

基于语言学特征向量和词嵌入向量的汉语动词事件类型预测*

刘洪超¹, 黄居仁¹, 侯仁魁^{2,1}, 李洪政³

(1. 香港理工大学中文及双语学系, 香港; 2. 鲁东大学文学院, 烟台, 264001; 3. 北京师范大学中文信息处理研究所, 北京, 100875)

摘要: 本文主要介绍汉语动词事件类型的预测。事件类型是根据内部时间结构对汉语动词进行的重要分类, 包括状态、活动、变化(完结和达成)。对汉语动词事件类型进行预测从理论上能够对以往语言学研究提出的特征进行验证, 从应用上可以服务于机器翻译等任务。本文基于两种方式构建词向量进行汉语动词事件类型的预测, 一种是根据语言学特征有监督地构建词向量; 另一种是利用 word2vec 无监督地构建词嵌入向量。通过多元逻辑回归、支持向量机和人工神经网络分类器对汉语动词事件类型进行预测, 最终实现了 73.6% 的总体准确率。

关键词: 事件类型; 汉语动词; 语言学特征; 词嵌入; 分类; 预测

Prediction of Mandarin verbs' event types based on linguistics features vectors and word embedding vectors

Hongchao Liu¹, Chu-Ren Huang¹, Renkui Hou^{2,1}, Hongzheng Li³

(1. CBS, The Hong Kong Polytechnic University, Hong Kong; 2. School of Chinese Language and Literature, Ludong University, Yantai, 264001; 3. Institute of Chinese Information Processing, Beijing Normal University, Beijing, 100875)

Abstract: This paper sets foot in the prediction of event types of Mandarin verbs. Event types are trisected into state, activity and transition or quartered into state, activity, accomplishment and achievement. Previous linguistics studies of event types of Mandarin verbs have come up with various features for different event types but none of them are validated by statistical methods or computational methods. Both supervised vectors and unsupervised vectors are about to applied to do the prediction. Among them, linguistics features are extracted to constructing supervised vectors while certain contexts of verbs are used to constructed word embedding vectors by word2vec. We achieve an overall accuracy of 73.6% using classifiers of multinomial regression, supporting vector machine and artificial neural network.

Keywords: Event type; Mandarin verbs; Linguistics features; Word embedding; Classification; Prediction

1. 引言

事件类型的研究由来已久, 可以追溯到古希腊和古印度时期。亚里士多德时期, 事件类型被明确提出(Binnick, 1991)。Vendler (1967) 根据[telic]、[durative]、[dynamic]这三个特征将事件类型划分为四种:

表格 1 Vendler (1967)事件类型

事件类型	特征			例子
	Telic	Durative	dynamic	
状态	-	+	-	know、love

* 本文受到国家社会科学基金(16BYY110)的支持, 特此感谢。

活动	-	+	+	run、drive a car
完结	+	+	+	build、write a letter
达成	+	-	+	die、recognize

事件类型实际上是反映的是事件的内部时间结构（internal temporal structure）。内部时间结构包括几个不同的阶段（phase）：起点（inception）、续段（duration）和终点（final）（Guo, 1993; Pustejovsky, 1991）。而事件类型的划分事实上就是根据事件在各个时间阶段上表现出的特点进行的。

首先，根据在事件发生的过程中，事件是否是具有同质性（homogeneity），将事件可以划分为状态的（stative）和非状态的（non-stative）。状态与其他三种事件类型的最大区别就在于其在发生过程中的同质性，状态没有起点和终点，在其发生的续段内，性质不发生任何变化：

(1) Mary know Tom.

