文章编号：1003-0077(2020)10-0039-12

基于事件属性的事件分类研究

王 亚[1,2]，曹存根[1]

(1. 中国科学院 计算技术研究所 智能信息处理重点实验室，北京 100190；2. 中国科学院大学，北京 100049)

摘 要：事件分类研究一直是计算机科学和语言学等学科的核心研究内容，针对动词语义层面上的分类问题，研究者们提出了不同的分类标准，而根据这些分类标准对动词进行分类会产生分类有交叉和分类粒度粗等问题。一个动词通常表示一个过程事件，该文以汉语世界中经常发生的过程事件为语义分类对象，从事件的定义中提取事件的特征属性，并给每个特征属性赋予权重，利用特征属性对顶层事件类包含的事件进行分类。该文采用框架的形式对事件进行语义描述，框架内容由事件的特征属性和私有属性组成。重点以“传播”类事件为例来阐述该文的分类方法，通过实际操作发现，利用该分类方法，可以得到一个比较清晰的事件语义分类结构。该文用描述逻辑来对事件及事件之间的分类关系进行形式化表示。根据该事件分类体系，可以有效获取事件属性相关的常识知识。

关键词：事件语义分类；特征属性；事件框架

中图分类号：TP391 **文献标识码**：A

Research on Categorization of Events Based on Event Attributes

WANG Ya[1,2], CAO Cungen[1]

(1. Key Laboratory of Intelligent Information Processing, Institute of Computer Technology, Chinese Academy of Sciences, Beijing 100190, China;
2.University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Aimed at the categorization of events with procedure, we propose an event semantic categorization method based on event attributes. We extract the characteristic attributes of an event from its definition and assign the weight to each characteristic attribute. We adopt the frame semantics to represent an event, which consists of characteristic attributes and private attributes. This paper utilizes the class of "dissemination events" as an example to demonstrate our categorization method. We prove that a clear semantic categorization structure of events can be obtained with this method. We use description logics to formalize the events and the relationships between these events. According to this event classification system, we can effectively acquire commonsense knowledge related to event attributes.

Keywords: semantic categorization of events; characteristic attribute; event frame

0 引言

认知即分类[1]，分类是人类认识客观世界的一种手段，分类问题是认知科学中最基本的问题。分类在计算机科学、语言学和心理学等学科的相关研究中均有广泛应用。目前，计算机技术越来越多地应用到语言研究中，而动词是构成语言的句子的中心，因此，为满足计算机处理语言信息的需要，词在语义层面上的分类特别是动词的语义分类变得尤为重要。在计算机科学的相关研究中，文本标注[2-3]、词义消歧[4]、机器翻译[5]、信息抽取[6]等任务都涉及到了动词的语义分类。

动词的语义分类是刻画动词语义特征的常用手段之一，而描述动词的语义特征一般涉及两个方面，一个是动词本身的语义特征，另一个是句法中的动

收稿日期：2019-10-11 **定稿日期**：2020-01-06

基金项目：国家重点研发计划(2017YFB1002300，2017YFC1700302)

词搭配成分的语义特征。因此,针对动词的语义分类,研究者们主要以动词本身的语义和动词的句法描述为分类依据,提出不同的分类角度。一些研究者以动词自身蕴含的语义为分类标准对动词进行分类[7-12];另一些研究者从以动词为中心的句法所包含的语义成分的角度对动词进行分类[13];还有一些研究者结合动词语义和动词的句法描述两个方面对动词进行分类[14-15]。其中,根据动词本身蕴含的语义进行分类是最常用的分类标准,但会伴随着分类、有交叉和分类标准不统一等问题。而利用句法描述进行分类,不能很好地反映动词的语义本质。

鲁川根据谓词蕴含的语义特征将谓词分为二十六个语义类[16],其中动词有二十个语义类,分别是"交易"类、"变化"类、"对待"类、"传播"类、"思想"类、"类同"类、"进展"类、"遭受"类、"自移"类、"活动"类、"给予"类、"索要"类、"搬移"类、"创建"类、"改变"类、"促使"类、"感知"类、"支配"类、"探询"类和"寻求"类。但鲁川仅对动词做了上层类的划分,没有对上层动词类包含的动词进行再分类。本文在鲁川二十个动词语义类的基础上继续进行动词分类研究,进一步细分这二十个上层动词类中的动词,并针对以往动词语义分类中存在的问题提出了利用动词的"特征属性"对汉语动词进行语义分类的方法。动词的特征属性即动词所具有的区分一个动词与另一个动词的显著性质,在理解一个动词时,我们通常只关注一个动词的特征属性。动词的定义是对动词语义的描述,通常包含较为完整的动词特征属性信息,因此能很容易地从动词的定义中获取动词的特征属性。

Rothstein将动词视为由事件构成的集合或者事件的类型[17],本文研究的重点是事件类(即不明确交代事件参与者、事件发生的时间和事件发生的地点等具体信息的事件类型)及其分类。事件类包括状态事件类和过程事件类,过程事件类通常用动词表示,本文研究的主要是过程事件类。为了与接下来的章节中提到的事件类相区别,我们将"过程事件类"称为"过程事件",简称"事件"。

本文将鲁川的二十个上层动词类称作顶层事件类,即对所有事件进行分类得到的第一层分类结果,根据事件具有的"特征属性"对各个顶层类中的事件进行再分类。以"传播"类事件为例,本文从这类事件的定义中提取了信源、信宿、传播媒介、信息种类、传播目的、传播时间等共十五个传播类事件包含的特征属性。但是当一个事件的定义中蕴含多个特征属性时,我们只能选择其中的一个特征属性作为对该事件进行分类的依据,即选择一个优先级最高的特征属性来分类该事件。为此我们给每个特征属性添加权重,一个特征属性的权重即该特征属性在一个事件类中的所有定义中的出现比例,权重越大,被选择的优先级越高。我们利用特征属性对事件进行语义分类,分类标准明确,能获得较清晰的分类结构。据本文作者所知,利用事件的特征属性对事件进行分类是一项具有创新性的工作,我们目前还没有发现根据事件的属性对事件分类的相关研究工作。获取事件的特征属性后,根据框架语义学[18]的思想,使用框架的方式对事件的语义进行描述,框架内容包含事件的特征属性和私有属性,私有属性是各个顶层事件类中的事件所具有的特殊属性,对事件进行语义分类即对事件框架进行语义分类。

