

基于深度信念网络的事件识别

张亚军, 刘宗田, 周文

(上海大学计算机工程与科学学院, 上海 200444)

摘要: 事件识别是信息抽取的重要基础. 为了克服现有事件识别方法的缺陷, 本文提出一种基于深度学习的事件识别模型. 首先, 我们通过分词系统获得候选词并将它们分为五种类型. 然后选择六种识别特征并制定相应的特征表示规则用来将词转化为向量样例. 最后我们通过深度信念网络抽取词的深层语义信息, 并由 Back-Propagation (BP) 神经网络识别事件. 实验显示模型最高 F 值达 85.17%. 同时, 本文还提出了一种融合无监督和有监督两种学习方式的混合监督深度信念网络, 该网络能够提高识别效果 (F 值达 89.2%) 并控制训练时间 (增加 27.50%).

关键词: 事件识别; 深度学习; 识别特征; 特征表示; 混合监督

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)06-1415-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.06.020

Event Recognition Based on Deep Belief Network

ZHANG Ya-jun, LIU Zong-tian, ZHOU Wen

(School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China)

Abstract: Event recognition is critical to information extraction. To overcome limitations of the exiting event recognition approaches, we proposed an event recognition model based on deep learning (DL-ERM). Firstly, we acquired candidate words through a word segmentation system and classified them into five categories. Then, we selected six recognition feature layers and constructed corresponding feature representation rules to convert words into vector samples. Finally, we employed a deep belief network (DBN) to extract deep semantic features of words, and used a back propagation neural network to identify events. The results of experiments show that the maximum F -measure is 85.17%. Furthermore, we presented a hybrid-supervised DBN, which combines the unsupervised and supervised learning. The novel DBN improves the recognition performance (89.2% F -measure) and effectively controls the training time (increased by 27.50%).

Key words: event recognition; deep learning; recognition feature; feature representation; hybrid supervision

1 研究背景

自然语言组织的文本通过事件来表达更高层次的语义信息. 识别文本中的事件就可以理解文本表达的确切含义, 有利于对文本进行结构化的分析, 为实现基于事件的自然语言处理打下坚实的基础^[1-5]. 在自然语言组织的文本中, 我们通过触发词来标识事件. 触发词是事件发生的重要标志, 其在句子中包含的信息量通常最大. 例如: 汶川发生地震. 地震就是这个句子中的触发词, 该事件属于地震事件, 它的地点要素是汶川. 在任何事件中, 触发词是必不可少的, 其他要素可以缺省. 所以事件识别的本质就是触发词的识别.

作为自然语言处理领域的热门研究方向, 事件识

别技术在自动文摘^[6,7]、自动问答^[8,9]和信息检索^[10-13]等领域有着广泛的应用. 现有的事件识别方法主要有基于规则以及浅层的机器学习两种. 基于规则的方法主要是通过手工自定义事件模板, 采用各种模式匹配算法将文本与自定义的事件模板进行匹配. 姜吉发^[14]提出了一种具有可移植性的事件模式的获取方法, 可以从一个原始语料中自动学出相应的事件模式. 陈群秀等^[15]提出了一种基于框架的信息抽取模式并建立了统一的灾难性事件框架, 利用框架的继承-归纳特性简化系统实现过程, 概括事件信息. 冯礼^[16]基于事件框架的突发事件信息抽取使用预定义的事件框架, 进行新闻要素的信息抽取, 同时探测事件框架中未事先设定的新侧面信息, 对事件框架进行补充. 基于浅层的机器

学习方法把重点放在分类器的构建和特征的发现、选择、组合上,把事件识别看成分类问题.2002年 Chieu 和 Ng 首次将最大熵分类器^[17]引入到事件抽取中,并用该分类器完成了对事件及其要素的识别.付剑锋等用依存分析发掘触发词与其它词之间的句法关系,并以此为特征在 Support Vector Machine(SVM)分类器上对事件进行分类,最终实现事件识别^[18].2006年 David Ahn^[19]结合 Megam 和 Timbl 两种方法分别实现了事件抽取中事件类别和元素的识别.在 Ahn 的方法中,最重要的一个步骤就是判断句子中的每个词是否是最能描述某个事件发生的触发词,如果是,则将其归为正例,并用一个多类型分类器对其进行分类,获得其所属事件类别,从而得出其所含的事件元素类型,用于构建识别每一类事件元素的分类器.

