到了,(姑娘,IS,一位 $\wedge$ 美丽的) $\rightarrow$ (她,给 $\wedge$ 带  
来,(幸福,IS,我的 $\wedge$ 一生的)。if((决定,IS,一个),给 $\wedge$ 带来,(幸福,IS,你的)), $\rightarrow$ (then)

((决定,IS,这个),是,明智的)。

[0313] 将各层面核心句义提取出来,得到

[0314] (我,做了,决定)(决定,是,(我,去∧工作,林场))

[0315] →(我,在,林场)(我,遇到了,姑娘)

[0316] →(姑娘,给∧带来,幸福)

[0317] if((决定,给∧带来,幸福)then((决定,是,明智的)

[0318] 将(姑娘,给∧带来,幸福)代入(我,遇到了,姑娘),得到

[0319] (我,在,林场)(我,遇到了,(姑娘,给∧带来,幸福)),再代入(决定,是,(我,去∧工作,林场)),得到

[0320] ((决定,是,(我,去∧工作,林场),(我,遇到了,(姑娘,给∧带来,幸福))

[0321] 取消因果关系链条中“结果”的行为主体,只保留原因的行为主体“决定”,简化得到

[0322] (决定,给∧带来,幸福),

[0323] 由if((决定,给∧带来,幸福)then((决定,是,明智的),将(决定,是,(我,去∧工作,林场))代入,得到

[0324] ((决定,给∧带来,幸福)→((我,去∧工作,林场),是,明智的)

[0325] 注:在因果关系链条中,作为“结果”的行为主体词(即主语),可由作为原因的行为主体词(即主语)替代,因为因果关系传递的原因,它的因果关系的本质意义不变。例如,“他的话把她说哭了,她生气地吃不下饭,于是几天后她瘦了三公斤”,这段话中的因果关系可以追溯为“他的话使她瘦了三公斤”。

[0326] 输出短语的中心意义:

[0327] 我去林场工作是明智的决定

[0328] =(我去林场工作,是,(决定,是,明智的)

[0329] 将都含有“决定”的语句再抽取出来,得到

[0330] 核心句义:

[0331] (我,去…工作,林场),((我,去…工作,林场),是,决定),(决定,是,明智的)

[0332] 设w1=(我,去…工作,林场)

[0333] w1(X)

[0334] 决定(X,Y,Z)

[0335] is(Y,Z)

[0336] 决定(X,Y)

[0337] 决定(Y)

[0338] 明智的(Z)

[0339] 用Prolog表示即

[0340] w1(X),decide(X,Y,Z),is(Y,Z),decide(Y),wise(Z).

[0341] 下面举例两种典型的推理模式,来说明本发明的推理方法。

[0342] 逻辑推理模式一:

[0343] 自然语言似谓词演算形式的自动推理算法模型(三段式形式逻辑):

[0344] 1.在句式中以个体替代类属作为对类属这一变量的赋值,也可用类属替代个

体——这要看推理目标的需要。终极的抽象概念(如“秩序”与“紊乱”，“好”与“坏”等成对的两极抽象概念)与被它所包括的具体概念之关系均如此。

[0345] 2.在句式中,如果说明子句与主体表达相矛盾,则句子无效。

[0346] 3.同一行为链中具有因果关系链的行为主体,在句式中可按照因果次序前者替换后者(这由因果关系的时空限定的唯一性所决定)。

[0347] 4.What<sub>1</sub> BE What<sub>2</sub>反过来也是一样,What<sub>2</sub> BE What<sub>1</sub>,二者等价。

[0348] 5.What<sub>1</sub> Do What<sub>2</sub>模式中,Do等价于What<sub>2</sub>。

[0349] 6.作为句中第三项的W<sub>j</sub>DW<sub>j+1</sub>都可以独立地提出来,有完整的意义,并与其他WDW进行归结合一并形成新句子。而且作为第三项,可以取消W<sub>j</sub>D,只保留W<sub>j+1</sub>,句子缩减为中心意义句。

[0350] 7.将句中最初切分为四种基本思维模式之一的DO的限定表述消去,并将同一句中最后一个DO的WHAT保留,如果WHAT是BE型的,则BE两边等值而形成两个简化句。

[0351] 8.一段话的分解充分的基本模式中的三项参数,除BE(即谓词)外,在同一情境中等价,即可以互相替代。可将这一段中每一句话中的相同的词用同一个等价的词替换。

[0352] 下面,我们详述借助本发明的方法,将自然语言切分后通过相同词替代及全称、单称替代进行推理。人工智能所谓的谓词演算的归一算法与之相同。

[0353] 【实例】假设有这样一些事实表述的句子(括号内为操作之表述):

[0354] (1)Tony是Graham的朋友;(Tony,是,Graham的(朋友))

[0355] (由原句“Tony是Graham的朋友”,得出

[0356] Tony是(Graham的)朋友。由这句话,因为没有特别说明,按照名字可知

[0357] Graham和Tony均是人。)——将句子切分为思维基本模式之一。

[0358] (2)Graham的所有朋友都喜欢酒;(Graham的(朋友),喜欢,酒)

[0359] (由原句“Graham的所有朋友都喜欢酒”,将

[0360] “所有……都”取消,获得——将句子中“所有……都”取消。

[0361] “Graham的朋友喜欢酒”,然后用——将句子中变成抽象(无具体限定)表述。

[0362] Tony置换“Graham的朋友喜欢酒”中的“Graham的朋友”,则得出

[0363] “Tony喜欢酒”。——将句子中“朋友”具体化。赋值,即给出确定值。

[0364] “所有……都”只是强调语气而已)

[0365] (3)每个喜欢酒的人都喝酒精饮料。(喜欢酒的(人),喝,酒精饮料)=((人,喜欢,酒),喝,酒精饮料);(人,喜欢,酒)与(Graham的(朋友),喜欢,酒)结构相同,则可以Graham的(朋友)替代“人”;而Tony,是,Graham的(朋友),所以Tony可以替代“人”。替换后有

[0366] (喜欢酒的(Tony),喝,酒精饮料),即

[0367] (Tony,喝,酒精饮料),同时,酒=酒精饮料。替代后又有

[0368] (喜欢酒精饮料的(Tony),喝,酒),即

[0369] Tony喜欢酒精饮料。

[0370] (由原句“每个喜欢酒的人都喝酒精饮料”,将

[0371] “每个……都”取消,获得——将句子中“每个……都”取消。

[0372] “喜欢酒的人喝酒精饮料”,——将句子中变成抽象(无具体限定)表述。

[0373] 因为“Tony是人”,和“Tony是喜欢酒的人”,用

[0374] Tony置换“喜欢酒的人喝酒精饮料”中的“喜欢酒的人”，则得出

[0375] “Tony喝酒精饮料”，同时，“酒是酒精饮料”。然后用

[0376] “酒精饮料”置换“Tony喜欢酒”中的“酒”，则有

[0377] “Tony喜欢酒精饮料”。——用等值的词置换。思维基本模式中两边是等值的。

[0378] “所有……都”只是强调语气而已)

[0379] (4)推理得出结论:Tony喜欢酒精饮料。

[0380] 这些事实表述的句子,因为相互有共同的构成主题,所以,可以构成一个复合句或段落:

[0381]  $S_1 = (Is, What_1, What_2(likes, What_2, What_3 \wedge (drinks, What_4, What_5)))$

[0382]  $What_1 = Tony; What_2 = Graham$ 的所有朋友; $What_3 = 酒; What_4 = 每个喜欢酒的人$ (自然包括Tony) =  $(likes, What_2, What_3); What_5 = 酒精饮料$ 。其中 $What_2$ 与 $What_4$ 包含 $What_1$ ,  $What_5$ 包含 $What_3$ ,分别是类属(抽象的概念)与个体的关系,二者相互依存和部分特性一样,故某些情形中可相互替代。在句式中可以个体替代类属作为对类属这一变量的赋值,也可用类属替代个体——这要看推理目标的需要。所以上式

[0383]  $= (Is, What_1, (likes, What_1, What_3 \wedge (drinks, (likes, What_1, What_3), What_5)))$