描述逻辑是一种用于知识表示的形式化语言[19],描述概念和概念之间的层次关系。描述逻辑具有严格的语义,对概念及概念分类都有很好的表示。因此,本文以描述逻辑(ALC)为基础,对事件及事件之间的关系等相关知识进行严格描述。

常识知识在自然语言处理[20]、计算机视觉[21]等领域都有着广泛应用,常识知识获取更成为制约人工智能发展的瓶颈问题。本文以事件属性为分类标准来对事件进行分类,因而在分类的过程中可以很自然地获取事件属性的相关常识知识,例如,蝉在夏天鸣叫(传播时间);人用嘴演唱歌曲(传播工具)等。在我们的事件分类结构中,子事件可以继承父事件的属性相关的常识知识,从而可以利用事件的语义分类实现常识知识的自动获取。我们还可以根据已构建好的事件分类体系检验已获取的常识知识是否完备,从而可以有效地避免常识知识遗漏的问题。

本文组织结构如下:第1节总结前人的语义分类研究工作;第2节对事件属性进行分类并介绍了特征属性的获取方法;第3节详细描述对事件进行语义分类的方法;第4节利用描述逻辑对事件及事件之间的关系进行表示;第5节介绍如何利用事件的语义分类获取事件属性相关的常识知识;最后总结本文的研究工作和遇到的问题。

1 相关研究

Levin认为动词的句法描述由其含义决定,动词不同的句法描述与其特定的语义相关联,如果两

个动词的句法描述包含相同的语义成分，则这两个动词属于同一类别[13]。一个动词通常具有多个语义成分，句法描述中包含的语义成分并不能充分反映动词的语义，而决定动词所属类别的通常是其强调的语义成分（即我们在本文中提出的事件的特征属性），而在 Levin 的分类系统中并没有体现这一特征。

WordNet[7]是一个在线的英语语义分类词典，其分类对象包括动词、名词、形容词和副词，被划分在同一个类别中的词组成一个同义词集合，一个同义词集合即代表一个词类，不同的词类位于分类系统中的不同位置，并以此形成一个分类结构。在 WordNet 中，动词被分成十五个顶层类，根据与这些顶层类之间的语义关系，不同的动词被划分到不同的同义词集合中。也就是说，WordNet 主要是根据语义关系对动词进行分类，这并不能清楚地描述动词自身的语义。

基于框架理论构建了 FrameNet 的研究者们认为，人们在很大程度上是通过词语所激活的框架来理解词语的含义[14]。通常情况下，激活一个框架的词语是动词，具有相似语义的动词会激活同一个框架。因而，FrameNet 根据动词所激活的框架来对动词进行分类，激活同一语义框架的动词被划分为同一个类别。FrameNet 要求一个动词类中的所有动词具有相同数量和类型的框架元素，即所有动词都具有相同的句法描述成分。由于汉语表达的多样性，具有相同语义的动词可能会有不同的句法描述，因而 FrameNet 的分类方法并不能很好地应用于汉语动词的分类中。此外，FrameNet 的分类比较粗糙，没有对同一框架中的动词进行细分，例如，其将“negotiate”（谈判）和“parley”（和谈）都分类到“discussion”这个框架中，但谈判与和谈（为恢复和平进行的谈判）具有上下位关系，谈判是和谈的上位动词，而 FrameNnet 并没有说明这种关系。

VerbNet[15]在 Levin 和 WordNet 动词分类的基础上，根据动词的语义和句法描述对动词进行分类，将具有相同语义和相同句法描述的动词归为一类。“具有相同的句法描述”这一分类依据与 FrameNet 类似，这里就不赘述了。VerbNet 还根据动词的意义对不同动词进行分类，与我们的分类方法相近，但动词意义包含的信息很多，VerbNet 并没有限制用于分类的意义范围，而我们限制了分类范围是从动词意义中提取的动词特征属性，分类标准更精确。另外，我们认为动词的不同句法描述刻画了动词的不同意义，用动词句法描述作为一个分类依据，其本质还是根据动词的意义进行分类。

HowNet[8]对大约 6 000 个汉字进行分析，进而从这些汉字的定义中抽取了一个有限的义原集合，并基于义原对词语进行分类，其他词语可以利用这些义原进行描述。与 HowNet 从动词的定义中提取动词义原的做法类似，我们从事件的定义中获取事件属性，并利用事件属性对事件进行分类，我们的事件属性与 HowNet 中的动词义原相似。但通过对事件进行考察我们发现，一个事件的定义往往会涉及其他事件，例如，答复的定义是“口头或书面回答别人的问题或要求”，答复具有“回答”这个义原，“回答”又具有“理睬”这个义原，因而就会产生一个事件用多个义原进行描述的问题，不利于对事件进行清晰的划分。

2　事件属性的内涵、分类与获取

2.1　事件属性的内涵与分类

事件属性是事件发生时伴随其产生的客观事实[22]，是事件具有的特殊性质。事件属性及其属性值是我们理解事件所需的基础知识，本文利用事件属性和其对应的属性值来描述一个事件。研究事件属性有助于我们挖掘事件本身的特性，对相关研究领域，特别是自然语言理解有着积极的意义[23]。根据属性的适用范围，可以将事件属性分为公有属性（public attributes）、私有属性（private attributes）和特征属性（characteristic attributes）。

定义 1（公有属性）　所有事件都具有的属性是公有属性。例如，事件的主体、事件发生的时间、事件发生的地点等。

定义 2（私有属性）　一个顶层事件类中的事件所具有的其他顶层事件类中的事件所不具有的属性是私有属性。例如，“传播”类的事件具有“信源”“信宿”等其他顶层类的事件所不具有的私有属性；“自移”类的事件具有“起点”“终点”等其他顶层类的事件所不具有的私有属性。

定义 3（特征属性）　一个事件所具有的并且在其定义中被明确描述的属性是特征属性。特征属性涉及事件的公有属性和私有属性，例如，“打招呼”在《现代汉语词典》中的定义是“用语言表示友好或礼貌”，因而“打招呼”具有“传播目的”（公有属性：表示友好或礼貌）和“信息载体”（私有属性：语言）两

个特征属性。

不同的事件具有不同的特征属性或者具有相同的特征属性但特征属性对应的属性值不同，因此，以事件的特征属性作为标准对事件进行分类，我们能区分所有不同义的事件，从而可以有效解决前人事件分类工作中存在的无法对某些事件进行分类或分类标准不统一的问题。