虽然现有的事件识别方法能够取得一定的效果,但是都存在着一些限制.基于规则的方法通常需要人工来制定相应的规则,而规则的合理性又依赖于对事件特征的分析.因此存在着特征提取困难、规则制定的过程复杂、可移植性差等问题.基于浅层的机器学习方法包括很多种,例如 SVM,最大熵模型以及神经网络等.这些方法同样存在着一定的缺陷,SVM 在处理多分类问题时存在困难,最大熵模型的优化过程较为复杂,简单的神经网络容易收敛到局部最小值等,从而极大的影响识别效果.近年来机器学习领域出现的深度学习(Deep Learning)方法^[20],通过学习一种深层非线性网络结构,实现复杂函数逼近,表征输入数据分布式表示,获得高维数据的特征函数.通过构建具有很多隐层的机器学习模型和海量的训练数据,学习更加抽象的特征,从而最终提升分类或预测的准确性.目前已有一些学者对该方法应用于文本挖掘领域做了一些尝试^[21,22],并取得了不错的效果.但在事件识别领域,还未见到用该方法抽取深层语义信息的相关报道.

基于上述原因,本文提出基于深度学习事件识别模型(Event Recognition Model based on Deep Learning, DL-ERM).研究 Deep Learning 深度学习结构的处理机制,探索面向事件识别的语义特征泛化表示,利用 Deep Learning 深度学习机制自动挖掘深层语义信息,研究深层语义信息在事件识别中的作用.本文通过特征分析的方法,对文本中的词从6个方面进行特征表示并且生成特征向量,最后由深度信念网络来分析特征向量并给出识别结果.经过实验分析,我们发现该模型的识别效果比现有的事件识别方法有一定的提升.而且本文提出的模型不仅可以直接用于事件识别,还可以用来处理其他事件要素的识别,最终实现基于深度学习的事件以及相关要素的自动标注.

2 深度信念网络

2.1 受限玻尔兹曼机

受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)是由 Hinton 等^[20]提出的一种生成式随机神经网络,如图1所示,该网络由可见单元和隐藏单元构成,两者都是二元变量,其状态取 $\{0,1\}$.整个网络是一个二部图,只有可见单元和隐藏单元之间才会存在边.

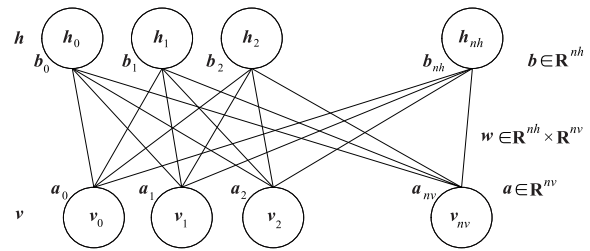


图1 RBM结构图

图1中的 v 和 h 分别表示可见单元和隐藏单元, w 表示可视层和隐含层之间的权值, a 和 b 分别表示可视层和隐含层的偏置量.利用联合概率分布公式可以由已知的可视层节点值得到隐含层的节点值^[20]:

$$P(\mathbf{h}_j = 1 | \mathbf{v}) = \left(1 + \exp(-\mathbf{b}_j - \sum_{i=1}^{nv} \mathbf{w}_{j,i} \mathbf{v}_i) \right)^{-1} \quad (1)$$

RBM 是对称网络,同理利用联合概率分布公式可以由已知的隐含层的节点值得到可视层的节点值:

$$P(\mathbf{v}_i = 1 | \mathbf{h}) = \left(1 + \exp(-\mathbf{a}_i - \sum_{j=1}^{nh} \mathbf{w}_{j,i} \mathbf{h}_j) \right)^{-1} \quad (2)$$

