

# 动词属性在自然语言处理当中的研究与应用

杨欢,许威,赵克,陈余

(西安电子科技大学 机电科学与技术研究所,陕西 西安 710071)

**摘要:**在语言学当中,动词跟受其支配的语义成分可以构成一个最基本的语义结构,动词与这些成分之间蕴含着语义以及语法关系。结合这一研究成果,提出动词属性的概念,分析其语义特征,并将其运用于自然语言处理模型的具体实践中,通过对以自然语言形式表达的用户需求中动词属性进行理解和分析,结合自然语言理解系统“动的”结构的运行,将最终分析结果转化成概念设计要求,较好地解决了动词属性结构与“动的”结构存在耦合的现象。

**关键词:**自然语言处理;动词属性;“动的”结构

**中图分类号:**TP182;H14

**文献标识码:**A

**文章编号:**1673-629X(2008)07-0233-04

## Research and Application of Verb Attributes in Natural Language Processing

YANG Huan, XU Wei, ZHAO Ke, CHEN Yu

(Inst. of Electromechanical Engineering, Xidian University, Xi'an 710071, China)

**Abstract:** In linguistics, the verb with its dominance semantic elements could constitute a basic semantic structure, verbs with these components contain grammatical and semantic relations. The researchers proposed the concept of the verb attributes, and analysed its semantic features, and their application in model of natural language processing of the specific practice among them. By understanding and analyzing the verb attributes in user requirements which is expressed in natural language style, and conversing them into the requirements of conceptual design with the whole system running. It preferably solves the problem of coupling between structure of verb attributes and Chinese structure “V+de”.

**Key words:** natural language processing; verb attributes; Chinese structure “V+de”

## 0 引言

动词和动词性结构是现代汉语语法研究的热门话题,也是语法研究的关键之一。有人说动词研究好了,语法问题就解决了一半<sup>[1]</sup>。古往今来,有很多学者对此进行了细致的研究。随着对动词及其相关认识的提高,人们发现动词的名词化处理在自然语言处理领域具有重要意义,例如“动的”结构以及下文将要讨论的动词属性结构都涉及到这一过程,它也成为当今自然语言处理领域研究的重要课题。另外,研究的深入使得自然语言处理模型对于动词的处理已不能仅仅围绕着施事和受事两方面展开,而是应该赋予它更为丰富的内容。笔者总结归纳了前人的研究成果,提出了“动词属性”的概念,并将其运用到自然语言处理模型的具

体实践当中,帮助解决有关动词与其相关成分的语义关系的处理问题。

## 1 动词属性及其分类

### 1.1 什么是动词属性

动词语义问题在当代语言研究中越来越引起人们的重视,研究也从传统语义学中以词义为主扩展到句子和篇章的层面。其实关于动词与其相关成分的语义关系的问题早已有学者问津。在汉语语法研究的早期吕叔湘先生就曾经提到<sup>[2]</sup>：“从语义方面看,名词和动词之间,也就是事物和动作之间,可以有多种多样的关系,决不限于施事和受事。”

人们知道,动词跟受其支配的语义成分可以构成一个最小的语义结构,如施事、受事、共事等关系,在语言学上,把这种语义结构称为动核结构。动核成分的性质决定了它可以跟多少或何种动元成分相搭配,制约了受其支配的语义成分的数量和性质。来看一个例子:我坐过磁悬浮列车三遍了。这句话当中动词“坐”

收稿日期:2007-10-30

基金项目:科技部科技型中小企业创新基金(01c26226111002)

作者简介:杨欢(1983-),男,陕西宝鸡人,硕士研究生,研究方向为人工智能、知识工程、创新技术;赵克,博士,教授,研究方向为人工智能、创新设计和知识工程。

是语义核心,叫动核;动核强制性联系的施事对象“我”称作主事动元,受事对象“磁悬浮列车”称为客事动元;“三遍”属于非强制性语义成分,它是“我坐过磁悬浮列车”基干句模的扩展,在这里叫状态元或状元。可以看到,状元的使用很灵活,可以根据表达需要来为基干句模添加状元成分。比如可以加入坐车的原因,坐车的时间等等<sup>[3,4]</sup>。

由此可见施事对象与受事对象是每一个动词所必然具有的特性,但实际上动词所具备的特性还远不止这两项。比如时间、地点、方式、原因、目的、结果、条件、过程等都是动作发生时客观存在的事实。另外对于某一类特殊的动词在发生的时候还会伴随其产生一些特有的性质,例如有关运动类的动词就潜在包含着速度这一特质。把这些动词发生时伴随其产生的客观事实称作该动词的属性。它们是该动词动元与状元的集合。对动词属性的研究有助于更深刻地挖掘动词本身具有的特性,对于自然语言理解有着积极的意义。