2.2 事件特征属性的获取

在一个顶层事件类中，特征属性用于区分两个事件，是对顶层事件类包含的事件进行分类的依据，获取事件的特征属性是我们对事件进行分类的一个前提。一个事件的定义是通过列出一个事件的基本属性来描述或规范一个动词的意义，包含完整的特征属性信息，因此，我们可以从事件的定义中获取事件的特征属性。此外，我们使用《现代汉语词典》来辅助获取各个事件的特征属性，《现代汉语词典》[24]是中国第一部规范性的语文词典，几乎涵盖了所有的汉语动词，对动词的定义也较全面，可以很好地满足查询事件定义的需求。

首先，我们查询《现代汉语词典》来获取事件的定义，例如，“打招呼”的定义是“用语言表示友好或礼貌”。通过该定义，我们可以获取“传播目的”（表示友好或礼貌）和“信息载体”（语言）两个特征属性。其次，如果一个事件的定义中出现了所属顶层事件类中的其他事件，则该事件继承其他事件的特征属性。例如，“道别”的定义是“分别时与人打招呼”，蕴含“传播时间”（分别时）这个特征属性，且定义中出现了“打招呼”这个事件，因而道别具有“传播目的”“信息载体”和“传播时间”三个特征属性，如图1所示。利用事件定义中出现的其他事件，我们还可以检验被分类的事件是否完备，如果定义中出现的其他事件不在我们已知的顶层事件类中，则将该事件添加到现有的分类体系中。

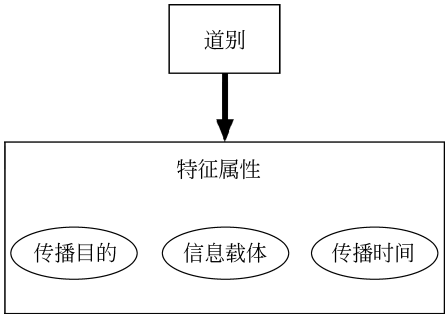

图1 “道别”的特征属性

表1给出了我们获取的“传播”类事件的十五个特征属性。

表1 “传播”类事件特征属性的定义

特征属性	含义	例子
信源	信息的发出者，包括信源种类、信源数量、信源情绪、信源态度等	欢唱、通报、搭讪
信宿	信息的接收者，包括信宿种类、信宿数量、信宿情绪、信宿态度等	报幕、汇报
传播媒介	信息传递的载体、工具或手段	演奏、致电
信息种类	信息的类别	宣判、发榜、开奖
信息量	信息的数量	沉默、细说
传播效果	传播行为在信宿身上引起的心理、态度或行为等的变化	热映、盗播
传播目的	实施传播行为的动机	警告、强调、和谈
传播范围	从同一信源发出的信息的信宿的数量、种类等	私议、颁布
传播次数	同一信源传播同一信息的次数	复议、首谈、重申
伴随动作	信源在传播信息的同时产生的动作	笑言、弹唱
传播距离	信源和信宿的距离远近	面谈、耳语
传播时间	实施传播行为的时间	夜谈、插话
传播地点	实施传播行为的地点	发帖、点映
传播速度	信源发出信息到信宿收到信息的时长	疯传
音量	信源传播信息时的声音大小	高呼、吼叫

除了具有特征属性外，事件还蕴含其他非特征属性，即私有属性。特征属性和私有属性都是刻画一个事件的语义所必不可少的属性，只不过特征属性比私有属性更重要。需要指出的是，事件的特征属性有可能是其他事件的私有属性，事件的私有属性也有可能是其他事件的特征属性。我们使用框架的形式对事件进行描述，框架名称用事件名称表示，框架内容由事件的特征属性和私有属性组成，事件的特征属性和私有属性是框架的槽，对应的属性值是框架的槽值，我们在事件框架中并没有列举出事件蕴含的所有私有属性，因为有些私有属性对应的属性值是未知的，这样的私有属性对理解事件是没

有意义的，因而我们在事件框架中不对其进行描述，如图 2～图 4 中的“欢唱”“独唱”和“合唱”的框架表示所示。

event 欢唱:
{
- ● **信源情绪**: 高兴
- ● 信息种类: 音乐
- ● 信息载体: 声音
- ● 传播工具: 嘴、音律
- ■ 信源种类: 非哑巴
- ■ 信源数量: 大于或等于1
- ■ 音量: 不能低于某一分贝

}

图 2　“欢唱”的框架表示

event 合唱:
{
- ● **信源数量**: 大于1
- ● 信息种类: 音乐
- ● 信息载体: 声音
- ● 传播工具: 嘴、音律
- ■ 信源种类: 非哑巴
- ■ 音量: 不能低于某一分贝

}

图 3　“合唱”的框架表示

event 独唱:
{
- ● **信源数量**: 1
- ● 信息种类: 音乐
- ● 信息载体: 声音
- ● 传播工具: 嘴、音律
- ■ 信源种类: 非哑巴
- ■ 音量: 不能低于某一分贝

}

图 4　“独唱”的框架表示

其中，圆点槽是事件的特征属性，方形槽是事件的私有属性。我们用特征属性来区分不同的事件，如欢唱具有独唱没有的“信源情绪”属性，独唱具有欢唱没有的“信源数量”属性，因此我们认为欢唱和独唱是两个不同的事件；独唱和合唱具有相同的属性，但两个事件的“信源数量”属性对应的属性值不同，独唱对应的属性值是“1”，合唱对应的属性值是“大于 1”，因此我们认为合唱和独唱也是两个不同的事件。

从特征属性的角度看，我们将具有如下关系特点的两个事件称为父事件和子事件。

（1）两个事件具有相同的特征属性，但存在一个事件的属性值是另一个事件属性值的子类的情况。例如，“表达”和“表白”都具有“信息载体”和“信息种类”这两个特征属性，前者的信息种类值是思想感情，后者的信息种类值是爱意，爱意是一种思想感情，因此“表达”是父事件，“表白”是子事件。

（2）一个事件只比另一个事件多包含一个特征属性，两个事件的其他特征属性相同且属性值也相同。例如，“道别”具有的特征属性为：传播目的（表示友好或礼貌）、信息载体（语言）和传播时间（分别时）；“打招呼”具有的特征属性为：传播目的（表示友好或礼貌）和信息载体（语言），因此“打招呼”是父事件，“道别”是子事件。