训练 Restricted Boltzmann Machine(RBM)目标就是为了求出使得联合概率 $p(\mathbf{v}, \mathbf{h})$ 值最大的参数 $\theta(\mathbf{w}, \mathbf{a}, \mathbf{b})$,传统的马尔可夫链蒙特卡罗算法训练时收敛速度很难保证,并且难以确定步长. Hinton^[20]提出利用 Contrastive Divergence(CD)准则,在保持精度的同时能够快速提高计算速度.利用 Kullback-Leibler 距离衡量两个概率分布的“差异性”,表示为 $KL(P \| P')$,如式(3)所示:

$$CD_n = KL(p_0 \| p_\infty) - KL(p_n \| p_\infty) \quad (3)$$

2.2 混合监督的深度信念网络

标准的 Deep Belief Network(DBN)是由若干层 RBM 和一层 BP(Back-Propagation)组成,RBM 层的训练是无监督的,最后通过 BP 网络来反向微调.这样的结构使得网络的训练速度提高,但是由于 RBM 层之间缺乏有监督训练,使得网络误差逐层向上传递,最终影响识别效果.因此本文提出一种改进型的混合监督 DBN 网络,其结构如图2所示.该网络为每一个隐含层都增

加一层 BP 网络,对通向这个 BP 网络的所有隐含层参数进行有监督学习.例如,第二层隐含层增加的有监督学习过程修改第一和第二层隐含层网络参数.这样改进后的 DBN 网络,同样能使得底层的网络参数得到充分的学习,缓解导数消亡问题,提高模型的识别效果.

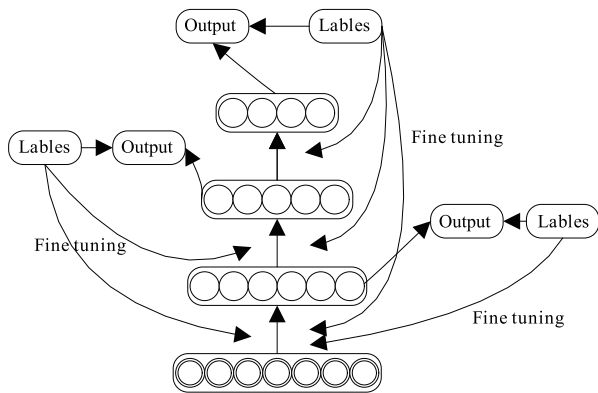


图2 混合监督的深度信念网络

混合监督的深度信念网络工作过程如下:第 1 步,首先无监督地训练每一层 RBM 网络,确保在特征向量映射到不同特征空间时,都尽可能多地保留特征信息.训练时使用小批量数据处理模式,通过 CD 算法加快收敛过程.训练后的 RBM 网络的相关参数将作为下一步有监督调整的输入.第 2 步是有监督的调整 RBM 网络.由于第 1 步的无监督训练中,RBM 训练过程中产生的误差可能会自底向上逐层传递,对最终的识别效果产生影响,所以需要现有的 RBM 层以及所有前层的 RBM 进行有监督的微调.有监督的调整工作主要是通过 BP 算法完成,通过整个 BP 网络自上而下的调整,确保误差控制在最低的范围,并且减少误差向上传递的可能,从而确保整个 DBN 的特征向量映射达到最优,经过有监督训练后的 RBM 网络参数将作为下一层 RBM 训练的输入.以上两个过程结束后一次混合监督的训练过程结束.对整个模型的所有层进行这种混合训练学习,就能实现系统识别效果的提升.