Mary 在任何时间都是认识 Tom 的，这个状态不随着时间发生变化。

其次，在非状态事件类型中，根据是否有界（telic），即是否具有起点和终点，可以将活动与其他两种事件类型区分开来。这里需要强调的是语言学上的有界和现实世界中的有界并不同。“run”本身的词义中并没有起点和终点，尽管一个人在现实世界中跑步有起点，最终也会有停止的终点。

最后，根据是否具有持续性（durative），可以将完成和达成区分开。达成没有持续性，其时间结构是点状，即起点和终点重合，因此也被成为点状事件（punctual）。完结事件，顾名思义，即事件发生并最终实现了某一个目标（attain a goal），比如“draw a circle”，既包含“draw”这个动作，又包含“a circle”这个最终的目标。

事件类型跟语言中的时间，尤其是体（aspect）有非常密切的关系。英语等屈折语言有相对丰富的时体标记，但是汉语中时体标记相对较少，因此，通过事件类型来认识汉语时体非常有必要，对于自然语言处理，尤其是机器翻译等有较为重要的意义。

展开讨论前，需要廓清本文的几个基本问题：

第一，动词、短语和句子都可以按照内部时间结构划分事件类型，事件类型具有组合性。我们同意 Pustejovsky (1991) 关于事件结构（event structure）的看法：事件结构与论元结构（argument structure）和物性结构（qualia structure）一样，都是词语语义实现（semantic specification）的层面之一，都具有递归性；由于递归性的存在，因此，从词语到短语和句子，都具有事件类型的属性。

第二，事件类型具有组合性。正如动词与动词的组合形成的不一定是动词性短语一样，词语的事件类型与其组成的短语的事件类型并不一定相同(Pustejovsky, 1991)。如“穿”的事件类型是“活动”，但是“穿着衣服”作为一个整体，其事件类型是“状态”。我们不同意将组成成分的事件类型与结构整体事件类型混淆的观点，如将“穿着衣服”中的“穿”的事件类型与整个短语的事件类型等同，认为其中的“穿”的事件类型是状态，单独说某人处于“穿”的状态没有意义，只能说某人处于“穿着衣服”的状态。

第三，事件类型所涉及的最小单位是词语的义项。事件结构是结构的语义层面之一，因此是由语义决定的，对于词语来说，就是由词义决定的。同一个动词有不同的义项，不同的义项可能对应不同的事件类型，从词语层面来讲，一个动词很可能同时具有多种事件类型。如动词“去”有三个义项，分别是“除去，去掉”（去皮）；“从所在地到别的地方”（去了三次）；“从事，做”（去讨论一下）。这三个义项分别对应了三种事件类型，第一个义项的事件类型是“状态”，第二个义项的事件类型是“变化”，第三个义项的事件类型是“状态”。

第四，本文不将体强迫现象（aspectual coercion）纳入考虑范围之内，如：

(2) He played the sonata for one day.

如果不考虑体强迫，(2)事实上不合语法。因为 *played the sonata* 是一个完结事件，因此不能受持续性状语 (*durative adverbials*) 修饰。但是如果谓语事件类型随着持续性状语进行调整，即赋予 *played the sonata* 一个反复义 (*iterative interpretation*)，则 (2) 又变得可以接受了。

体强迫 (*aspectual coercion*) 是非常有趣的现象，但是其事件类型的预测更多的与语用相关，超出了本文的研究范围。本文主要涉及的是非体强迫句中的汉语动词的事件类型。

本文研究的是汉语动词的事件类型，事件类型的组合性及其向量化表示将在另外一篇文章讨论。

2. 汉语动词事件类型的标注

汉语动词事件类型的标注主要包括了三个步骤，即动词的选取，事件类型标注体系的确定以及事件类型的标注。下文将分别介绍。

本文所选择的动词全部来自《汉语动词用法词典》(Meng, 1987)。词典从《现代汉语词典》(中国社会科学院语言研究所词典编辑, 1990)选择 1223 个动词，共划分出两千多个义项。

虽然词典中的动词都是以义项存在，但是语料库中的词语却只有词形及其词性标记，语料库中的词语及其词性标记与义项之间并没有规律性的对应关系，因此，只能对所有的动词，按照在语料库中的词性标记重新建立词条，对这些词条进行逐一标注。本文所使用的语料库为台湾中央研究院语料库(K.-J. Chen, Huang, Chang, & Hsu, 1996)，建立词条主要包括以下几步：

第一步，从台湾中央研究语料库中抽取《汉语动词用法词典》收录的所有词形及其 PoS。如果一个词形对应多个 PoS 则视为不同的词处理。如“长”有两个词性标记，VC 和 VH 分别对应 *zhang3* 和 *chang2*。这实际上是两个词，需要分别记录。