[0384]  $= (Is, What_1, (drinks, (likes, What_1, What_5), What_5))$

[0385] 句子中已有了 $(likes, What_1, What_5)$ 这一结论,这句话意指“Tony是喝酒精饮料——Tony喜欢酒精饮料”。

[0386] 如果推理得出结论:Graham的所有朋友喜欢酒精饮料,则在句中是以 $What_2$ 替代 $What_1$ ,所以上式

[0387]  $= (Is, What_2, (likes, What_2, What_3 \wedge (drinks, (likes, What_2, What_3), What_5)))$

[0388]  $= (Is, What_2, (drinks, (likes, What_2, What_5), What_5))$

[0389] 如果将 $What_3$ 替代 $What_5$ 。(句子中以个体替代类属作为对类属这一变量的赋值),所以上式

[0390]  $= (Is, What_2, (drinks, (likes, What_2, What_3), What_3))$

[0391]  $= (Is, What_2, (drinks+likes, What_2, What_3))$ ,这里 $Is, What_2$ ,可省去,即有:

[0392]  $= (drinks+likes, What_2, What_3)$

[0393]  $(drinks, (likes, What_2, What_3), What_3)$ 意指“Graham的所有朋友喜欢酒——因而饮酒”。推理得出结论:Graham的所有朋友饮酒。

[0394] 句子的主题词(表达的主体)由动机和语境决定。上述是以Tony为主题,则推理结论表达Tony。以Graham的所有朋友为主题,则有了“Graham的所有朋友饮酒”的推论。

[0395] 逻辑推理模式二

[0396] 由自然语言信息转换的似谓词演算形式进行演算的推理算法模型:

[0397] (1)按时间优先原则,取首句第一组数列中的第三项减次句中第一组数列中的第一项,如结果为0,并且首句第一组数列中的第二项减次句第二组数列中的第二项的结果也为0,则将次句第二组的第一项替代首句第一组数列的第三项;新的首句第一组数列构造完成;

[0398] (1.1)如未有上述结果,则取首句第一组数列中的第三项减次2句第三组数列中的第一项,如结果为0,并且首句第一组数列中的第二项减次2句第三组数列中的第二项的结

果为0,则将次2句第三组的第一项替代首句第一组数列的第三项;新的首句第一组数列构造完成;

[0399] (2)取第一组数列中的第三项减去已成功完成步骤1运算的次组数列中的第一项,如结果为0,并且第一组数列中的第二项减去已成功完成步骤1运算的次组数列中的第二项的结果为0,则将已成功完成步骤1运算的次组数列中的第一项替代第一组数列的第三项;新的第一组数列构造完成;

[0400] (3)继续上述过程直至无法完成上述过程,中止;输出新的第一组数列,这即是推理结果;

[0401] (4)如果所选的一组数列无法完成上述过程,则选择次一组数列进行上述过程;

[0402] (5)有推理结果或无推理结果,输出结论。

[0403] 【实例】将下面自然语句完全地转换为谓词形式并完成推理:

[0404] 任何通过历史考试和中奖的人是快乐的,

[0405] 任何努力学习或者幸运的人能够通过所有考试,

[0406] 约翰不学习但是很幸运,

[0407] 任何幸运的人能中奖,

[0408] 推理证明:约翰是快乐的。

[0409]  $\forall x (\text{pass}(x, \text{history}) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x))$

[0410] 下面是中文句子与谓词形式完全对应过程,并可知能将二者通过机器自动转换。

[0411]

**S1=任何通过历史考试和中奖的人是快乐的**

**= (任何通过历史考试和中奖的人, 是, 快乐的)**

**=任何 ((通过历史考试和中奖的人), 是, 快乐的)**

**= $\forall$ 人 ((人, 通过, 历史考试)  $\wedge$  (人, 中, 奖)), 是, 快乐的)**

**= $\forall$ 人 (通过 (人, 历史考试)  $\wedge$  中 (人, 奖)), 是, 快乐的 (人))**

**= $\forall$ 人 (通过 (人, (考试, 是, 历史))  $\wedge$  中 (人, 奖)), 是, 快乐的 (人))**

**= $\forall$ 人 (通过 (人, 考试)  $\wedge$  通过 (人, 历史))  $\wedge$  中 (人, 奖)), 是, 快乐的 (人)),**

**S1= $\forall$ 人 (通过 (人, 考试)  $\wedge$  通过 (人, 历史))  $\wedge$  中 (人, 奖)  $\rightarrow$  (是) 快乐的 (人))。**

[0412] 从此句可得到最基础的核心句义,即IF(人,通过,历史考试) $\wedge$ (人,中,奖)THEN(人,是,快乐的),由三项组成的表述式,与谓词演算直接转换。

[0413] 以x取取代抽象主语词“人”,另外取“通过(人,考试)”和中(人,奖),得到

**S1= $\forall x$  (通过 (x, 考试)  $\wedge$  中 (x, 奖))  $\rightarrow$  快乐的 (x))**

[0414]

**= $\forall x$  (pass (x, exam)  $\wedge$  win (x, lottery)  $\rightarrow$  happy (x))**

[0415] 注:(考试,是,历史)表示“考试”与“历史”等价,为了简洁,可取“通过(人,考试)”,当然,也可取“通过(人,历史)”。

[0416]

$$\begin{aligned}
S1 &= \forall x (\text{pass}(x, \text{history exam}) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x)) \\
&= \forall x (\text{pass}(x, \text{history\_exam}) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x)) \\
&= \forall x (\text{pass}(x, \text{is}(\text{exam}, \text{history})) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x)) \\
S1 &= \forall x (\text{pass}(x, \text{history}) \wedge \text{win}(x, \text{lottery}) \rightarrow \text{happy}(x))
\end{aligned}$$

[0417] 将下面自然语句完全地转换为谓词形式:

[0418] ②任何努力学习或者幸运的人能够通过所有考试。

[0419]  $\forall x \forall y (\text{study}(x) \vee \text{lucky}(y) \rightarrow \text{pass}(x, y))$ 

[0420] “能够”、“原来” = “→”。另外，“决定”的行为同时隐含着已经“计划”了什么。所以，“决定‘WHAT’” = “计划‘WHAT’”。这是描述行为时，因行为的连续性、因果关系而决定的概念之间的关系。

[0421]

S2=任何努力学习或者幸运的人能够通过所有考试

$$\begin{aligned}
&= \forall ((x \text{人, 努力, 学习}) \vee (y \text{人, 是, 幸运的}), \text{能够通过, 所有考试}) \\
&= \forall x \forall y ((\text{努力}(x, \text{学习}) \vee \text{幸运的}(y) \rightarrow ((x, y), \text{通过, 考试})) \\
&= \forall x \forall y ((\text{努力}(x, \text{学习}) \vee \text{幸运的}(y) \rightarrow \text{通过}((x, y), \text{考试})) \\
&= \forall x \forall y ((\text{strive}(x, \text{study}) \vee \text{lucky}(y) \rightarrow \text{pass\_exam}((x, y)) \\
&= \forall x \forall y ((\text{strive}(x, \text{study}) \vee \text{lucky}(y) \rightarrow \text{pass}(x, \text{exam}) \vee \text{pass}(y, \text{exam}))
\end{aligned}$$

[0422] 从此句可得到最基础的核心句义,即IF(x人,努力,学习)∨(y人,是,幸运的)THEN((x,y)人,通过,考试),由三项组成的表述式。

[0423] “通过考试”与“通过”在句中等价,为了简洁,可取“pass(x,y)”。

[0424] S2= $\forall x \forall y (\text{strive}(x, \text{study}) \vee \text{lucky}(y) \rightarrow \text{pass}(x, y))$ 

[0425] 将下面自然语句完全地转换为谓词形式:

[0426] ③约翰不学习但是很幸运,

[0427]  $\neg \text{study}(\text{john}) \wedge \text{lucky}(\text{john})$ 

[0428] S3=约翰不学习但是很幸运

[0429] =((约翰,不,学习)∧(约翰,是,很幸运))

[0430] =((约翰,不,学习)∧(约翰,是,(幸运,是,很的)))