3 基于深度信念网络的事件识别模型

DL-ERM 主要包括语料选择、预处理、特征向量生成、深度分类器四个主要模块,见图 3. 预处理模块主要功能是对语料中的句子进行分词,分词工具为 LTP^[23],分词后按照语料中标注类型给词分类.特征向量生成模块是将词按照六个特征层(词性、依存语法、长度、位置、词与核心词的距离和触发词频度)生成特征向量,特征向量是二值模式.深度分类器主要功能是通过训练语料来生成稳定的 DBN 网络,并对测试语料中的触发词进行识别,本文中的深度分类器有两种,它们的学习方式分别是无监督以及混合监督.

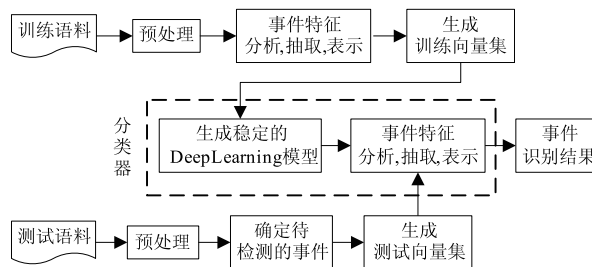


图3 DL-ERM模型框架

3.1 语料选择

本文使用的语料库 CEC 2.0 是由上海大学语义智能实验室标注.包括 333 篇关于地震、火灾、交通事故、恐怖袭击以及食物中毒新闻文本.通过半自动的方式标注了事件触发词、参与者、对象、时间、地点等要素.语料选择包含两部分:训练语料以及测试语料.两者的比例大约为 5:1.训练语料包括五大类 271 篇语料,其中地震 51 篇、火灾 61 篇、交通事故 70 篇、食物中毒 50 篇、恐怖袭击 39 篇,包含事件触发词 4690 个.测试语料包括五大类 62 篇语料,其中地震 12 篇、火灾 14 篇、交通事故 15 篇、食物中毒 11 篇、恐怖袭击 10 篇,包含事件触发词 1301 个.

3.2 预处理系统

预处理系统的主要功能是将语料库中的 XML 格式语料文档进行分析处理,并通过分词系统得到相关元素分词后的词性、依存语法等,最后保存为 XML 文件.本文中所有词根据事件要素的定义分为 5 种类型,分别是触发词、参与者、对象、时间和地点.首先使用分词系统进行切分,切分后与该语句中原来标注的触发词对比,如果相同标记该词为触发词.其他要素也通过该方式设置类别.预处理系统的结果将作为特征表示以及向量生成的基础.

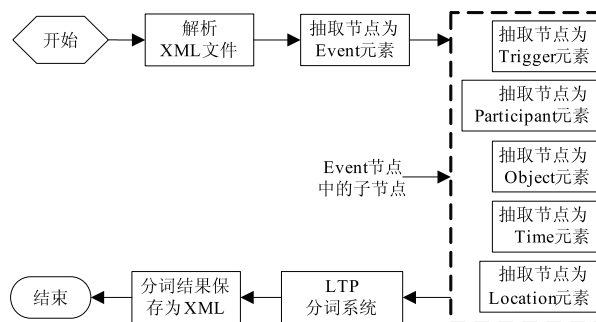


图4 预处理流程图

3.3 深度分类器

深度分类器以词的特征向量作为输入,词的分类作为训练以及测试的对比标准.深度分类器训练模式是批量数据训练模式. DL-ERM 中的深度分类器有两

种,一种是无监督的 DBN 网络(标准的 DBN)分类器,由于采用无监督的方式,所以训练的时间短,效率高,但是识别效果较差.另一种混合监督的深度分类器,该分类器结合无监督以及有监督两者训练方式来减少误差向上层传递的可能,该分类器在提升识别效果的同时训练时间的增长在可控的范围内.算法 1 详细地描述了混合监督的深度分类器的工作过程.