父事件与子事件之间具有上下位关系，因而子事件可以直接继承父事件的特征属性和私有属性，利用子事件与父事件之间的上下位关系，我们由父事件的特征属性和私有属性可以自动获取子事件的特征属性和私有属性。

3　事件的划分

首先，我们利用顶层事件类中的事件所具有的特征属性，对顶层事件类中的事件进行分类；然后，根据子事件类的特征属性对子事件进行分类；最后，我们根据特征属性和相应的属性值来区分最底层事件子类中的各个事件，如图 5 所示。

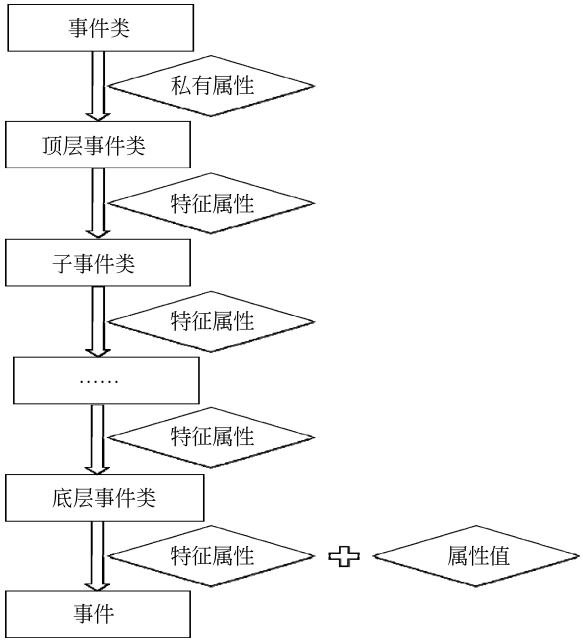

图 5　事件的分类依据

3.1　事件特征属性的权重计算

如上所述，从事件的定义中，通常可以获取多个特征属性，那么应该选择哪个特征属性作为对事件

进行分类的依据呢？通过对“传播”类事件进行详细的考察可以发现，一些特征属性在整个传播类中出现的比例明显大于其他特征属性，我们认为这些特征属性更能代表传播类事件，在对事件进行分类时，会优先于其他特征属性被选中作为分类标准。

但对于从特征属性的角度对事件进行分类来说，在将特征属性出现的比重作为其权重时，有两种计算权重的方法：第一种方法是，计算一个特征属性在顶层事件类中出现的比重，即特征属性在顶层事件类中出现的次数除以该顶层事件类中包含的事件数目，且特征属性的权重在整个分类过程中不会改变；第二种方法是，先计算特征属性在顶层事件类中出现的比重作为第一次分类的权重，然后在剩下的未分类事件中再重新计算各个特征属性出现的比重作为第二次分类的权重，特征属性的权重在整个分类过程中会发生改变。本文采用的是第一种权重计算方法，因为我们是从语义层面对事件进行分类，一个事件的语义是固定不变的，特征属性用于描述事件语义，特征属性的权重代表一个特征属性在事件语义中所占的比重，因此，特征属性的权重也应该是固定不变的。

我们设某一特征属性的权重为 w，该特征属性在一个顶层事件类中出现的次数为 f，这个顶层事件类中包含的事件数目为 n，如式(1)所示。

$$w=\frac{f}{n} \tag{1}$$

例如，传播类有 100 个事件，“传播目的”在 53 个事件的定义中出现了，则“传播目的”的权重是 0.53；而“信息载体”在 96 个事件的定义中出现了，则其权重是 0.96。

当“传播目的”和“信息载体”同时在一个传播类事件的定义中出现时，我们优先选择权重更大的“信息载体”作为对该事件进行分类的依据，把权重第二大的作为对事件进行再分类的标准，以此类推，直到事件的所有属性都已被使用，则结束对该事件的分类。例如，“打招呼”具有“传播目的”和“信息载体”两个特征属性，我们根据“信息载体”特征属性分类“打招呼”，因为“信息载体”的权重比“传播目的”的权重大。

表 2 给出了“传播”类各个特征属性的权重。

表 2 “传播”类具有的特征属性及其权重

特征属性		频数	权重	优先级
信源	信源种类	27	0.113	7
	信源数量	28	0.117	6
	信源情绪	16	0.067	12
	信源态度	23	0.096	9
信宿	信宿种类	26	0.108	8
	信宿数量	22	0.092	10
	信宿情绪	0	0.000	
	信宿态度	0	0.000	
传播媒介	信息载体	96	0.400	2
	传播工具	36	0.150	4
	传播手段	35	0.146	5
信息种类		164	0.683	1
信息量		8	0.033	16
传播效果		1	0.004	19
传播目的		50	0.208	3
传播范围		24	0.100	9
传播次数		11	0.046	15
伴随动作		6	0.025	17
传播距离		3	0.013	18
传播地点		13	0.054	14
传播速度		0	0.000	
音量		19	0.079	11
传播时间	时刻	14	0.063	13
	时长	1		

通过分析表 2 可以发现，“信源态度”和“传播范围”的权重是相同的，如果一个传播类事件同时具有这两个特征属性，那我们该优先选择哪一个呢？构成传播类事件的基本要素是信源、信宿、信息、媒介和反馈[25]，当属于基本要素的特征属性与不属于基本要素的特征属性的权重相同时，我们优先选择属于基本要素的特征属性，即优先选择“信源态度”这个特征属性；当属于基本要素或不属于基本要素的两个特征属性的权重相同时，就会产生分类有交叉的问题，此时，根据具体的应用来决定优先选择哪一个特征属性作为分类依据或随机选择一个特征属性作为分类依据。

3.2　事件的分类

获取了顶层事件类的所有特征属性并计算出每个特征属性的权重后，我们利用权重最大（即优先级最高）的特征属性对顶层事件类中的事件进行划分，得到顶层事件类的第一个子事件类，然后利用权重第二大的特征属性对剩余没有被划分到第一个子事件类中的事件进行分类，得到顶层事件类的第二个子事件类，以此类推，直到所有的事件都已被分类，此时我们完成了该顶层事件类第二层子事件类的划分。我们还可以继续对第二层子事件类中的事件进行分类，获取第三层子事件类，方法如第二层子事件类的划分。我们根据以下准则确定事件的分类粒度。