[0431] 从此句可得到最基础的核心句义,即(john,不,学习)∨(john,是,幸运的),得到

S3= $\neg \text{study}(\text{john}) \wedge \text{lucky}(\text{john})$ 

[0432] ④任何幸运的人能中奖,

 $\forall x (\text{lucky}(x) \rightarrow \text{win}(x, \text{lottery}))$ 

S4=任何幸运的人能中奖

=∇(幸运的人→中奖)

[0433]

=∇((人,是,幸运的)→(人,中,奖))

=∇x(幸运的(x)→中(x,奖))

=∇x(lucky(x)→win(x,lottery))

[0434] 下面看推理证明:

- [0435] 由sentence(x,中,奖):-WBW(x,是,幸运的).
- [0436] (john,是,幸运的):- (john,是,x),得到
- [0437] sentence(john,中,奖):-WBW(john,是,幸运的).
- [0438] 再由sentence(x,是,快乐的):-WDW(x,通过,考试),WDW(x,中,奖).得到
- [0439] sentence(john,是,快乐的):-WDW(john,中,奖)
- [0440] 推理证明:约翰是快乐的。
- [0441] 将上例S1、S2、S3、S4分别泛化为句式,则有
- [0442] S5=任何通过历史考试和中奖的人是快乐的。
- [0443] ……的人→“人”泛化为变量x;x前是以“的”为标识的一个或多个WDW或WBW限定成分。
- [0444] ……的人是→“是”泛化为变量BE;
- [0445] ……的人是快乐的→“快乐的”泛化为变量y,y域为限定单词类集。
- [0446] 换一下表达方式,有:
- [0447]

**S1**=任何通过历史考试和中奖的人是快乐的

= (任何通过历史考试和中奖的人, 是, 快乐的)

= (任何(通过历史考试和中奖的人), 是, 快乐的)

= ( $\forall$ 人((人, 通过, 历史考试)  $\wedge$  (人, 中, 奖)), 是, 快乐的)

= ( $\forall$ 人(通过(人, 历史考试)  $\wedge$  中(人, 奖)), 是, 快乐的(人))

= ( $\forall$ 人(通过(人, (考试, 是, 历史))  $\wedge$  中(人, 奖)), 是, 快乐的

(人))

= ( $\forall$ 人(通过(人, 历史)  $\wedge$  中(人, 奖)), 是, 快乐的(人))

= ( $\forall$ x(通过(x, 历史)  $\wedge$  中(x, 奖)), 是, 快乐的(x))

=  $\forall$ x (pass (x, history)  $\wedge$  win (x, lottery)  $\rightarrow$  happy (x))

- [0448] 从此句可得到最基础的核心句义,即IF(人,通过,历史考试) $\wedge$ (人,中,奖)THEN(人,是,快乐的),由三项组成的表述式,与谓词演算直接转换。

[0449] sentence(人,是,快乐的):-WDW(人,通过,历史考试),WDW(人,中,奖).

[0450] WDW(人,通过,历史考试):-

[0451] WDW(人,中,奖):-

[0452] 回应这句话,可以从上面三个基本句子引起。

[0453] 注:(考试,是,历史)表示“考试”与“历史”等价,为了简洁,可取消“考试”一词。并将“人”一词替换为变量x。

**S1**= $\forall$ x (pass (x, history exam)  $\wedge$  win (x, lottery)  $\rightarrow$  happy (x))

[0454]

= $\forall$ x (pass (x, history\_exam)  $\wedge$  win (x, lottery)  $\rightarrow$  happy (x))

[0455] 比较上式 $\forall$ x (pass (x, history)  $\wedge$  win (x, lottery)  $\rightarrow$  happy (x))

[0456] (五)句子生成

[0457] 句式切分的回溯过程就是新句子的生成过程。在句式切分基础上,依据相同字词的替代而完成的逻辑推理过程,将产生新的最简句式(最简思维模式)的集合,进行切分的逆向回溯,就会产生一个新的句子作为输出结果,从而实现机器与人的逻辑思维的结果相

同。

[0458] 1.生成句子就是以最简单的基本思维模式为基础,通过增加WHAT——名词或代词主语和DO——动词谓语的限定成份而逐步形成。设定determiner[x]为限定词集合,把限定词集合与WHAT和DO结合在一起,四种基本思维模式就组成了自然语言句式。生成语句是语句切分后的反操作,一种回溯方式。由于我们的计算机的数据库都以最简式——即似谓词演算的形式构成,所以可以很便捷地把知识重新组合成语句输出。句式切分的回溯过程就是新句子的生成过程。在句式切分基础上,依据相同字词的替代而完成的逻辑推理过程,将产生新的最简思维模式的集合,进行切分的逆向回溯,就会产生一个新的句子作为输出结果(与人的逻辑思维的结果相同)。

[0459] 【实例】依据语句内在相同成分而形成的复合语句

[0460] 依据最简式或简单的语句内在相同成分而形成复合语句,即通过逆分解过程回溯并泛化生成句式。

[0461] 例如下面的语句:

[0462] 建筑风格是维多利亚式。

[0463] 按照句式切分方法,可以得到

[0464] 底层:(建筑,有,风格)

[0465] 顶层:(风格,是,维多利亚式)

[0466] 建筑(x)=building(x)

[0467] 风格(x,y)=style(x,y)

[0468] 维多利亚式(y)=victorian(y)

[0469] 风格(z).=style(z)

[0470]  $\forall x \exists y \forall z \text{ building}(x) \wedge \text{style}(x, y) \wedge \text{style}(z) \rightarrow \text{victorian}(y)$

[0471] 泛化句式是

[0472] (what1,HAS,what2)和

[0473] (what2,IS,what3)

[0474] is\_contained\_in(what3,what2):-location(what3,what2).

[0475] (what3属于what2类集。)

[0476] 回溯得到语句“what1 what2 IS what3”。

[0477] 逆分解过程回溯上面的例句:

[0478] 将(风格,是,维多利亚式)代入(建筑,有,风格)中,操作步骤是

[0479] (1)将(建筑,有,风格)复原,取消“有”及括号内逗号,得到(建筑风格);

[0480] (2)由(风格,是,维多利亚式),取消括号内逗号,替换(建筑风格)中的“风格”,得到:

[0481] (3)建筑风格是维多利亚式。

[0482] 注:上述泛化句式的各项可能是词组或语句。

[0483] 采用本发明的方法,可以实现对自然语言整篇的进行切分,这是实现机器思维与人类思维完全一致的基础。

[0484] 采用本发明的方法,可以实现对两种不同的语言通过切分后进行机器自动翻译。本发明方法依据人类思维的本质模式,所以,不同语言体系只有表达方式的细节差异,基本



结构是相同的。切分后的最简思维模式之集合会有另一种语言相对应,翻译即是将最简思维模式中的词汇相应转换为匹配的另一语言的词汇。

[0485] 采用本发明的方法,可以实现对数学的机器思维完全以自然语言表达并自动完成解决问题过程。通过泛化处理,可以归纳出各种类型题的题目的自然语言似谓词形式转换后的解题模式。

[0486] 采用本发明的方法,切分后的语句变成了由三项字词构成的最简思维模式的集合,运用搜索匹配识别,进行递归替代等计算过程而实现自动推理、联想等思维过程。

[0487] 采用本发明的方法,通过词(概念)的格式塔结构维度的相似度,机器可以自动搜索识别,找出同类属的概念(词)和相反的概念(词),表述行为的系列词汇也可以同样的方法寻出,并作为学习结果储存入字词数据库。

[0488] 采用本发明方法,对自然语言表述进行似谓词形式的转换,泛化的基础上实现机器自动编程。

[0489] 本发明的一种自然语言机器识别系统,包括人机接口模块、句子切分模块、中央处理单元、句式合成模块及数据库模块,所述的句子切分模块和句式合成模块分别与所述中央处理单元的输入端和输出端通过电信号相连,所述的数据库模块为多库协同模块,包括字词库管理模块、知识库管理模块、情景库管理模块、多元语义网络库管理模块及隐喻网络库管理模块。