算法 1 混合监督深度分类器工作过程

```

Begin
1: load NE5; //加载输入特征向量.
2: layerNum = 6; rbmNum = 4; //确定参与训练的特征层数和 DBN 网络深度.
3: train_x ; test_x ; train_y ; test_y ; //训练集与测试集以及两者对应的分类.
4: numepochs = 1 ; batchsize = 100 ; momentum = 0; alpha = 1; //初始化网络参数.
5: eventType = 3; //待识别目标的类型值,3 表示触发词.
6: dbn_sizes //设置 dbn 网络层数以及各层的对应的输入输出单元数量.
7: dbn = dbnsetup( dbn, train_x, opts ); //初始化 DBN 网络
8: dbn = dbntrain( dbn, train_x, train_y, opts, eventType ); //调用 DBN 训练函数.
9: //执行 dbntrain 算法
10: n = numel( dbn. rbm ); //判断 DBN 网络深度.
11: x = train_x; //保存初始训练样本.
12: dbn. rbm{1} = rbmtrain( dbn. rbm{1}, x, opts, eventType ); //无监督训练.
13: for i = 2 : n
14: z = i - 1; //设置混合监督深度信念网络停止工作的判断条件.
15: if( z < n ) //最顶层反向微调由整个 DBN 网络顶层的 BP 算法处理.
16: nnhybrid = dbnunfolttonnhybrid( dbn, eventType, z ); //由 DBN 相关网络参数,设置有监督调整所使用的 BP 网络初始化参数.
17: nnhybrid. activation_function = 'sigm'; //设置 BP 网络的激活函数
18: nnhybrid = nnttrainhybrid( nnhybrid, train_x, train_y, opts ); // BP 对所有前层 RBM 进行有监督与无监督训练共同构成混合监督.
19: for m = 1 : z //遍历并获取有监督训练后的 RBM 层相关参数
20: colNum = size( nnhybrid. W{m}, 2 );
21: temp = nnhybrid. W{m};
22: dbn. rbm{m}. b = temp( :, 1 : 1 );
23: dbn. rbm{m}. W = temp( :, 2 : colNum );
24: end
25: x = rbmup( dbn. rbm{ i - 1 }, x ); //将 RBM 输出作为下一层的输入.

```

```

26: dbn. rbm{ i } = dbntrain( dbn. rbm{ i }, x, opts ); //进行下一层 RBM 训练.
25: end
27: nn = dbnunfolttonn( dbn, eventType ); //初始化 BP 反向微调网络
28: nn. activation_function = 'sigm'; //设置 BP 网络的激活函数
29: nn = nnttrain( nn, train_x, train_y, opts ); // BP 反向微调
30: [ er, bad ] = nntest( nn, test_x, test_y, opts, eventType );
31: R = ; P = ; F; // 计算结果,评估系统性能.
End

```

3.4 特征抽象

根据对触发词语义特征分析的结果,本文定义了 6 个不同的特征抽象层,将事件触发词的识别问题转换为特征分类问题.由于深度学习可以更加高效地挖掘数据特征,因此可以更加有效地实现事件识别.

词性抽象层(L1):词性是词最基本的语法特征.在句子中,语法成分对词性有较强的限制.例如名词、代词可以在句子中充当主语成分,但副词和连词则不可以.所以将词性作为抽象特征符合文本语义信息的表达特征以及人类在识别事件时的基本认知.通过对 CEC2.0 中触发词的词性统计分析,我们发现触发词的词性分布集中度非常高,主要以动词和名词为主,其中有 84% 是动词,14% 是名词.因此,将词性作为抽象特征层能够提高触发词识别的准确性.

依存语法抽象层(L2):依存语法认为句子中的述语动词是支配其他成分的中心,而它本身却不受其他任何成分的支配,所有的受支配成分都以某种依存关系从属于其支配者.依存语法直接描述词与词之间的语义角色关系,具有非常高的语义表现性.触发词作为句子中信息量最大、表达信息最清晰的词,在句子中所扮演的角色在一定意义上就如同依存语法中述语动词.通过对 CEC2.0 中的触发词依存语法的统计分析,我们发现 62% 的触发词在句子中的依存语法角色是 head,18% 为动词宾语.依存语法中的述语动词和句子中的触发词大多情况下是一致的.