第二步，抽取第一步保留的动词在台湾中央研究语料库中的所有例句，按照句子和《汉语动词用法词典》中的义项描述判定动词的语义或义项归属。如“除”有两个记录，一个 PoS 是 VC，对应的例句表示的意义是“去掉”；另一个的 PoS 是 VJ，对应的例句表示的是“用一个数把另一个数分为若干份”。对应这两个词性标记分别建立词条。

建立词条之后再对照 Guo (1993)标注的动词事件类型进行标注就变得相对简单了。Guo (1993)对《汉语动词用法词典》中的所有义项进行了事件类型的标注。下表是其事件类型及其语言学特征：

表 1 Guo (1993)事件类型及其语言学特征

类别	时间结构	小类	v+了	v+TP	v+着	在/正在+v	v+过	例子
Va	无限结构		-	-	-	-	-	是
Vb	前限结构		I	I	-	-	-	知道
		1	I	I	-	-	+	相信
Vc	双限结构	2	I	I	+	-	+	有
		3	I,F	I	+	-	+	醉
		4	I,F	I	+	+	+	等
Vd	后限结构	5	I,F	I,F	+	+	+	吃
		1	F	F	+	+	+	提高
		2	F	F	-	+	+	离开
Ve	点结构		F	F	-	-	+	来

注释: I表示“开始”(inception); F表示“结束”(final); TP表示“时间短语”

Guo (1997)又进一步合并了上表中的事件类型, 形成了一个事件类型体系:

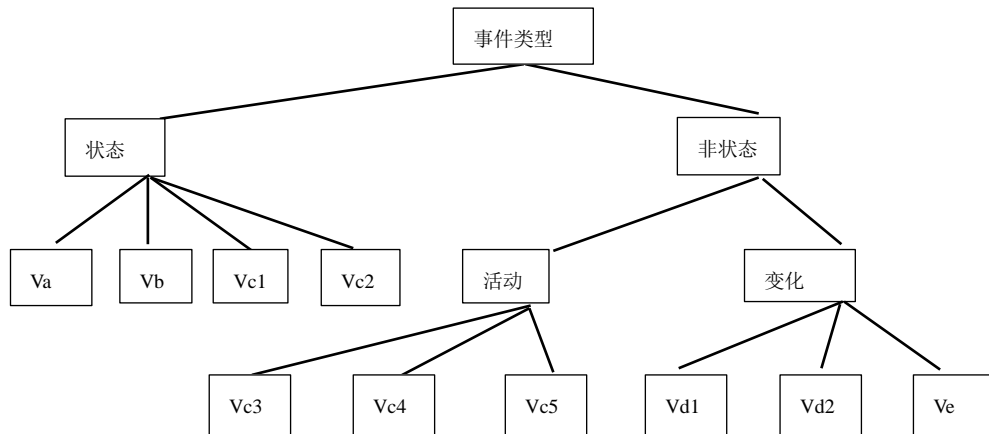


图 1 Guo (1997)的事件类型分类体系

本文并没有照搬 Guo (1993)的事件类型体系, 而是对 Guo (1993)的事件类型体系进行了验证和调整。

Guo (1993)的事件类型划分是以动词能否与其选择的体标记词 (aspectual marker) 共现以及共现后表示的语义划分的事件类型, 那么这种划分就是可以用计算方法验证的。验证实验的基本假设是按照 Guo (1993)对事件类型及其与特征词的共现情况构建矩阵, 那么按照语义距离形成的层级分类体系应当与 Guo (1997)的层级分类体系相似。

下表是按照表 1 构建的矩阵:

表 2 Guo (1993)的事件类型小类与其语言学特征矩阵

小类	了	了_I	了_F	T	T_I	T_F	着	在	过
va	0	0	0	0	0	0	0	0	0
vb	1	1	0	1	1	0	0	0	0
vc1	1	1	0	1	1	0	0	0	1
vc2	1	1	0	1	1	0	1	0	1
vc3	1	1	1	1	1	0	1	0	1
vc4	1	1	1	1	1	0	1	1	1
vc5	1	1	1	1	1	1	1	1	1
vd1	1	0	1	1	0	1	1	1	1
vd2	1	0	1	1	0	1	0	1	1
ve	1	0	1	1	0	1	0	0	1