## 1.2 动词属性的分类

分类是自然语言理解工作当中的重要内容。分类是否合理直接影响到计算机工作的成效<sup>[5]</sup>。动词属性是动词知识的组成部分,根据它本身的复杂性有必要对其进行分类。对动词属性的合理分类可以有效地处理有关动词的语义结构。下面就讨论一下动词属性的分类。

事实上目前关于动词以及动元、状元的分类并没有完全一致的意见<sup>[3]</sup>。不同的分类标准得到不同的分类结论,每一种分类方法都有它自己的道理,谁也说服不了谁。这里也是提出了一种分类标准,即按照动词属性的从属范围划分。

按照动词属性的从属范围可以把它分为一类动词共有的属性和单个动词具有的特有属性。

(1)公有属性是每个动词都包含的属性,它包括施事、受事、时间、地点、方式、原因、目的、结果、条件、过程等。这些属性普遍存在于每个动词知识当中,只是并不是每一个属性在动作发生时都会被具体描述,而仅仅是在需要时才会出现在语句当中。

\* 施事、受事是动词属性的必选项。二者是一个动作发生时必然涉及的要素,否则该动作没有执行和承受的载体。若不考虑省略现象,所有动作都有其执行者和承受者,实际语言运用中,由于省略现象的普遍存在,施事或受事常常可能省略,给人造成某些动词无施事或受事的印象,其实它们都是必选项,至少其中之一不可缺少,其中施事表示在执行该动作时的动作执行者,它从动作执行者的具体要求和数量方面来限制,受事表示在执行该动作时的动作承受者,它从动作承

受者的具体要求和数量方面来限制。

\* 时间、地点、方式、原因、目的、结果、条件和过程等是动词属性的可选项。这些属性在动作发生时都是存在的,但这些属性并不是陈述该动作的必然载体,只是使该动作描述更具体的相关要素。

(2)特有属性是一个或一类动词所具有的私有属性,如行程问题中,运动类动词如走、跑、出发、去、行驶、到达等动词都具有速度属性。特有属性是这个或这类动词与其它动词互相区别的关键属性。但是特有属性并不是动词发生时的必选项,它同除施事、受事之外的公有属性一样只是让动作的描述更加丰富具体。

动词属性分类示意图如图 1 所示。

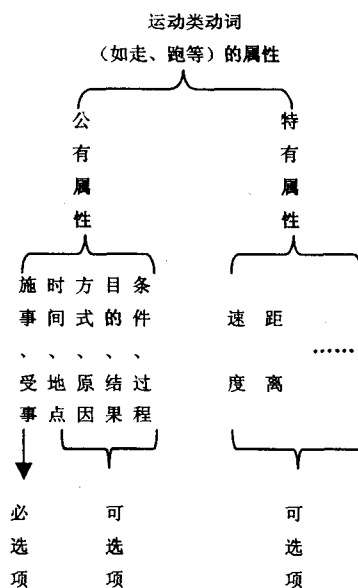


图 1 动词属性分类

## 2 动词属性处理的语义特征

### 2.1 动词的属性不参与本动词的语义搭配

通过以上对动词属性的分类,参与动词语义搭配的属性是本动词的施事和受事属性,除此之外的动词属性不会参与到本动词的语义搭配,从一定意义上讲它们是动词语义搭配成功后生成事件的相关描述。

虽然这些属性不参与到本动词的语义搭配中,一般情况下会参与到其它动词的语义搭配关系中,这应该是在一个动词的描述中着重描述该属性的目的所在。

例如:小明买烤鸭的地点是东大街。

运动类动词

(如走、跑等)的属性

本句中“小明”是“买”动词的施事属性,“烤鸭”是“买”的受事属性,“地点”是“买”的其它属性,可以看到“小明买烤鸭”已经构成了完整的事件,“地点”并没有

参与其语义搭配,而是参与到后面的语义搭配当中,构成了另一具体事件“地点在东大街”。

## 2.2 包含动词的属性的动词组合会产生多余的动元

由于动词的属性包含在一组动词组合中,而这些属性不参与到本动词的语义搭配,因此会造成产生多余的动元的情况。这些动元与该动词具有属性关系,并且应该充当该动词之后的动核结构的动元。

例如:中国队出线的希望大增。

本例中“中国队”是前一动核结构“中国队出线”的动元,而作为“出线”的属性的“希望”则作为后动核结构“希望大增”的动元。

## 2.3 动词属性与名词的关系

动词的属性与名词和名词的属性有密切关系。动词的属性表现为名词,名词的属性也表现为名词,有时二者可能产生重合。如在行程问题中,运动类动词如走、跑、出发、去、行驶、到达等动词都具有速度属性;同时在叙述时通常会说“人的速度”、“车的速度”等,也就是说人和车同样具有速度属性,这是因为人和车运动时所产生了速度,即人和车执行了运动的动作从而继承了运动的速度属性。