长度抽象层(L3):词的长度与其在句子中扮演的角色之间的有着紧密联系.例如,在句子中主语和宾语通常有比谓语更大的长度,因为谓语主要用来表示动作以及状态,这类词通常比较简练且表达明确.通过对 CEC2.0 中的所有语料中触发词长度进行分析,81% 的触发词长度为 2,12% 的长度为 1.可见,触发词的长度特征集中度较高,非常适合作为识别触发词的特征.

触发词位置层(L4):索引是词在句子中的位置特征,该特征会影响词所要表达的语义信息.词在句子中的位置是人类在文字书写过程中的逐渐形成的一种规范,各种语言按照自身的语法逻辑来确定词在句子中的位置.所以词的索引对于判别词在句子中的角色具

有非常积极意义. 考虑到句子的长度不一样, 在本文中只使用 0 到 10, 共 11 个索引位置来标识该抽象层. 通过对 CEC2.0 中的所有语料文本中触发词的位置分析, 我们发现 16% 的触发词索引在句子的第一位, 12% 在句子的第二位. 由于分布的集中度不是非常高, 所以该特征对于触发词识别的贡献度相对较低.

与核心词的距离层(L5): 核心词主要指的是句子中主要成分的词, 通常用来表达一个句子最核心的内容, 而事件中的触发词同样用来表达句子中最重要的成分, 所以两者有很高的契合度. 一个词与核心词的距离越近, 那么它是触发词的可能性就越大. 在本文中核心词指的依存语法角色为 head 的词. 通过对 CEC2.0 语料的分析, 12% 的触发词与核心词的距离为 0, 27% 的触发词与核心词的距离为 1.

触发词词频抽象层(L6): 通过统计分析 CEC2.0 中的触发词, 我们发现语料中有大量的触发词重复出现. 有的触发词出现的频率较高, 有的触发词出现的频率较低. 本文将触发词频率分为 hp、a、b、c、d、e、f、g 八个等级, 等级的定义根据下表 1.

表 1 触发词词频等级规则表

序号	特征层	等级	定义
101	L6	hp	Frequency > = 1
102	L6	a	Frequency > 1 and Frequency < = 5
103	L6	b	Frequency > 5 and Frequency < = 10
104	L6	c	Frequency > 10 and Frequency < = 20
105	L6	d	Frequency > 20 and Frequency < = 30
106	L6	e	Frequency > 30 and Frequency < = 40
107	L6	f	Frequency > 40 and Frequency < = 50
108	L6	g	Frequency > 50

3.5 特征表示与样例生成

特征表示主要采用二值表示模型. 触发词词频特征向量值的生成方式是累加的方式, 其他 5 种抽象层的特征向量值生成方式相同. 下面分别描述各个抽象层特征表示方式以及特征向量值的生成方式.

词性抽象层特征表示: 词性特征向量共 29 个维度, 表示 29 种词性. 如果候选词的词性与特征向量中某个维度表示的词性一致, 那么该候选词在这维的特征向量值为 1, 其他 28 维特征向量值为 0.

依存语法抽象层特征表示: 在该特征层中, 向量维度为 14, 表示 14 种依存语法属性.

长度抽象层的特征表示: 在该特征层中, 向量维度为 10, 分别表示候选词的长度为 1 到 10. 如果长度超出 10, 那么该候选词的所有特征向量值为 0.

触发词位置层特征表示: 在该特征层中, 向量维度为

11, 11 个维度表示 0 到 10 共十一个索引位置. 如果位置索引超出 10, 那么该候选词的所有特征向量值为 0.

与核心词的距离层特征表示: 在该特征层中, 向量维度为 11, 分别表示候选词的到核心词的距离为 0 到 10. 如果距离超出 10, 那么该候选词的所有特征向量值为 0.

触发词词频抽象层特征表示: 在该特征层中, 向量维度为 8, 表示触发词的 8 个等级. 该特征层的向量表示采用的方式是累计. 例如, 候选词“地震”的等级 e, 那么它的特征向量值为“0 0 1 1 1 1 1 1”.