注释: I表示“开始”(inception); F表示“结束”(final); T表示“时间短语”

以此矩阵为基础进行的层次聚类如下:

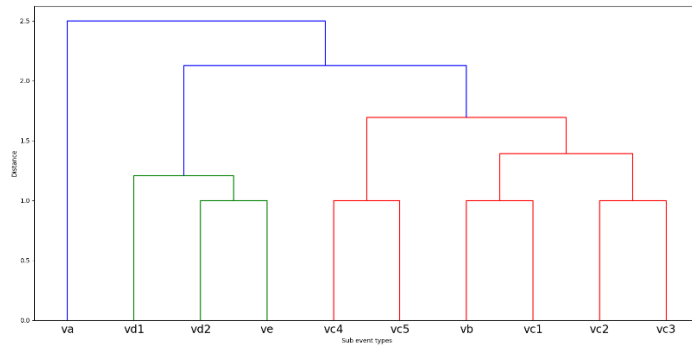


图 2 Guo (1993)的典型事件类型聚类结果

我们用同表象相关系数 (CPCC, cophenetic correlation coefficient) 来衡量聚类的好坏。CPCC 主要是衡量未经建模的原始矩阵和建模之后的矩阵之间的相似度, 后者代表的是层次聚类算法的矩阵(Hou & Jiang, 2014); Hou, Yang, and Jiang (2014)。实际上 CPCC 可以表示层次聚类算法所使用的矩阵在多大程度上保留了原始矩阵的信息。这个值越接近 1, 表明层次聚类采用的矩阵与原始矩阵之间的相似度越高。

上图结果的 CPCC 为 0.85 表明聚类结果令人满意。从上图可以看出, vb 与 vc1 的语义距离最近, vc2 与 vc3 的语义距离最近, vb、vc1、vc2 和 vc3 又聚成一个大类; vc4 与 vc5 的语义距离最近, 两者组成一个类别; vd2 和 ve 的语义距离最近, vd1、vd2 和 ve 组成了一个类别; va 单独成类。

这基本上与 Guo (1997)的分类一致。vb、vc1、vc2 和 vc3 组成了“状态”; vc4 和 vc5 组成“活动”; vd1、vd2 和 ve 组成“变化”。

但是也有几点不同:

首先, Guo (1997)将 vc3 归入 activity, 但是根据语义距离的计算, vc3 与 vc2 的相似性更大, 而不是 vc4。Guo (1993)事实上指出过, vc1、vc2 和 vc3 在以往的研究中都属于状态动词, 但是他认为 vc3 的“动作性已经非常强”。可是事实上, 如果严格按照 Guo (1993)提出的标准进行划类, 则类似“爱”、“病”和“承担”这类 vc3 词, 事实上应当与 vc2 和 vc1 归入一类, 同属于“状态”。

其次, Guo (1993, 1997)都将 va 归入“状态”, 但是层次聚类将其单独列为一类。va 是一个比较特殊的类别, 其内部的词如“是”、“等于”和“敢”等词表示的是一种绝对的状态, He (1992)将之称为“绝对状态动词” (absolute stative verb), 这类词与其他状态动词的最大区别是无法用“很”进行修饰。

再次, Guo (1993, 1997)将“活动”和“变化”合并为“非状态动词”, 而层次聚类的结果表明, “状态”跟“活动”的距离更近。

这实际上不能说是两者的不同, 因为按照 Vendler (1967)提出的标准, 完全可以将“状态”和“活动”合并为“无界动词” (non-telic verb), 而将“变化” (包括“完结”和“达成”) 称为“有界动词” (telic verbs)。由于 Vendler (1967)提出了 telic、dynamic(或 stative)和 durative 三个要素来区分三者, 所以怎么样合并“状态”、“活动”和“变化”三者实际上取决于采取什么样的观察角度, 即首先使用哪个要素来对三者进行划类。如果按照 Guo (1993)的标准, 从语义距离上说, “状态”跟“活动”更相似, 也就是 Guo (1993)的标准实际上更接近于使用[telic]作为首次划类的标准。