从这个角度来讲可以得出这样的结论:参与动作的名词可以继承动词的一些属性,从而也具备了这些属性特征。这样可以将名词的属性分为两类,即名词本身固有的属性和继承动词而得到的属性。对名词属性的分类同样很有必要,它对名词块的处理和名词的知识表示起着重要作用,如:

小明的考试成绩很高。

小明的消费金额为2000元。

小明的吃饭速度很快。

在这些例子中,“考试成绩”、“消费金额”、“吃饭速度”等并不是小明的真正属性,而是小明参与“考试”、“消费”、“吃饭”等动作所产生的属性,这些属性最好不要放入名词“人”的外延槽中,否则必然造成名词知识的庞大与冗余,而是应当作为动词的属性分别存储在动词的静态知识中,在“人”参与相关动作时,才继承给“人”。

## 3 动词属性和“动的”结构的耦合

### 3.1 “动的”结构

“的”字结构是一种比较特殊的结构。名词、形容词、动词三大词类几乎能平等地与助词“的”构成“的”字结构,但它们在“的”字结构主语句中的出现频率很不一样,“V的”最高,“A的”次之,“N的”很少。“的”字结构是现代汉语中最常用的一种短语,它使用的范围非常广泛,语义结构也非常复杂。

“动的”结构是“的”字结构的一类,作用相当于名词,在句中主要充当主语和宾语。这类短语很有特色,在现代汉语中十分常见<sup>[6]</sup>。将“动的”结构分为V+的+N、V+N+的+O、N+的+V三大类。

例如:三边相等的三角形是等边三角形。

人民生活水平的提高原因是多方面的。

### 3.2 动词属性结构和“动的”结构存在耦合

动词属性结构和“动的”结构某些时候在结构搭配上非常类似,仅从句法结构上判断很难区分它们。

例如在语言中,有一类非常常见的句子,它的结构为“名词1+动词+的+名词2+其它”,名词1是动词的施事,而名词2是动词的属性,如:中国队出线的希望大增。

再如:小明的消费地点在东大街。本句的结构为“名词1+的+动词+名词2+其它”。同样的,名词1是动词的施事,而名词2是动词的属性。

可以看出,这两句话分别与3.1节当中的两个“动的”结构的例句在搭配关系上如出一辙,但是它们不能算作“动的”结构,也不能按照“动的”结构的分析方法分析。二者的关键区别在于动词的属性不参与到本动词的语义搭配中,只充当该动词之后的动核结构的动元,是“单用”;而动的结构的多余核心词一般要参与到本动词的语义搭配中,也充当该动词之后的动核结构的动元,是“复用”。

## 4 动词属性与“动的”结构的处理与实现

### 4.1 动词属性与“动的”结构耦合部分句法结构划分

自然语言的灵活性与复杂性使得自然语言处理模型的建立非常困难。由于篇幅限制这里仅将较为常见的几类动词属性与“动的”结构存在耦合的句法结构总结列举出。

按照动词属性在句子中的作用,可以将存在耦合的结构可分为以下几类:

#### (1) 动词属性类型1。

结构特征如“名词+动1+名词+的+名词+动2”。这种结构前名词块的名词的个数大于0,后名词块名词的个数大于1。动词组合结构的动词有主语,动词组合结构后名词块的前一个名词做“动的”的宾语,动词组合结构后名词块的后一个名词做下一个动词组合的主语。

例如:小明买烤鸭的地点是东大街。

#### (2) 动词属性类型2。

结构特征如“名词+动1+的+单个名词+动2”。这种结构前名词块的名词的个数大于0,后名词块的名词的个数等于1。动词组合结构的动词有主语,动

词组合结构后唯一的一个名词做下一个动词组合的主语。

例如:小明消费的地点是东大街。

### (3) 动词属性类型 3。

结构特征如“名词 + 的 + 动 1 + 名词 + 动 2”。这种结构前名词块的名词的个数大于 0, 后名词块的名词的个数大于 0。动词组合结构的动词有主语, 动词组合结构后唯一的一个名词做下一个动词组合的主语。