[0490] 以下为本发明的多库协同模块的建立方法:

[0491] 建立字词库管理模块的新代码字词库

[0492] 1.采用自然数编码的新代码字词库,所述新代码编码数据库通过手工输入或利用现有的开放代码源的字词库输入字词的方式来建立。

[0493] 建立数字代码字词库,并能输出相应数字(如:0,1,2,……9)用本身数字作代码,另加数位标识代码。数字以数位标识和单数字共同定义。词义特定的如“是、在、的、地”等和标点符号,以及词类确定的(如植物名称、专有名词等)用某个区间的数字作代码。作为表达行为的词,在句式位置决定其词性(动词性或名词性),这部分词在字词库的构成及形式有不同情形而有不同输出。

[0494] 在字词库建立各种行为动作模式,作为最简式中间项“DO”的组成部分;

[0495] 考虑到为词的常规搭配及固定词组,建立字词库初期赋词代码数字之间留足间隔。

[0496] 在构造字词库时,将词之间的固有搭配变为连续的两个或多个数之数列,固有搭配的词组或成语也具有唯一性,那么,对于原有字词库中没有的统计概率大的搭配一起的词,将其作为固有搭配的词组进入字词数据库。这将成为机器自动学习方式之一。

[0497] 建立了字词库后,该库具有纳入输入字词、进行搜索比对,并输出相应数字代码的功能。

[0498] 2.建立句式切分模块。句子切分模块与词性分类的字词库相联,并有分类的寄存地址所构成的句式。分层将数列句子分成由三项的数或数组组成的数列。

[0499] 由字词库管理模块输出的数组替代原句子,由之一个句子变为一组数列,并进入句式切分模块。句式切分模块具备搜索比对功能,按自然顺序逐一进行。识别采用减法的计算,结果为0即成功,字词库功能之一是比对识别。

[0500] 由两个字以上构成的词组比对识别的算法模型:

[0501] 1.按先后次序,先在作为词组代码的数之区间判别,再依次将作为词组代码的一数减字词数据库该区间首遇的第一个代码数,差为零,则该词组第1个字为首遇的数代码定义的字;如果差不为零,则余数依次减后面的代码数,直至差为零,另一字识别方式相同,由之,词组均得到比对识别。如果差数为非0,则判断字词库无此词组。

[0502] 2.如果1.判断未完成与该词组比对识别。则转入下一词组比对;

[0503] 2.1.重复上述1.过程。如果仍未完成与词组比对识别,则

[0504] 3.重复上述过程,直至完成比对识别。如果比对成功或比对失败,则

[0505] 4.输出结果。

[0506] 以似谓词演算形式定义并诠释一个词,建立自然语言理解的字词库

[0507] 词的定义是最主要的基础工作,应在高度统一的层面开始,逐层将格式塔结构维度确定为某词的基础意义。层面越低越具体化——这就是现在词典的表述方式。

[0508] 利用一系列的谓词形式定义一个词,再构造出相应的例句句式,建立自然语言理解的词典。以此方法建立常用字词典。例如,在《汉字源流字典》p22有“才”的释义,按我们的方法可表达为:

[0509] 原义:“才”像草木之初,由之可知,衍生出的意义都有隐喻的特性。

[0510] (才,像,草木之初)→(才,是,初始)→(才,是,刚刚)

[0511] (才,像,草木之初)→(才,是,质性)→(才,是,能力)→(才,是,有能力的人)

[0512] (才,像,草木之初)→(才,是,质性)→(才,是,才能)→(才,是,有才能的人)

[0513] (才,像,草木之初)→(才,是,草木)→(才,是,木材)∧(才,是,材)→(才,是,材料)→(才,是,棺材)

[0514] 一个词以似谓词演算形式表示是为了转换为谓词形式,从而进行谓词演算和用Prolog表示。这样,词与词之间的匹配就变成了谓词演算的递归、合一的替换操作。

[0515] 上面的每一条关于“才”的定义可以写成谓词形式

[0516] BE(y,Z),或

[0517] Z(y).

[0518] “人”与“才”在语句中的匹配就有了内在的逻辑运算的依据。在涉及“人”或拟人化的描述中,有关“人”或类人者之质性的描述,“人”或类人者的词与“才”匹配。譬如,人之有才能的人,组合成“人才”、“才子”、“男才女貌”、“你太有才了!”等匹配。

[0519] 下面看如何利用上面词之定义的方法,通过同义词替换,自动生成同义句:

[0520] “你太有才了!”原句为“你太有能力了!”由(才,是,能力)有(能力,是,才),“是”两边等价。将(能力,是,才)代入“你太有能力了”,以“才”替换“能力”一词,就得到“你太有才了!”

[0521] 用Prolog的语言来表达,有

[0522] [DO](x,y):-[DO](x,z),BE(y,z).

[0523] [DO](x,z),

[0524] BE(y,z).

[0525] 建立字词库(词典)的算法模型:

[0526] 1.选择一个简单字词进行定义(如上面“才”之定义),建立词库,依次将第一个词

定义中出现的词加以定义；

[0527] 2. 当下词定义完毕, 重新进行1. 操作。

[0528] 按此算法, 将上面“才”之定义中出现的第一个词“像”加以定义, 然后依次是“草, 木, 之, 初, ……”, “棺”。此轮定义完毕, 则从第二个出现的词“草”开始新的定义过程, 此轮定义完毕后, 又重新进行相似操作。此算法经过编程, 将可以利用已有电子辞典而自动获取该辞典中的释词而建立词库。

[0529] 词典的另一个用途是传统的查字。由于词典采用谓词形式建立, 即一个词都以语句的形式存在于数据库中。作为例句的句式则可直接输出并自动依照当前语句替换相关词汇, 谓词形式为这种替换提供了可能。

[0530] 建立词汇数据库

[0531] 将词汇按事物类别分列构成“词林”, 并将常用词组合列入单词名下。参考《诗韵合璧》作法, ([1]清·汤文璐编. 诗韵合璧. 上海: 上海古籍书店, 1982.) 摘取书中p1内容如下。

[0532] 根据《诗韵合璧》词林典股目录·天文门, 扩大词汇:

[0533] 天·日月

[0534] [日] = {(春日, IS, (日子, IS, 春季的)) = (春日, IS, 日子) ∧ (日子, IS, 春季的), (夏日, IS, 日子) ∧ (日子, IS, 夏季的), (秋日, IS, 日子) ∧ (日子, IS, 秋季的), (冬日, IS, 日子) ∧ (日子, IS, 冬季的)}

[0535] 下面词的数据库组成形式与[日]相同。

[0536] [月] = {新月, 残月, 月桂, 中秋月} ∧ {戴月, 岁月, 日月}

[0537] [星] = {天河, 云, 庆云, 云峰} ∧ {愁云, 云山, 云雨, 云飞, 云锁, 星光, 披星}

[0538] 天·风雨

[0539] [风] = {春风, 夏风, 秋风, 冬风}

[0540] [雨] = {夜雨, 喜雨, 黄梅雨, 久雨, 春雨, 夏雨, 秋雨, 冬雨}

[0541] [雷] = {}

[0542] [电] = {}

[0543] [虹] = {}

[0544] [霞] = {}

[0545] [露] = {}

[0546] [霜] = {}

[0547] [雪] = {喜雪, 春雪}

[0548] [雾] = {}

[0549] [霁] = {}

[0550] [烟] = {游丝} 空中飘浮着游丝, 游丝即细细的烟。

[0551] [晴] = {}

[0552] [阴] = {}

[0553] 建立情景库

[0554] 建立情景词汇和专门词汇数据库的算法模型:

[0555] 1. 情景构成(字词库或知识库构成): 以谓词演算形式的表达式聚集而构成。包含: 事物命名; 时间与空间特性; 限定词; 回应语句自动生成及输出, 等等。应能通过机器自

动学习增加知识。

[0556] 2. 将描述情景中各主要物体及相关事件的字词以数列表达, 作为词库的子库。此子库中各字词有紧密的相关性。

[0557] 3. 情景库由情景特征词汇为构成基础。在此基础上, 增加可被用于新的句子、段落生成的描述句及段落, 此由数列表达。

[0558] 4. 大的方面分, 词义分为: 表事、物命名的; 表时间、空间存在、位置、数量特性的; 表行为动作的; 表状态的; 表抽象限定的如“是”、“在”等。