所以实际需要解决的问题是 va 和 vc3 的归类问题。后文将通过多元逻辑回归分析证明, va 和 vc3 归入 state 是较为合适的。

所以最终, 本文所采用的分类体系实际上如下:

- A. 状态: va, vb, vc2, vc3
- B. 活动: vc4, vc5
- C. 变化: vd1, vd2, ve

最后一个问题是 Guo (1993)只是对占数据集绝大多数的 10 个典型类别进行了划类，对于剩余的 29 个其他例外类别并没有进行划类。之所以称之为例外类别，并不是因为这些类别不能通过选择的语言学特征来衡量，而是它们与选择的特征词的共现情况与 10 个典型的类别都不一致，所以无法对之进行归类。

但是通过层次聚类则可以解决这一点，层次聚类会按照语义距离对所有的类别进行聚类。

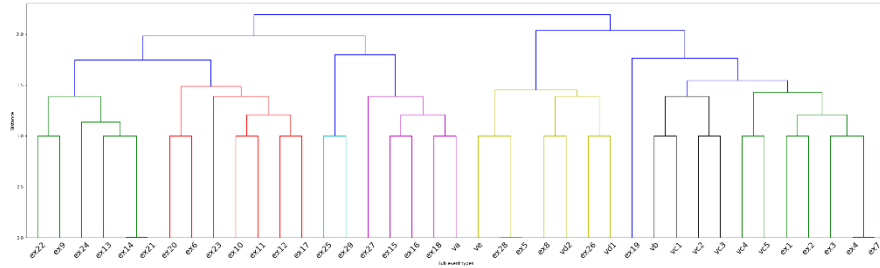


图 3 Guo (1993)所有事件类型的聚类结果

CPCG 为 0.75，表明聚类结果也是令人满意的。例如，ex1 与 vc4 和 vc5 的语义距离最近，因此可以归入“活动”，ex1 包括的词如“表决”、“挑拨”以及“提倡”；再如 ex18 与 va 的语义距离最近，可以归为一类，ex18 包含的词如“叫”（他叫张三）、“愿意”以及“尊敬”等。

如此一来，所有的例外类别也都按照与典型类别的语义距离划入了相应的典型事件类型中去。本文所使用的标注体系也最终确定下来。我们采用人工标注的方式，对由台湾中央研究院语料库和《汉语动词用法词典》对照产生的各个词条进行了事件类型的标注，基本统计情况如下：

表 3 事件类型分布状况

	状态	活动	变化	总计
数目	387	978	245	1610

3. 汉语动词事件类型的预测

事件类型的预测是非常困难也很有意义的任务(Tommaso & Huang, 2012)。其困难在于即便是人工对其判定，未经训练的标注者准确率也非常低，只有经过专门的语言学训练的可靠标注者，准确率才令人满意。Tommaso and Huang (2012)采用“众包”（crowdsourcing）的方式对意大利语进行事件类型的标注试验，经过与金标注（golden standard）对比发现，普通标注者的准确率只有 43%，而经过语言学训练的可靠标注者（trusted rater）准确率可以达到 93%。

Siegel (1998)是较早进行英语事件类型预测的研究之一。他选择了十四个语言学特征用于构建矩阵，在此基础上采用了决策树（decision tree）、逻辑回归（logistics regression）和基因编程（genetic programming）分类器实现了动词事件类型的预测，下表是其对“状态”和“非状态”（他称之为 event）的区分：

表 4 Siegel (1998)区分“状态”和“非状态”结果

	Overall accuracy	状态		非状态	
		Recall	Precision	Recall	Precision
Decision tree	93.9%	74.2%	86.4%	97.7%	95.1%

Genetic programming	91.2%	47.4%	97.3%	99.7%	90.7%
Logistic regression	86.7%	34.2%	68.3%	96.9%	88.4%
Baseline	83.8%	0.0%	100.0%	100.0%	83.8%