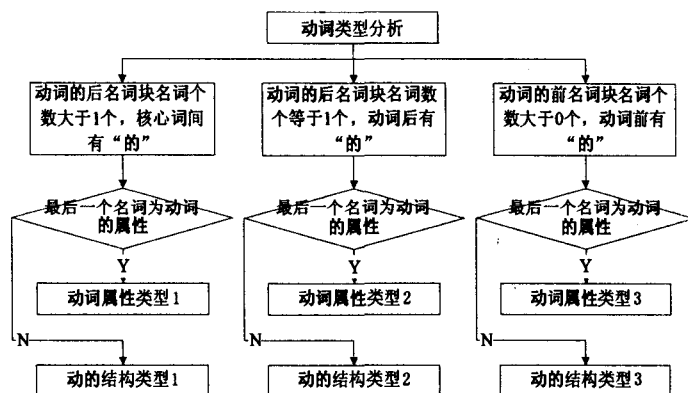
例如:小明的消费地点是东大街。

## 4.2 具体处理办法

实际上解决动词属性与“动的”结构耦合现象的关键在于分辨核心词是否是动词的属性。若是则进入到对应的动词属性处理模块, 不是再进入到相应的“动的”结构处理模块。处理的总体流程图如图 2 所示。

## 5 结 语

语言学的研究发现了动词意义中丰富的语义、句法属性。结合这方面的研究成果, 提出了动词属性的概念, 并对动词属性的分类和语义特征进行讨论, 将其运用到自然语言处理模型的具体实践当中, 针对动词属性结构与“动的”结构存在耦合的现象提出了一种解决办法。当然, 此解决方案还不能涵盖所有的耦合情况, 中间的某些情况可能仍需要细分, 并在以后调试过程中逐渐优化。



“动的”结构、动词属性的分类流程

图 2 总体流程

## 参考文献:

- [1] 马庆株. 汉语动词和动词性结构(一编)[M]. 北京: 北京大学出版社, 2005.
- [2] 吕叔湘. 汉语语法分析问题[C]//汉语语法论文集. 北京: 商务印书馆, 1979.
- [3] 冯丽萍. 动词的语义特征及其在句子加工中的心理现实性[J]. 语言文字应用, 2006, 12: 172-176.
- [4] 陈昌来. 现代汉语语义平面问题研究[M]. 上海: 学林出版社, 2003.
- [5] 石纯一, 黄昌宁, 王家钦. 人工智能原理[M]. 北京: 清华大学出版社, 1993.
- [6] 裘荣荣. “动 + 的”短语的表意功能[J]. 修辞学习, 1999 (1): 43-44.

(上接第 232 页)

现有方案中资源只能被添加入系统而未提供删除方案; 方案不支持模糊关键字搜索等。

## 参考文献:

- [1] Yiu W P K, Jin Xing, Gary H S C. Chan. Distributed Storage to Support User Interactivity in Peer-to-Peer Video Streaming[C]//ICC'2006. Istanbul: [s. n.], 2006: 55-60.
- [2] 阳天保, 张修如, 贾丽会. 基于 P2P 视频点播系统模型及算法研究[J]. 计算机技术与发展, 2006, 16(10): 45-48.
- [3] Castro M, Druschel P, Kermarrec A, et al. Splitsream: High-bandwidth multicast in a cooperative environment[C]//ACM SOSP. New York, US: ACM, 2003.
- [4] Tran D, Hua K, Do T. ZIGZAG: An Efficient Peer-to-Peer Scheme for Media Streaming[C]//IEEE Infocom'2003. San Francisco, California, USA: IEEE, 2003.
- [5] Wan Kan Hung, Loeser C. An overlay network for replica placement within a P2P VoD network[J]. Journal of High Performance Computing and Networking, 2004(7): 1-10.
- [6] Stoica I, Morris R, Karger D, et al. Chord: A Scalable Peer to peer Lookup Service for Internet Applications[C]//SIGCOMM'2001. San Diego, California, US: ACM, 2001.
- [7] Ratnasamy S, Francis P, Handley M, et al. A Content Addressable Network[C]//SIGCOMM'2001. San Diego, California, US: ACM, 2001.
- [8] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems[C]//Proceedings of IFIP/ACM Middleware 2001. Heidelberg, Germany: ACM, 2001.
- [9] Rowstron A, Druschel P. Storage Management and Caching in PAST, a Large-Scale, Persistent Peer-to-Peer Storage Utility[C]//Proc. of ACM Symposium on Operating Systems Principles (SOSP). Banff, Alberta, Canada: ACM, 2001.
- [10] Ng E. GNP software[EB/OL]. 2003. <http://www-2.cs.cmu.edu/eugeneng/research/gnp/software.html>.
- [11] Dabek F, Cox R, Kaashoek F, et al. Vivaldi: A Decentralized Network Coordinate System[C]//SIGCOMM'04. Portland, Oregon, US: [s. n.], 2004.
- [12] Cohen B. Incentives build robustness in Bit Torrent[J]. Economics of Peer-to-Peer Systems, 2003(6): 55-61.