（1）一个事件类只有一个事件时，没有再对这个事件进行进一步分类的必要，即当某一子事件类只有一个事件时，不再对该事件进行细分。

（2）本文构建的事件分类体系要应用于事件属性相关的常识知识获取，具体内容在第 5 节中有详细介绍，因此，为了更好地进行常识知识获取，父事件和子事件必须属于同一类，即如果当前分类使父事件与子事件被划分到不同的事件类中，则停止对相应父事件和子事件的分类。

图 6 是对“传播类”事件进行分类得到的部分分类结果。首先利用“信息种类”等权重较大的特征属性对传播类事件进行第二层子事件类的划分，然后再对第二层子事件类包含的事件进行进一步的分类，例如，对以“信息载体”作为分类依据得到的第二层子事件类进行再分类，可以得到第三层子事件类的分类结果。

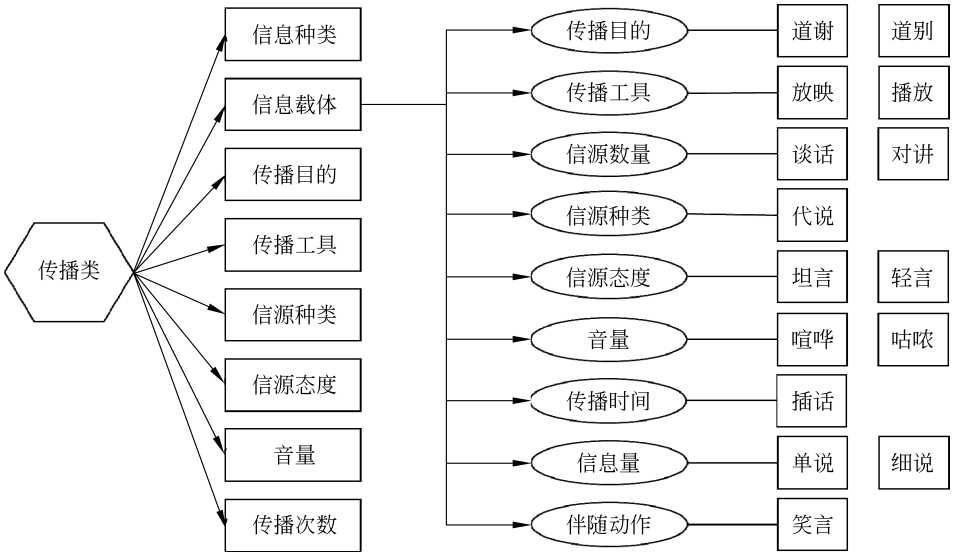

图 6　“传播”类事件的部分分类

目前为止，我们一共对二十个顶层事件类包含的 3 217 个事件进行了分类，这些事件都是从《现代汉语词典》中选取并根据动词的语义进行扩展得到的，基本涵盖了汉语世界中所有的过程事件，这些事件的具体分类层次如表 3 所示。需要指出的是，由于汉语动词自身的复杂性和表达方式的多样性，很多事件还具有同义事件，例如，“打酱油”中的“打”事件是“购买”事件的同义事件；“打毛衣”中的“打”事件是“编织”事件的同义事件；“打论文”中的“打”事件是“打印”事件的同义事件。因此，加上每个事件的同义事件，我们实际上一共对一万多个事件进行了分类。

表 3　事件的分类层次

顶层类	二层类	三层类	四层类	事件总计
传播类	10	112	118	240
对待类	100	14	132	246
改变类	11	129	66	206
搬移类	3	104	88	195
探询类	8	43	109	160
类同类	3	7	4	14
思想类	2	7	42	51
寻求类	2	21	11	34

续表

顶层类	二层类	三层类	四层类	事件总计
自移类	8	205	52	265
进展类	4	30	81	115
索要类	3	42	32	77
给予类	13	41	52	106
变化类	14	146	13	173
活动类	2	19	194	215
交易类	8	24	79	111
遭受类	4	18	176	198
感知类	2	8	54	64
创建类	12	145	16	173
促使类	13	107	46	166
支配类	15	296	97	408

4 基于描述逻辑的事件表示

4.1 描述逻辑

描述逻辑的基本构造元素是原子概念(atomic concept)和原子关系(atomic role),概念和关系可以利用构造子(constructor)在原子概念和原子关系的基础上进行描述。概念是个体的集合,关系表示个体之间的二元关系。ALC是最基本的描述逻辑形式语言,功能与语法较为简单,但足以满足我们对事件进行表示的需求。下面简要介绍ALC语法和知识库相关的概念。本文涉及的主要是概念,概念主要有以下语法形式:

$$C::=A\mid T\mid\bot\mid\neg C\mid C\cap\neg D\mid C\cup D\mid\forall R\cdot C\mid\exists R\cdot C\mid\geqslant nR\mid\leqslant nR$$

其中,A 是原子概念,C 和 D 表示概念,R 表示关系;T 是全概念,是其他概念的父概念;$\bot$ 是底层概念,是其他概念的子概念;$\neg$ 表示否定,$\cap$ 表示合取(conjunction),$\cup$ 表示析取(disjunction),$\exists$ 表示存在限制,$\forall$ 表示任意限制。例如,Person 和 Female 是原子概念,hasChild 是原子关系,Person $\cap$ Female 表示女性,Person $\cap\forall$ hasChild · Female 表示所有孩子均为女性的人,Person $\cap\geqslant$ 2hasChild 表示至少有两个孩子的人。

一个描述逻辑的知识库 K 包含两部分:TBox 和 ABox。TBox 是术语公理(terminology axiom)的集合,用来表示领域中的概念以及概念之间的关系,具有两种知识表示的形式,即 $C\subseteq D$ 和 $C\equiv D$,$C\subseteq D$ 表示概念 C 包含于概念 D,$C\equiv D$ 表示概念 C 等价于概念 D。ABox 是断言公理(assertion axiom)的集合,有两类断言,即 $C(a)$ 和 $R(b,c)$,a、b 和 c 是个体名,前者是概念断言,表示个体 a 属于概念 C,例如,Person(Tom)表示 Tom 是一个人;后者是关系断言,表示个体 b 和个体 c 具有关系 R,例如,hasChild(Mike, Tom)表示 Mike 有个孩子 Tom。