但是也存在一些问题。

正如 Zarcone and Lenci (2008)所指出的, Siegel (1998)分开两次进行分类试验, 没有测试模型在多分类任务上的总体准确率。第二次试验通过分类回归树 (classification and regression tree, CART) 达到了最高的准确率 74% (Baseline:63.3%); 两次试验采用的均是 hold out 检验法, 即简单的将数据集分为了训练集和测试集, 这种方法对模型的验证说服力不强, 应当进行 K-fold cross validation。

Zarcone and Lenci (2008)采用了两种方法构建用于事件类型分类的特征。一种是有监督的方法, 一种是无监督的方法。其中有监督的方法通过对动词标注大量的句法信息包括论元结构等构建向量矩阵, 最终实现了 85.4%的总体准确率; 采用无监督的方法构造向量矩阵主要是基于分布式语义模型, 通过抽取上下文(一定窗口范围内)的共现词频率来构建矩阵, 最终实现了 72.5%的准确率。

但是关于汉语动词事件类型预测的研究几乎没有。而且汉语的识别难度要高于英语和意大利语。这两者都是屈折语, 带有丰富的屈折语标记, 所以可以直接使用这些标记进行预测, 取得较高的识别效果。但是汉语几乎没有这样的标记, 根据以往的语言学研究如 Tai (1984)、Deng (1985)、P. Chen (1988)、He (1992)和 Guo (1993, 1997)等, 只能发现“很”、“在”、“正在”、“着”、“了”、“过”等几个体标记 (aspectual marker)。

本文通过两种方式构建了动词事件类型的特征矩阵。一种是使用语言学特征构建向量矩阵, 另一种是使用 word2vec 构建向量矩阵。下面将分别介绍通过这种方式构建向量矩阵实现动词事件类型的预测过程。

对于语言学特征, 本文选择了“很”、“在”、“正在”、“着”、“了”、“过”以及动词在台湾中央研究院语料库中的频率信息。值得注意的是, 除了“很之外”, 其他几个体标记都有同形词, 如“在”既可以是介词(在家里), 也可以是体标记(在吃饭), 依靠词性标记可以将之区分开, 特征词及其词性标记如下:

表 5 特征词及其词性标记

特征词	PoS	解释	例子
很	Dfa	动词前程度副词	很喜欢
在	Di	体副词	在跑步
正在	D	副词	正在吃饭
着	Di	体副词	走着
了	Di	体副词	完成了
过	Di	体副词	去过

前文提及了只有将 va 和 vc3 归入“状态”才是较为合适的做法, 这可以通过构建的语言学特征向量矩阵进行验证。既然 Guo (1993)提出的事件类型都是按照其提出的语言学特征进行归类, 那么如果依靠这些特征构建出的动词的特征向量应该可以实现动词事件类型的预测。我们采用动词与目标词的共现频率作为特征向量值。共现频率的计算是在台湾中央研究院的精标语料库上实现的。我们采用了多元逻辑回归分析来验证两者之间的关系。

实验的基本假设是, 如果 va 和 vc3 归入“状态”才是较为合适的, 那么使用选择的语言学特征就可以拟合出比随机模型好的多元逻辑回归模型; 同时, 将 va 和 vc3 归入其它事件类型, 则不能实现比随机模型好的多元逻辑回归模型。

多元逻辑回归是逻辑回归的一种, 使用最大似然估计结出相关参数。使用多元逻辑回归

对自变量和因变量进行分析时，需要先进行似然比检验，以确定特征的有效性，其基本原理就是通过引入了特征的模型和未引入特征的随机模型进行对比。由于两者似然值的差值服从 χ^2 分布，因此可以借助该统计检验量来验证模型的有效性。

下面是将 va 和 vc3 归入“状态”时的似然比检验结果：

表 6 将 va 和 vc3 归入“状态”时的似然比检验结果

Log likelihood	Log likelihood null	LLR p-value
1467.3	-1500.5	1.491e-09

p 值表明模型是有效的，我们也检验了将 va 和 vc3 放入其他类别，p 值都为 1.00，这表明模型都是无效的。由于篇幅原因，不再具体展开。由此证明，va 和 vc3 应当归入“状态”。

作为对比，本文引入了词性标记的语言学特征向量。基于这些语言学特征向量矩阵，除多元逻辑回归外，本文进一步使用支持向量机和人工神经网络分类器对标注的数据集进行了分类试验，所有的实现使用的都是十折交叉检验。下面是各个分类器在选择不同特征的情况下实现的总体准确率（overall accuracy）：