[0559] 参考图文词典建立分类的情景词汇和专门词汇数据库。譬如参考《英汉图文对照词典》, 建立起非常丰富的语境数据库和生成句子用的词汇数据库, ([2]《英汉图文对照词典》翻译组. 英汉图文对照词典. 上海: 上海科学技术出版社, 1984.) 摘取书中p265咖啡室, 茶室内容如下:

[0560] (咖啡店(室), 有, X1|X2|X3|...|Xn[X])

[0561] [X]表示情境中构成部分或特征集合。

[0562] 柜台(糕点柜台){X1}, ——{X1}是关于柜台特性、特征及样式的列表, 下面{X}也均是同类的列表。就样式(style)而言, 如果需要“样式”输出, 可自动联结。——(咖啡店(室), 有, 柜台(糕点柜台)[X])

[0563] 咖啡壶{X2}, ——(咖啡店(室), 有, 咖啡壶[X])

[0564] 放货币盘子{X3}, ——(咖啡店(室), 有, 放货币盘子[X])

[0565] 蛋糕{X4}, ——(咖啡店(室), 有, 蛋糕[X])

[0566] 带搅奶油蛋白甜饼{X5}, ——(咖啡店(室), 有, 带搅奶油蛋白甜饼[X])

[0567] 制糕点学徒{X6},

[0568] 柜台女服务员{X7},

[0569] 报架{X8},

[0570] 壁灯{X9},

[0571] 拐角软垫沙发{X10},

[0572] 咖啡室小圆桌{X11},

[0573] 大理石台面{X12},

[0574] 女侍应生{X13},

[0575] 托盘{X14},

[0576] 一瓶柠檬水{X15},

[0577] 柠檬水杯{X16},

[0578] 对局下棋者{X17},

[0579] 咖啡餐具{X18},

[0580] 一杯咖啡{X19},

[0581] 方糖小钵{X20},

[0582] 奶油壶{X21},

[0583] 咖啡室顾客{X22},

[0584] 先生{X23},

[0585] 女士{X24},

- [0586] 正在看报的人{X25},
- [0587] 报纸{X26},
- [0588] 报夹{X27}。
- [0589] 以上词汇表达那个时代英国咖啡室,茶室主要内容特征。
- [0590] 任何词汇的意义都与语境相关,所以,非常重要是,指明词义是在何种情境中才成立的。这是词汇数据库中的词的标识之一。
- [0591] 建立隐喻网络库
- [0592] 建立用于隐喻生成的词汇数据库,也需要象上面那样去做。
- [0593] 建立可输出的能用于生成语句的知识库。
- [0594] 将知识全部表达为谓词形式,以此建立能用于生成语句的知识库可以很方便地组合成语句输出。知识库中知识的谓词形式不一定非基本思维模式的最简式不可,也可部分为复式。
- [0595] 如果是WBW的简式,即只有形容词和名词组成的词组NP,那么由ADJ N有
- [0596] (noun(x),BE,adj(y)),或(noun(x),BE,noun(y))。当名词为常项N时,由
- [0597] (noun(N),BE,adj(y))或(noun(x),BE,noun(y)),
- [0598] 可知不同的形容词或名词y同属于N的类属集合。按照上式,将在文章中搜索到的y自动地归属到N的类属集合,由此实现机器自动学习。
- [0599] 如
- [0600] N=装饰风格[古罗马式,拜占庭式,维多利亚式,洛可可式,巴洛克式,中国汉代式,中国明式,中国清式……]
- [0601] 实例:设输入语句(建筑,有,风格)和(风格,是,简洁的)。这两条语句由原句“建筑有简洁的风格”切分而来。
- [0602] 在建筑条目下有建筑造型样式的知识,其中有一条对建筑的定义是:
- [0603] (建筑,有,风格)——知识语句存在形式是(建筑,有,风格)
- [0604] 依据“建筑”、“风格”这两个词可以展开相关知识的自动搜索。
- [0605] 而在艺术条目下有建筑造型样式的知识,其中有“风格”词目,风格包含许多不同的样式,关于建筑造型的有:
- [0606] 建筑风格[古罗马式,拜占庭式,维多利亚式,洛可可式,巴洛克式,现代式,中国汉式,中国藏式,……]
- [0607] 知识语句存在形式是(风格,是,样式(x))。
- [0608] 其中的样式之一是“现代式”,关于“现代式”的知识语句存在形式是(现代式,是,简洁的)。
- [0609] 还需要有一个另外的约束条件才能在(风格,是,样式(x))中确定与知识语句(建筑,有,风格)的最佳匹配。这另一个约束条件就来自另一个输入的语句(风格,是,简洁的)。
- [0610] 当搜索进行到知识语句(风格,是,样式(x))类集之后,还可以再搜索到下一层面的知识语句类集,这是定义样式的状态类属集合:
- [0611] (样式(x),是,adj(y))
- [0612] 按照相同词进行递归操作,可以得到
- [0613] (现代式,是,简洁的)——“是”两边等价。

- [0614] 代入(风格,是,简洁的),得到
- [0615] (风格,是,现代式)
- [0616] 再与(建筑,有,风格)进行回溯,得到新的语句:
- [0617] 建筑风格是现代式。
- [0618] 以此语句作为输出。
- [0619] 这样由输入的原句“建筑有简洁的风格”,得到输出语句“建筑风格是现代式”。
- [0620] 7. 建立多元语义网络库管理模块
- [0621] 以下语句表述一条知识:
- [0622] 小燕从春天到秋天占有一个巢。
- [0623] 传统方法是
- [0624] 二元关系的合取:
- [0625] ■start(占有1,春天),finish(占有,秋天),
- [0626] ■owner(占有1,小燕),owner(占有1,巢1)
- [0627] 采用本发明的方法,完全用似谓词演算形式的自然语句表示,并进一步以Prolog语言自动编程,将一种知识片断建立起来,构成语义网络的一部分:
- [0628] 小燕从春天到秋天占有一个巢
- [0629] =(小燕从春天到秋天,占有,一个巢)
- [0630] =((小燕,(时,从,春天(到,秋天)),占有,一个巢)
- [0631] =(小燕,占有,一个巢) $\wedge$ (时,从,春天(到,秋天))
- [0632] =(小燕,占有,一个巢) $\wedge$ (小燕,从,春天(到,秋天)) $\wedge$ (占有,从,春天(到,秋天)) $\wedge$ (时,从,春天(到,秋天))———①
- [0633] 补充:(小燕,占有,一个巢) $\rightarrow$ (燕,占有) $\rightarrow$ (燕,是,占有者),———(自动生成)
- [0634] (燕,是,鸟)。
- [0635] 利用Prolog语言自动编程,将上式①表述的知识片断建立起来,构成关于鸟类的语义网络的一部分。
- [0636] “小燕从春天到秋天占有一个巢”程序这样编写:
- [0637] program swallows(x):-own(swallow,一个巢),own\_start(swallow,春天),own\_finish(swallow,秋天),start(own,春天),finish(own,秋天),start(time,春天),finish(time,秋天),owner(swallow),swallows(x).
- [0638] swallows\_report:-
- [0639] write('Known swallows are:'),nl,swallows(x),write(x),nl,
- [0640] fail.
- [0641] 把这个程序调入Listener中,运行swallows\_report.。
- [0642] 采用本发明方法,对自然语言表述进行似谓词演算形式的转换,泛化的基础上实现机器自动编程。
- [0643] 本发明的方法在隐喻、联想机制中的表达
- [0644] 按照本发明方法,首先以格式塔结构维度为基础将信息与知识形式化,与推理机制结合,并以前者为基础进行联结、转换,实现联想的思维过程和结果。而且,一般在一个语句中词汇多数都有扩展意义,这用格式塔结构维度将它表达出来,“词”的格式塔结构维度