4.2 基于描述逻辑的“传播”类事件表示

事件包含一组属性,不同事件具有的特征属性不同或者具有的特征属性相同但特征属性对应的属性值不同。我们以“传播类”事件为例来介绍如何利用描述逻辑对事件进行表示,传播类定义为概念:

$$\text{DisseminationEvent}\equiv\text{Event}\cap\text{Event}_1\cap\text{Event}_2\cap\text{Event}_3\cap\text{Event}_4\cap\text{Event}_5\cap\text{Event}_6\cap\text{Event}_7\cap\text{Event}_8\cap\text{Event}_9\cap\text{Event}_{10}\cap\text{Event}_{11}\cap\text{Event}_{12}\cap\text{Event}_{13}\cap\text{Event}_{14}\cap\text{Event}_{15}$$

,其中:

$$\text{Event}_1\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{InformationSource}$$

$$\text{Event}_2\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{InformationHome}$$

$$\text{Event}_3\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{DisseminationVolume}$$

$$\text{Event}_4\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{InformationType}$$

$$\text{Event}_5\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{InformationSourceAmount}$$

$$\text{Event}_6\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{AccompanyingAction}$$

$$\text{Event}_7\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{DisseminationPurpose}$$

$$\text{Event}_8\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{DisseminationFrequency}$$

$$\text{Event}_9\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{DisseminationDistance}$$

$$\text{Event}_{10}\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{DisseminationTime}$$

$$\text{Event}_{11}\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{DisseminationArea}$$

$$\text{Event}_{12}\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{DisseminationScope}$$

$$\text{Event}_{13}\equiv\text{Event}\cap\exists\,\text{hasAttribute}\cdot\text{Dissemina-}$$

tionSpeed

$Event_{14} \equiv Event \cap \exists hasAttribute \cdot DisseminationEffect$

$Event_{15} \equiv Event \cap \exists hasAttribute \cdot InformationChannel$

传播类包含的属性是传播类的私有属性。传播类事件也是概念，例如，“打招呼”≡DisseminationEvent∩ ∃hasDisseminationPurpose·(=“表示友好或礼貌”)∩ ∃hasInformationChannel·(=“语言”)，该形式化表示中的属性是事件所蕴含的特征属性。

传播类是事件类的一个子类，即 DisseminationEvent⊆Event；“打招呼”事件又是传播类的一个子类，即：“打招呼”⊆DisseminationEvent。事件与事件之间也具有包含关系，例如，“道别”⊆“打招呼”；事件与事件之间还具有等价关系，例如，“道别”≡“告别”，这种包含关系和等价关系可用于描述传播类事件的分类关系。

2.2 节中介绍了两种不同形式的事件之间的包含关系，分别用描述逻辑表示即：

(1) 两个事件具有相同的特征属性，但存在一个事件的属性值是另一个事件属性值的子类：

$E_1 = hasAttribute_1 \cdot (= value_1) \cap \cdots \cap hasAttribute_i \cdot (= value_i)$

$E_2 = hasAttribute_1 \cdot (= value_1) \cap \cdots \cap hasAttribute_i \cdot (= value_{i'})$

其中，$1 \leqslant i \leqslant 15$。如果 $value_i \subseteq value_{i'}$，我们称 $E_1 \subseteq E_2$。例如，“表白”⊆“表达”。

(2) 一个事件只比另一个事件多包含一个特征属性，两个事件具有的其他特征属性相同并且属性值也相同：

$E_1 = hasAttribute_1 \cdot (= value_1) \cap \cdots \cap hasAttribute_i \cdot (= value_i) \cap hasAttribute_{i+1} \cdot (= value_{i+1})$

$E_2 = hasAttribute_1 \cdot (= value_1) \cap \cdots \cap hasAttribute_i \cdot (= value_i)$

其中，$1 \leqslant i \leqslant 14$。我们称 $E_1 \subseteq E_2$。例如，“道别”⊆“打招呼”。

以上的描述逻辑表示中，我们设定两个事件中的 $hasAttribute_i$ 和 $value_i$ 表示的属性和属性值相同，E_1 和 E_2 表示两个不同的事件。

5　基于事件分类的常识知识获取

由于常识知识本身的隐含性(人所共知的常识不会明确地被表述)和人类的报告偏倚(不同寻常的事实被更多的描述)等，使得常识知识获取成为人工智能领域急需解决的重要问题。研究者们探索了各种方法获取常识知识[26-29]，但至今还没有发现尝试利用分类的方法进行常识知识获取的相关研究工作。通过实际操作我们发现，利用前人提出的动词分类方法都不能很好地进行常识知识获取，而本文构造的事件分类体系可以为常识知识获取提供以下两个支持：

(1) 我们的事件分类依据“特征属性”为获取事件属性相关的常识知识提供了一个角度；

(2) 事件之间的上下位关系可以实现事件属性相关的常识知识的自动获取。

5.1　根据事件属性的常识知识获取

根据事件属性对事件进行语义分类的过程也是获取事件属性相关的常识知识的过程。我们利用事件的特征属性来对事件进行语义分类的，那么在分类前，我们需要知道该事件具有哪些特征属性。例如，要对“道别”进行分类，我们首先要获取它的三个特征属性，即传播目的、信息载体和传播时间。因此，对于“道别”事件，我们可以获取其相应的属性相关的常识知识，即“道别的目的是表示友好或礼貌”、“主体用说话的方式道别”和“道别发生在两个主体即将分别时”。

因为我们总结了传播类事件的所有特征属性，除了可以获取事件本身蕴含的特征属性相关的常识知识外，还可以将其他特征属性(私有属性)作为提示信息来获取更多的常识知识，例如，从“信息种类”的角度进行提示，我们可以获取“道别的内容是将要离开的话”；从“传播工具”的角度进行提示，我们可以获取“主体用嘴说话道别”，如图 7 所示。

5.2　根据事件分类关系的常识知识获取

在本文的事件分类结构中，从事件的特征属性角度来对事件进行划分的，因此，同一类事件具有一个或多个相同特征属性相关的常识知识，例如，“道别”和“道谢”都是根据“信息载体”这个特征属性进行分类而划分到同一类的，都具有信息载体相关的常识知识。如果“道别”或“道谢”没有获取信息载体相关的常识知识，则说明产生了常识知识遗漏，在常识知识获取的过程中，可以利用事件的语义分类进行常识知识遗漏的检验，从而获取更多的常识知识。