表 7 基于语言学特征向量的试验结果

	MNLogit	SVM	ANN
Adverb	65.40%	65.53%	63.54%
Adverb + freq	66.21%	61.93%	57.14%
PoS	67.83%	67.89%	67.83%
Adverb + PoS	68.82%	66.46%	64.91%
Adverb + freq + PoS	69.32%	61.93%	60.19%
Baseline		60.7%	

注释：MNLogit 表示多元逻辑回归；SVM 表示支持向量机；ANN 表示人工神经网络

仅仅使用体标记副词（aspectual adverb）基于多元逻辑回归了 65.4% 的准确率；作为对比的词性标记（PoS）基于多元逻辑回归实现了 67.83% 的准确率，单单从数字上看，似乎使用词性标记效果好于使用体标记副词，但是使用配对样本检验的结果表明两者的差异并不具有显著性：

表 8 MNLogit(adv)和 MNLogit(PoS)试验结果统计

	N	Mean	Std. err	Std. err mean
MNLogit(adv)	10	0.6540	0.0321	0.0101
MNLogit(PoS)	10	0.6783	0.0277	0.0088
差异	10	-0.0242	0.0459	0.0145

表 9 MNLogit(adv)和 MNLogit(PoS)试验结果配对样本 t 检验结果

T	P	95% confidence interval of the difference	
		lower	upper
-1.67	0.129	-0.0570	0.0086

所以两种特征在预测动词事件类型的能力上都相似。将所有特征考虑进来，实现了最高 69.32% 的整体准确率，与其他情况进行配对 t 检验表明，差异是显著的。

基于 aspectual marker 等共现词来判定动词的事件类型实际上在某种程度上是根据上下

文来预测动词的事件类型，只是这里的上下文是经过筛选的上下文。筛选工作是由语言学家进行的，但是这些信息很有可能是不全的，缺失的信息在一定程度上会影响动词事件类型的预测准确率。

既然很难将能够影响动词事件类型的所有上下文信息依靠人工的方式挑选出来，那么解决这个问题一个办法是把上下文的所有信息都用来预测动词的事件类型。其理论基础就是分布式语义假设。

Distributional semantic hypothesis 认为出现在相似的上下文中的词，在语义上更加相近(Harris, 1954)。基于分布式语义假设构建的词与上下文共现词形成矩阵就称为分布式语义模型。由于基于上下文共现词不需要人工的选择，因此是一种非常好的无监督构建向量矩阵的方法，并且目前在多种任务上都被证明有相当不错的表现(Liu, Karl, Enrico, & Huang, 2016; Santus, Yung, Lenci, & Huang, 2015)。

基于分布式语义假设，我们的推论是，上下文相似的动词，其事件类型也相似。我们选择了与当前动词的距离为 5(左 5, 右 5)的所有的上下文，通过 word2vec 构建了 500 维的向量矩阵。Word embedding 向量是在 sinica corpus(K.-J. Chen et al., 1996)以及 Chinese Gigaword corpus(Hong & Huang, 2006)上训练完成。所有的词性标记都与 Sinica corpus 一致。

分别采用多元逻辑回归、神经网络和支持向量机作为分类器进行了试验，模型的整体准确率如下：

表 10 基于词嵌入向量的动词事件类型预测效果

	MNLogit	SVM	ANN
Overall accuracy	69.88%	73.60%	72.24%
Baseline	60.70%		

可以看出，使用支持向量机作为分类器达到了最高的分类准确率 73.60%，比 baseline 高 12.9 个百分点，比使用语言学特征的模型的最高准确率 69.32%，高 4.3 个百分点。下面我们使用配对 t 检验检查差异是否是显著的：

表 11 SVM (word2vec) 和 MNLogit (adv+freq+pos)的配对 t 检验结果

T	P	95% confidence interval of the difference	
		lower	upper
2.89	0.018	0.0093	0.0764