正是表述的词的“语境”。从下面用BNF(Bachus-Naur Form)表述式来表示“我在读书”句子的理论分析,可见其独特性。

[0645] a.首先按照词义和位置进行转换

[0646] 赋值语句::=匹配<WHAT的定义>|设<WHAT的定义>

[0647] b.WHAT(1,2)的定义::=<WHAT(1,2)的名称><具有><格式塔结构维度集合>|<格式塔结构维度集合><属于><WHAT(1,2)的名称>

[0648] c.动词::=具有|属于

[0649] d.格式塔结构维度集合::=<维度1>,<维度2>,<维度3>,...<维度n>

[0650] e.格式塔结构维度::=<空间限定>,<时间限定>,<量>,<相互关系>,……

[0651] f.WHAT(1,2)名称::=(WHAT(1,2))<设定的符号>|(WHAT(1,2))<设定的符号>的格式塔结构维度集合

[0652] g.选择新的WHAT(R)的定义::=<WHAT(1)or WHAT(2)的名称><具有><格式塔结构维度集合>|<格式塔结构维度集合><属于><WHAT(1)or WHAT(2)的名称>

[0653] h.回应赋值语句::=设<what'(1,2)的定义><具有><格式塔结构维度集合>|<格式塔结构维度集合><属于><what'(1,2)>

[0654] i.由省缺的原WHAT(1,2)的格式塔结构维度中匹配提取某维度作为新句子的议题。

[0655] (g)~(i)的行为由“语境”决定。语境由已设定的模式中按目标状态、时间状态和空间状态等综合因素决定激活其一。

[0656] 【实例】首一中一尾(3段组合)模式是隐喻式表述,是抽象模式上一中一下(以大地重力为参照系,垂直向)、前一中一后(水平向)、左一中一右(水平向)的限定形态之一。这三种形态模式统一表达为一端(start)—中间(middle)—另一端(end)的离散的空间限定的终极抽象形式。用形式化的自然语言的表述方法,这属于关系概念集合smea。

[0657] 设Q为n维关系元,则有

$$Q = \begin{bmatrix} \text{smea}, W_1, \text{BE}, \text{SME}_1 \\ W_2, \text{BE}, \text{SME}_2 \\ \dots & \dots & \dots \\ W_N, \text{BE}, \text{SME}_N \end{bmatrix}$$

[0658]  $Q = (\text{smea}, (W_1, \text{BE}, \text{SME}))$  和n维关系元

[0659]  $\text{SME} = [\text{start}] \wedge [\text{middle}] \wedge [\text{end}]$

[0660] [start],[middle],[end]分别是一端(start)—中间(middle)—另一端(end)的概念集合。

[0661]  $W_1$ 是WHAT的一个常项。

[0662] 如婆媳关系:母—子—儿媳;社会成员角色分类关系:精英分子—社会大众—弱势群体。此二例均为隐喻式表述。

[0663] “不受苦中苦,难为人上人”则属于两极概念的结构模式。此类可表示为:

[0664] 不受苦中苦,难为人上人

[0665]

$$= \neg (\dots, \text{受}, \text{苦中苦}) \rightarrow (\dots, \text{难为}, \text{人上人})$$

$$= \neg (\dots, \text{受}, (\text{苦}, \text{BE}, \text{苦中更苦})) \rightarrow (\dots, \text{为} (\text{为}, \text{BE}, \text{难}), (\text{人}, \text{BE}, \text{人中更优}))$$

$$= \neg (\dots, \text{受}, \text{苦中更苦}) \rightarrow (\text{为}, \text{BE}, \text{难}) \wedge (\dots, \text{为}, \text{人中更优})$$

[0666] 两个相对矛盾终极概念表达为一端(start)——另一端(end),由上述方法表示即

[0667]  $SE = [\text{start}] \wedge [\text{end}]$ [0668]  $Q = (SE, (W_1, BE, SE)_a, w)$ 和n维关系元

$$Q = \begin{bmatrix} \text{smea}, W_1, BE, SE_1 \\ W_2, BE, SE_2 \\ \dots \dots \dots \\ W_N, BE, SE_N \end{bmatrix}$$

[0669] 隐喻式推理

[0670] 这是联想、隐喻的形成方式。

[0671] 在隐喻的源体与目标体关系中,相对应的组成元素即相似的格式塔结构维度。找出隐喻的可自动分解的操作模式就解决了创新性思维的可行模式问题。利用格式塔结构维度部分相似来达到创新。从思维的角度看,创新是隐喻的结果。

[0672] 我们在此以相互关系的征象之一——形态的“聚”的格式塔结构维度集合来加以说明:

[0673] “聚”的关系征象:①基元的数量2以上;②相互依存性;③形态相向的态势;④互补性;⑤相似性;⑥有序性。

[0674] 显然,直线形态不具有“聚”的特征,所以变曲的河流才有“聚气”的功能,这已反映在中国古代建筑“风水术”的相关理论中。我们再以情感之表述词“爱”的格式塔结构维度集合来加以说明:

[0675] 1)“爱”的格式塔结构维度:①非平衡性。即“爱”是倾向性的,偏向的。②控制关系。即对心灵的影响,对其有约束作用。③主体与客体(或称施方与受方)。施方受“爱”的控制最强。④相互作用关系,付出与回报同时存在。“爱”是付出(心灵的或行为的),但同时也使施爱方得到“愿望的满足”,即付出“爱”的愿望得到实现。⑤在刺激——反应关系中,“爱”最先显现为一种接受刺激的倾向,然后才引起生理——心理情感的相关激发体验,即进入“爱”的状态。这种接受刺激的倾向被称为“态度”。

[0676] 2)“恨”与爱方向相反,部分结构维度相同。但第⑤点不同,它具有排斥刺激的倾向。

[0677] 一个事物A由一系列由符号系统的概念所命名的中间过程和事物所组成,例如一座房屋,一台机器等等。将一个符号作为上述表示方式中的一个格式塔结构维度,则该事物就可以由一系列的格式塔结构维度所定义。显然,这是另一层面的格式塔结构维度表示方式,每一个维度又有与之相似的更深层面的格式塔结构维度集合所定义。如此递归至上述的最基础的格式塔结构维度表达方式。呈现为树形构成结构。变化某一层面的某个或某些维度的量值或进行相似事物的代换,就能使事物A发生形态甚至性质变异,而转变为另一个事物B,这就是我们创造思维的“机械化”方式。循着这一思路,利用人工智能技术对此进行表述,例如利用现在关于知识的产生式表示法,框架表示法和语义网络表示法,只是表达的内容有所不同。



[0678] 相似的概念在其格式塔结构维度集中,维度相同相似的数量大,例如,“肥”与“密”;“瘦”与“疏”作为现象描述时,都是对表面形态体量的描述,在量的维度是“多”或“少”的界定,它们每组都是相似的。而对于“全”与“缺”,除了“多”或“少”的界定,还有量的变化的表述,“缺”隐含着量未达标准。这两个概念更加抽象,并不是直接地对表面形态的表述。这组相反概念组与上述概念组部分相似维度数量是不同的,差异性更大些。差异性大的相似概念相互替代,创新性和创新程度大,但也可能被认为荒谬。下面以实例看看:

[0679] 隐喻的逻辑

[0680] 以“月光如水”来说明,

[0681] 月(“月”的特性集合——格式塔结构维度集合,就是“月”的抽象的语义网络):

[0682] 在“阴阳”概念上属“阴”。

[0683] 其形态与物理属性:(1)有光、但柔和;(2)圆形或扁圆形;(3)不发热;(4)光色与水色相似;(5)弱。

[0684] 其空间属性:(1)与太阳同在天上;(2)尺度较太阳为小;

[0685] 其时间属性:夜间

[0686] 其环境属性:(1)静寂;(2)温度低;(3)[视觉环境]模糊,故统一性极强。

[0687]  $G_{月} =$ 

月, 阴阳的隐喻属性, 阴
形态特征, 柔和的光
物理属性, 不发热
颜色, 偏蓝
强度, 弱
空间限定, 天上和形状确定
尺度, 比太阳小
时间限定, 夜间
物理环境属性, 静寂和温度低
视觉环境属性, 模糊