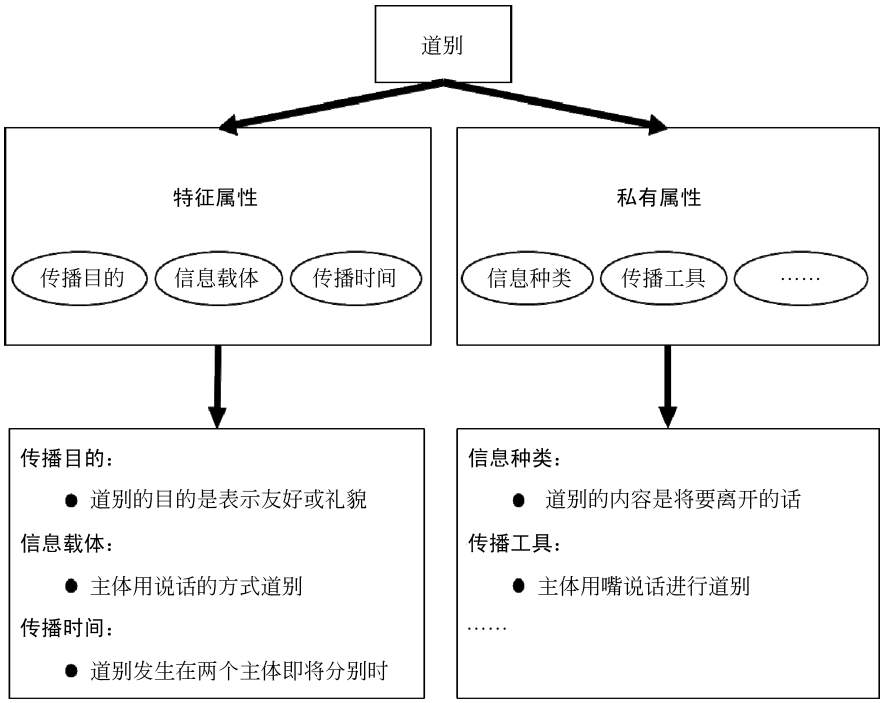

图 7 “道别”的常识知识获取

2.2 节中介绍了父事件与子事件的概念，父事件与子事件之间具有上下位关系，子事件可以直接继承父事件的特征属性和私有属性，因此，子事件也可以直接继承父事件属性相关的常识知识。利用子事件与父事件之间的上下位关系，我们由父事件的属性相关的常识知识可以自动获取子事件的属性相关的常识知识。例如，“欢唱”“高唱”“独唱”“唱歌”和“重唱”是“演唱”的子事件，如果已经获取了“演唱”的属性相关的常识知识，那么我们就可以自动获取其子事件的属性相关的常识知识，如图 8 所示，图中的垂直四点表示的是子事件直接从父事件继承的常识知识。需要说明的是，子事件可以根据自身需要修改从父事件继承的常识知识，如图中所示，“高唱”“独唱”和“唱歌”分别修改了其父事件中“音量”“信源数量”和“信息种类”方面的常识知识，这样做可以使子事件获取到的常识知识更准确；此外，除了从父事件继承的常识知识外，子事件还可能有自己特有的常识知识，如图中所示，“欢唱”和“重唱”分别从“信源情绪”和“传播次数”角度获取了父事件中没有的常识知识。

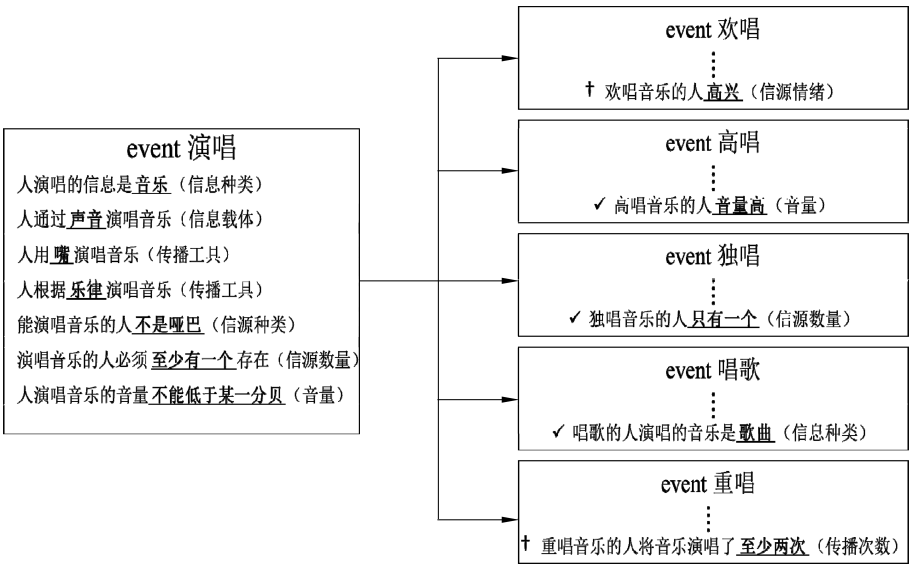

图 8 “演唱”的子事件的常识知识获取

6　总结

分类是人类认识客观世界的一种常用方法，本文以事件属性为分类依据对日常生活中发生的中文事件进行分类，以事件的特征属性为分类角度对顶层事件类包含的事件进行分类。我们利用《现代汉语词典》从事件的定义中获取事件的特征属性，当一个事件的定义中蕴含多个特征属性时，选取权重最大的特征属性作为对该事件分类的标准，权重次之的特征属性可以作为对该事件进行再分类的依据。一个特征属性的权重为该特征属性在一个顶层事件类包含的所有事件中出现的比例。通过对“传播”类事件进行全面的考察发现，利用事件的特征属性可以对事件进行很好的分类，不存在分类标准难以把握和分类有交叉等前人分类工作中存在的问题。为了形式化表述事件及事件之间的关系，我们利用描述逻辑对事件以及事件之间的关系进行了表示。此外，我们还将本文构建的事件分类系统应用到常识知识获取中，为常识知识的获取提供了一个新的获取途径。

但是，在分类的过程中，我们也发现了一些问题，例如，现实世界中发生的事件种类繁多，我们很难获取所有的事件，当有新的事件添加到我们的分类体系中时，一些特征属性的权重和优先级有可能会发生变化，相应的分类结果也会改变。另外，当从定义中提取事件的特征属性时，有可能会遗漏某些特征属性，这也会影响对事件进行分类的结果，这些都是我们后续要重点研究和完善的工作。

参考文献

[1] Cohen H. Handbook of Categorization in Cognitive Science[M]. Amsterdam: Elsevier, 2005.