统计结果显示，使用无监督方式构造的上下文向量构建的模型，其预测准确率显著地高于使用语言学特征构建的模型。

下面报告每个小类的分类效果：

表 12 基于 word2vec 的支持向量机分类结果

事件类型	precision	recall	f1-score	support	Overall accuracy
状态	0.72	0.46	0.56	387	
活动	0.73	0.93	0.82	978	73.60%
变化	0.79	0.38	0.51	245	

4. 结语

本文分布基于语言学特征向量和词嵌入向量实现了动词事件类型的预测，本文的主要贡献有：

第一，基于 Guo (1993)对《汉语动词用法词典》重新标注和验证了动词事件类型，形成了总量达到 1610 的汉语动词事件类型数据集，可以直接用于相关模型的训练和评价、；

第二，实现了汉语动词事件类型的预测，最高准确率达到 73.6%。

我们的未来工作将集中于短语和句子事件类型的识别。

参考文献:

- Binnick, R. I. (1991). *Time and the verb: A guide to tense and aspect*: Oxford University Press.
- Chen, K.-J., Huang, C.-R., Chang, L.-P., & Hsu, H.-L. (1996). Sinica corpus: Design methodology for balanced corpora. *Language*, 167, 176.
- Chen, P. (1988). 论现代汉语时间系统的三元结构. [On the Ternary Structure of Temporal system of Mandarin Chinese]. *Chinese Language*, 6, 22.
- Deng, S. (1985). The temporal structure of Verbs in Chinese. *Language Teaching and Linguistic Studies*(4), 7-17.
- Guo, R. (1993). 汉语动词的过程结构. [The process structure of Mandarin verbs]. *Chinese Language*, 6, 410-419.
- Guo, R. (1997). 过程和非过程——汉语谓词性成分的两种外在时间类型. [Process and non-process: two types of the external temporal structure of Mandarin predicates]. *Chinese Language*(3), 162-175.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3), 146-162.
- He, B. (1992). *Situation types and aspectual classes of verbs in Mandarin Chinese*.
- Hong, J.-F., & Huang, C.-R. (2006). *Using chinese gigaword corpus and chinese word sketch in linguistic research*. Paper presented at the The 20th Pacific Asia Conference on Language, Information and Computation (PACLIC-20). November.
- Hou, R., & Jiang, M. (2014). Analysis on Chinese quantitative stylistic features based on text mining. *Digital Scholarship in the Humanities*, fqu067.
- Hou, R., Yang, J., & Jiang, M. (2014). A Study on Chinese Quantitative Stylistic Features and Relation Among Different Styles Based on Text Clustering. *Journal of Quantitative Linguistics*, 21(3), 246-280.
- Liu, H., Karl, N., Enrico, S., & Huang, C.-R. (2016). *EVALution-MAN: A Chinese Dataset for the Training and Evaluation of DSMs*. Paper presented at the Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Paris, France.
- Meng, C. (1987). *Chinese verb usage dictionary* (Vol. 612): The Commercial Press.
- Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, 41(1), 47-81.
- Santus, E., Yung, F., Lenci, A., & Huang, C.-R. (2015). EVALution 1.0: an Evolving Semantic Dataset for Training and Evaluation of Distributional Semantic Models. *ACL-IJCNLP 2015*, 64.
- Siegel, E. V. (1998). *Linguistic indicators for language understanding: using machine learning methods to combine corpus-based indicators for aspectual classification of clauses*. (Ph.D), Columbia University.
- Tai, J. (1984). Verbs and times in Chinese: Vendler's four categories. *CLS 20/2. Papers from the parasession on lexical semantics, Chicago, IL: Chicago Linguistic Society*. (289-296) Tenny, Carol 1992. « The Aspectual Interface Hypothesis », in: I. Sag & A, 92(5), 75-78.
- Tommaso, C., & Huang, C.-R. (2012). *Sourcing the Crowd for a Few Good Ones: Event Type Detection*. Paper presented at the COLING (Posters).

Vendler, Z. (1967). *Linguistics and philosophy*: Cornell University Press.

Zarcone, A., & Lenci, A. (2008). *Computational Models for Event Type Classification in Context*.

Paper presented at the LREC.

中国社会科学院语言研究所词典编辑. (1990). *现代汉语词典*: 商务印书馆出版社.