[0688] 水(“水”的特性集合——格式塔结构维度集合,就是“水”的抽象的语义网络):

[0689] 在“阴阳”概念上属“阴”。

[0690] 其形态与物理属性:(1)有光(反射光),但柔和;(2)有线状纹理(当风吹动时);(3)不发热,可使人产生凉的记忆;(4)、光色与月色相似;(5)弱;(6)可有声响;(7)清晰或模糊(依光照而定)。

[0691] 其空间属性:(1)、随处而形,有依附性;(2)、尺度不定

[0692] 其时间属性:(1)或静止或流动

[0693] 其环境属性:地上的诸多因素而定。

[0694]  $G_{水} =$  水, 阴阳的隐喻属性, 阴  
 形态特征, 柔和的光  
 物理属性, 不发热  
 颜色, 偏蓝  
 强度, 弱  
 空间限定, 随处而形  
 尺度, 不定  
 时间限定, 静止或流动  
 物理环境属性, 诸多因素而定  
 视觉环境属性, 模糊或清晰

[0695] 比较“水”与“月”之属性, 可见二者在光、色、温度、柔弱、模糊等诸方面同一, 即 $G_{月}$ 与 $G_{水}$ 相似度大, 故可以形成隐喻的关系。可用 $G_{月} \cong G_{水}$ 表示格式塔维度相似数量大的关系, 简称 $G_{月}$ 与 $G_{水}$ 相似。

[0696] 设 $g$ 为格式塔结构分维度。

[0697] 如果以似谓语形式表达, 即

[0698]

$G_{月} =$  月,  $g_{隐喻}$  (隐喻, 是, 阴 or 阳), 阴  
 $g_{特征}$  (特征, 属, 形态), (光, 是, 柔和)  
 $g_{特性}$  (特性, 属, 物理), 不发热, (光, 是, 不热)  
 $g_{颜色}$ , 偏蓝  
 $g_{强度}$ , 弱  
 $g_{空间性}$ , (月, 在, 天上) 和 (形状, 是, 确定)  
 $g_{尺度}$ , (月, 比...小, 太阳)  
 $g_{时间性}$ , 夜间  
 $g_{物理}$  (特性, 属, (环境, 是, 物理)), 不发热, (夜, 是, 静寂和温度低)  
 $g_{视觉}$  (特性, 属, (环境, 是, 视觉)), 模糊

[0699]	G <sub>水</sub> = 水, g <sub>隐喻</sub> (隐喻, 是, 阴 or 阳), 阴
	g <sub>特征</sub> (特征, 属, 形态), (光, 是, 柔和)
	g <sub>特性</sub> (特性, 属, 物理), 不发热
	g <sub>颜色</sub> , 偏蓝
	g <sub>强度</sub> , 弱
	g <sub>空间性</sub> , 随处而形
	g <sub>尺度</sub> , 不定
	g <sub>时间性</sub> , 静止或流动
	g <sub>物理</sub> (特性, 属, (环境, 是, 物理)), 诸多因素而定 (不确定)
	g <sub>视觉</sub> (特性, 属, (环境, 是, 视觉)), 模糊或清晰

## [0700] 【隐喻生成实例】

[0701] 虚拟(非感觉性)的思维中的信息或情感像一件东西,语言是思维的载体,词语像装载东西的容器,交谈像信息或情感的传递,按隐喻生成机制,用语言交流信息和情感,这个复杂概念由三个相关联的隐喻有机地组成一个整体(引自《语法隐喻和隐喻语法》(沈家煊讲稿p1),这就是下面的三个相关联的隐喻陈述:

[0702] ①信息或情感是一件东西

[0703] ②词语是装载东西的容器

[0704] ③交谈过程是物件的传递过程

[0705] 下面我们看利用自然语言如何从①、②及补充的另外一些相关知识得出③:

[0706] 1. 相似性:(信息或情感,像,一件东西)→隐喻:(信息或情感,是,一件东西)

[0707] 2. 相似性:(词语,像,装载东西的容器)→隐喻:(词语,是,容器)

[0708] =相似性:(词语,像,(容器,装载,东西))→隐喻:(词语,是,容器)

[0709] =相似性:(词语,像,容器)∧(容器,装载,东西)→隐喻:(词语,是,容器)→(词语,装载,东西)

[0710] 词语∈语言,思维单元∈思维

[0711] (语言,是,载体)∧(载体,装载,思维)

[0712] (词语,属于,语言)→(词语,装载,思维单元)

[0713] (容器,属于,载体)→(容器,装载,东西)

[0714] 3. 相似性:(交谈,像,信息或情感的传递)→隐喻:(交谈,是,信息或情感的传递)

[0715] =相似性:(交谈,像,(信息或情感的传递))→隐喻:(交谈,是,(信息或情感的传递))

[0716] =相似性:(交谈,像,(…,传递,信息或情感))→隐喻:(交谈,是,(…,传递,信息或情感))

[0717] =相似性:(交谈,像,传递)∧(…,传递,信息或情感)→隐喻:(交谈,是,传递)∧(…,传递,信息或情感)

[0718] 代入1. 式,得到

[0719] 相似性:(交谈,像,传递) $\wedge$ ( $\dots$ ,传递,一件东西) $\rightarrow$ 隐喻:(交谈,是,传递) $\wedge$ ( $\dots$ ,传递,一件东西)

[0720] 由物件 $\in$ 东西,代入上式得到

[0721] 相似性:(交谈,像,传递) $\wedge$ ( $\dots$ ,传递,物件) $\rightarrow$ 隐喻:(交谈,是,传递) $\wedge$ ( $\dots$ ,传递,物件) $\rightarrow$ (交谈过程,是,传递过程) $\wedge$ ( $\dots$ ,传递,物件)

[0722] ( $\dots$ ,传递,物件)是对“传递”的限定,“传递”可以变换为“物件的传递”,所以,回溯得到

[0723] 交谈过程是物件的传递过程。

[0724] 泛化上述的变换,

[0725] ( $\dots$ ,DO,W)等价于WS DON

[0726] WS是原名词形容词化

[0727] DON是原动词名词化

[0728] 创造思维最终目标以计算机的无与伦比的记忆、联结和极速来非常高效地模仿实现人的创造思维。

[0729] 本发明将自然语言似谓词演算形式转换后可以实现机器自动学习:

[0730] 人类知识由各种具体景象中获得,并通过发现它们的相似性而归纳出类属,即一些具有相似性的不同事物从某种角度看,被认定为一些相同的东西。按照本发明方法,通过词(概念)的格式塔结构维度的相似度,机器可以自动搜索识别,找出同类属的概念(词)和相反的概念(词),表述行为的系列词汇也可以同样的方法寻出,并作为学习结果储存入字词数据库。

[0731] 本发明可以实现机器通过对某事实表述的学习,从中找寻因果关系并推理得出某个结论:

[0732] 进行这类推理,需要有知识库提供不同的但是意义相同的概念(词)和因果关系句子最简似谓词演算形式联结。

[0733] 本发明的自然语言似谓词演算形式转换后可以实现机器自动编程:

[0734] 11.自动编程

[0735] 11.1 程序“自变量为a时,执行p,q这种谓词,自变量为b时,执行r,s这种谓词”。

[0736] pred(a):-p,q

[0737] pred(b):-r,s

[0738] Pred(abbr.断言;宣布;谓语(predicate))

[0739] first(初始)

[0740] rest(剩余部分)

[0741] cons(construction的缩写,建立,解释)

[0742] 初始(first)元素为a,剩余的元素(rest)为[b,c],这时的表可以表示为[a|[b,c]]。若将这种表示进行复原,即从表中取出初始元素first,和初始元素以外的元素rest,或者构成以谓词描述的cons,则会变成下列形式:

[0743] first first([X|Y],X)

[0744] rest rest([X|Y],Y)

[0745] cons cons(X,Y,[X|Y])

[0746] 11.2 需要编制的程序

[0747] 11.2.1 自动生成词库和知识库

[0748] 设从辞典中抽取一词“读——看着文字念出声来或学习”，把这个词及其释义转换为以Prolog建立的词库中的词。

[0749] 【实例】从已有辞典得到“读,看着文字念出声来或学习”,先转成英文:

[0750] Read is Look at words and sound or learn.