[2] Li Ji Hong, Wang R B, Wang W L, et al. Automatic Labeling of Semantic Roles on Chinese FrameNet[J]. Journal of Software, 2010, 21(4):597-611.

[3] Park S B, Yoo E, Kim H, et al. Automatic Emotion Annotation of Movie Dialogue Using WordNet[M].Intelligent Information and Database Systems. New York: Springer, 2011.

[4] Wang C Y. Knowledge-based sense pruning using the hownet: an alternative to word sense disambiguation [D]. M. Phil.Thesis, Hong Kong: Hong Kong University of Science and Technology, 2002.

[5] 詹卫东，刘群. 词的语义分类在汉英机器翻译中所起的作用以及难以处理的问题[C]. 全国计算机语言学联合学术会议. 1997.

[6] Bai R, Wang X, Liao J. Extract semantic information from wordnet to improve text classification performance[M]. Advances in Computer Science and Information Technology.New York: Springer,2010: 409-420.

[7] Fellbaum C, Miller G. WordNet: An Electronic Lexical Database[M]. San Mateo:MIT Press, 1998.

[8] Dong Z, Qiang D. HowNet-A Hybrid Language and Knowledge Resource[C]//Proceedings of the 2003 International Conference on Natural Language Processing & Knowledge Engineering. NJ: IEEE, 2003: 820-824.

[9] 马洪海. 汉语框架语义研究[M]. 北京：中国社会科学出版社，2010.

[10] 刘海琴. 现代汉语位移动词研究[D]. 上海：复旦大学硕士学位论文，2011.

[11] 冯璐璐. 现代汉语认知心理动词句研究[D].南京：南京师范大学硕士学位论文 2015.

[12] 吴剑锋. 言语行为与汉语句类研究[M]. 上海：上海交通大学出版社，2015.

[13] Levin B. English verb classes and alternations: A preliminary investigation [M]. Chicago: University of Chicago Press, 1993.

[14] Lowe J B. The Berkeley FrameNet Project[J]. Proceedings of the COLING ACL,1998, 47(4):86-90.

[15] Kipper K, Dang H T, Palmer M. Class-Based Construction of a Verb Lexicon [C]//Proceedings of the17th National Conference on Artificial Intelligence & 12th Conference on Innovative Applications of Artificial Intelligence. 2000:691-696.

[16] 鲁川. 知识工程语言学[M]. 北京：清华大学出版社，2010.

[17] Rothstein S. Structuring events: A study in the semantics of lexical aspect[M]. Hoboken: John Wiley & Sons, 2008.

[18] Fillmore C J. Frame Semantics[M].Encyclopedia of Language & Linguistics, 2006:613-620.

[19] Baader, Franz. The description logic handbook: Theory, implementation and applications[M]. Cambridge University Press, 2003.

[20] Chen J, Chen J, Yu Z. Incorporating structured commonsense knowledge in story completion[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 6244-6251.

[21] Zellers R, Bisk Y, Farhadi A, et al. From recognition to cognition: Visual commonsense reasoning [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 6720-6731.

[22] 马庆株. 汉语动词和动词性结构[M]. 北京：北京大学出版社，2005.

[23] 杨欢，许威，赵克. 动词属性在自然语言处理当中的

研究与应用[J]. 计算机技术与发展，2008，18(7)：233-236.

[24] 中国社会科学院语言研究所词典室. 现代汉语词典，第 6 版[M]. 北京：商务印书馆，2012.

[25] 郭庆光. 传播学教程(第 2 版)[M]. 北京：中国人民大学出版社，2011.

[26] Speer R，Havasi C，Lieberman H. Analogy Space：Reducing the Dimensionality of Common Sense Knowledge[C]//Proceedings of the 23rd AAAI Conference on Artificial Intelligence. 2008.

[27] Tandon N，De Melo G，Weikum G. Acquiring comparative commonsense knowledge from the web[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. 2014.

[28] Boia M，Musat C C，Faltings B. Acquiring commonsense knowledge for sentiment analysis through human computation[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. 2014.

[29] Cambria E，Nguyen T V，Cheng B，et al. GECKA3D：A 3D Game Engine for Commonsense Knowledge Acquisition[C]//Proceedings of the 29th International Flairs Conference. 2016.

王亚(1988—)，博士，工程师，主要研究领域为常识获取、人工智能。
E-mail：wangya@ict.ac.cn

曹存根(1964—)，博士，研究员，主要研究领域为大规模知识处理、人工智能。
E-mail：cgcao@ict.ac.cn

(上接第 38 页)

[7] Landauer T，Dumais S. A solution to Plato's problem：The latent semantic analysis theory of acquisition，induction and representation of knowledge[J]. Psychological Review，1997：104 (2)，211-240.

[8] Pantel P. Inducing ontological co-occurrence vectors [C]//Proceedings of the 43rd Conference of the Association for Computational Linguistics，ACL'05 2005：125-132. Morristown，NJ，USA：Association for Computational Linguistics.

[9] Harris Z. Distributional Structure[C]//Proceedings of the papers in structual and transformational lingatstics，1970：1775-794.

[10] 梅家驹，竺一鸣，高蕴奇，等.同义词词林[M].上海：上海辞书出版社，1983.

[11] 哈工大社会计算与信息检索研究中心. 同义词词林扩展版[EB /OL]. [2019-09-13]. http://www.datatang.com/data/42306/

[12] Bengio Y，Ducharme R，Vincent P，et al.A neural probabilistic language model[J].Journal of Machine Learning Research，2003，3 (6) ：1137-1155.

[13] Mikolov T，Chen K，Corrado G，et al. Efficient Estimation of Word Representations in Vector Space [J/OL].arXiv preprint arXiv：1301.3781v3，2013.

龙从军(1978—)，博士，副研究员，主要研究领域为藏语计算语言学。
E-mail：longcj@cass.org.cn

刘汇丹(1982—)，博士，副研究员，主要研究领域为多语言处理。
E-mail：huidan@iscas.ac.cn

周毛克(1995—)，硕士研究生，主要研究领域为藏语自然语言处理。
E-mail：zmk_cass@126.com