[0751] 问题1:能否直接联结电子辞典而将汉字对应转换为英文词?如果不行,则建立英汉对应的

[0752] 字库(是字库而不是辞库,如“读——read”对应)。

[0753] 2.先人工将read的定义转成似谓词演算形式:

[0754] Read is Look at words and sound

[0755]  $=(\text{read, is, (X1, Look at, words)} \wedge (\text{X1, sound}) \vee (\text{X1, learn}))$

[0756]  $=(\text{read, is, W1} \wedge \text{W2} \vee \text{W3})$

[0757]  $=(\text{read, is, W1}) \wedge (\text{read, is, W2}) \vee (\text{read, is, W3})$

[0758]  $W1 = (\text{X1, Look at, words})$

[0759]  $W2 = (\text{X1, sound})$

[0760]  $W3 = (\text{X1, learn})$

[0761] 3.进而机器自动直接转换为似谓词演算形式并编程:

[0762] 似谓词演算形式均是

[0763]  $(N1, DO, N2),$

[0764] 或  $(N1, IS, N2),$

[0765] 或  $(N, IS, WHERE),$

[0766] 或  $(N, IS, STATE),$

[0767] 最后分解后的似谓词演算形式均是最简式,即最多只有三个单词,且第二个词是谓词。一个语句可能分解为数个“与”关系的最简式,并由它们构成完整句义。

[0768] 按最简式一组三词的次序转换为谓词形式并编程,完成一组后,再进行另一组的相同操作,如此到完毕。

[0769]  $\text{significance}(\text{read}): -W1, W2$

[0770]  $\text{significance}(\text{read}): -W3$

[0771]  $W1: -\text{Look\_at}(X, \text{words}), \text{person}(X)$

[0772]  $W2: -\text{sound}(X), \text{person}(X)$

[0773]  $W3: -\text{learn}(X), \text{person}(X)$

[0774]  $\text{person}(X)$

[0775] 如此就是“读——read”在我们自建的词库中的形态,所有的最简式都有自己的编号Wj,以避免在定义其他词时出现重复。其他词均如此建立。

[0776] 以上对本发明实施例所提供的自然语言机器识别方法及系统进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发

明的限制。

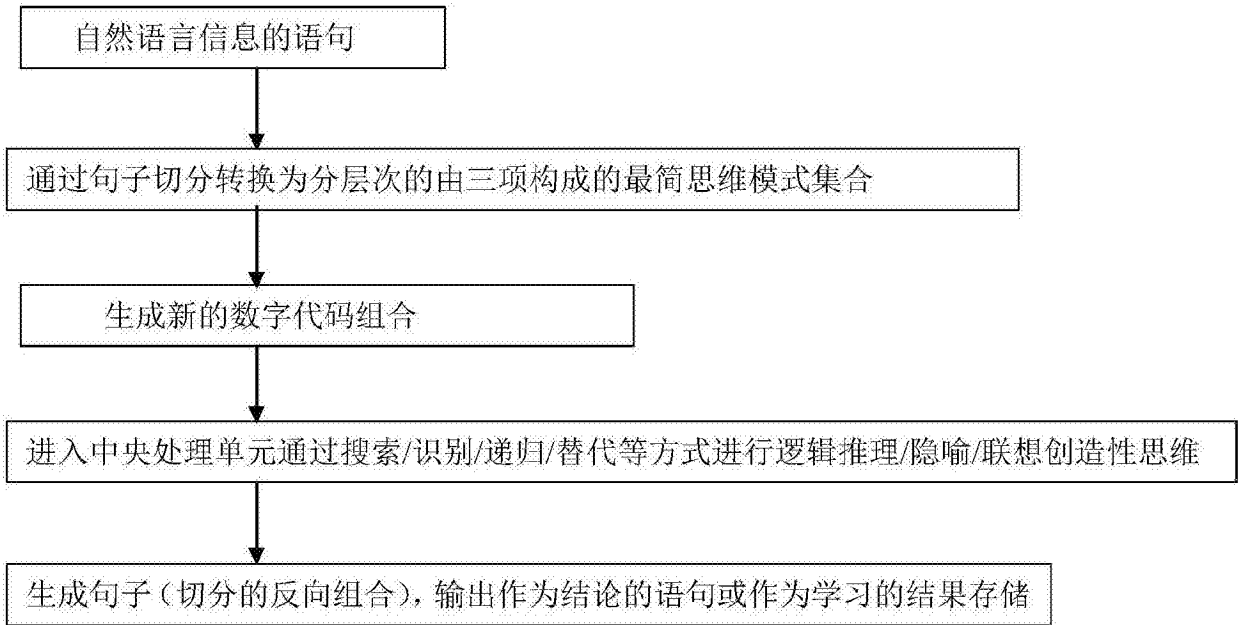


图1

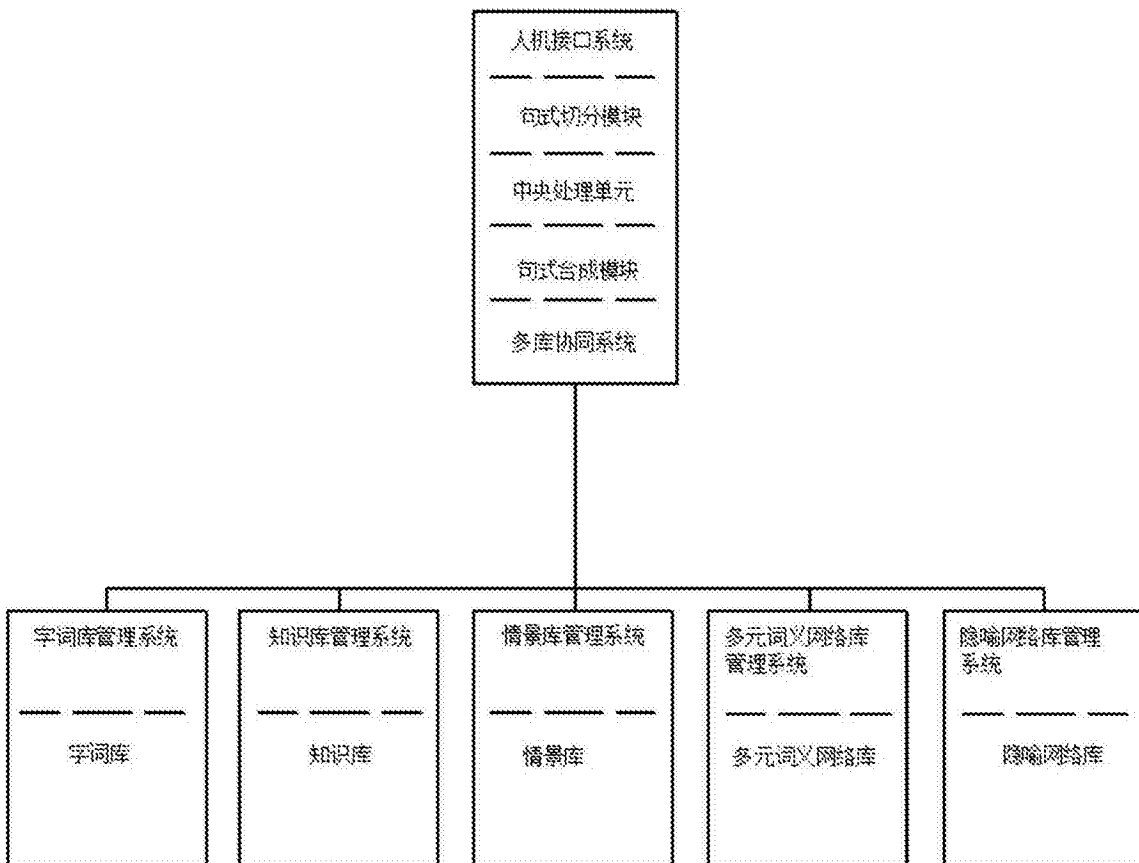


图2