根据上面的特征表示方式, 我们将例句: “昨晚警方当场逮捕嫌疑犯”进行特征表示. 表示的结果如下表 2 所示. 样例中的类型 1 表示参与者, 2 表示时间, 3 表示触发词, 4 表示地点.

表 2 特征表示样例

候选词	特征抽象层	特征向量			类型
昨晚	L1 +	0000000000000000010000000000			2
	L2 + L3 +	00000010000000	0100000000		
	L4 + L5 + L6	1000000000	0001000000	00000000	
警方	L1 +	0000000000010000000000000000			1
	L2 + L3 +	10000000000000	0100000000		
	L4 + L5 + L6	0100000000	0010000000	00000000	
当场	L1 +	0001000000000000000000000000			4
	L2 + L3 +	00000010000000	0100000000		
	L4 + L5 + L6	0010000000	0100000000	00000000	
逮捕	L1 +	000000000000000000000000010000			3
	L2 + L3 +	00000000000001	0100000000		
	L4 + L5 + L6	0001000000	1000000000	00000011	
嫌疑犯	L1 +	0000000000010000000000000000			1
	L2 + L3 +	01000000000000	0010000000		
	L4 + L5 + L6	0000100000	0100000000	00000000	

4 试验结果

本文选取 CEC2.0 中文语料作为实验数据, 然后按 3.4 节给出的特征表示提取事件识别所需要的语义信息, 按照 3.5 节生成训练样例及测试样例, 分别交由 DL-ERM 进行学习、测试, 由深度分类器进行分类判断, 完成事件识别. 在实验中, 我们采用召回率(R)、准确率(P)和 F 系数来评价事件识别的结果, 同时使用 FI、FIR、CFI、CFIR 对两种分类器的识别效果进行对比.

4.1 无监督深度分类器的识别效果

我们将对不同特征组合后生成训练及测试样例, 分别利用 DBN_i(i 表示 DBN 包含的 RBM 层数, 分别取值为 1, 2, 3, 4, 5, 7, 8) 完成 2 组实验, 并进行比较. 第 1

组实验在相同的特征层情况下,采用不同 RBM 层数,主要验证增加 RBM 层数对事件识别性能的影响;第 2 组实验在使用的 RBM 层数相同的情况下,验证引入不同语义特征抽象层对事件识别性能的影响;

表 3 给出第 1 组实验的结果.从中可以看出:在一定范围内增加 RBM 的层数,模型的识别效果有较好的提升,原因在于模型通过多层映射单元提取出主要的特征信息,其精确度要优于单层结构.但是,在 RBM 层数大于 4 之后,模型的识别效果却随着层数的增加而有所降低,这说明模型的识别效果与 RBM 层数增长之间并非成正比关系,所以实际应用中需要通过大量的测试分析选择合适的 RBM 层数.同时,在本组实验中,随着 RBM 层数的增加,系统训练时间也随之增加.原因是层数增加后,用于完成训练的神经网络节点增加,组合训练工作量大幅增加,这对系统平台的计算负载提出了更高要求.

表 3 不同 RBM 层无监督的 DL-ERM 模型识别效果对比

模型 + 特征层	R	P	F	FI	FIR(%)
DBN1 + L[1-6]	80.41	78.51	79.45	0.00	0.00
DBN2 + L[1-6]	83.20	78.70	80.89	1.44	1.81
DBN3 + L[1-6]	84.70	82.80	83.74	2.85	3.52
DBN4 + L[1-6]	87.32	83.12	85.17	1.43	1.71
DBN5 + L[1-6]	86.13	83.21	84.64	-0.53	-0.62
DBN6 + L[1-6]	85.21	81.21	83.16	-1.48	-1.75
DBN7 + L[1-6]	83.13	81.32	82.22	-0.94	-1.13
DBN8 + L[1-6]	82.59	80.12	81.34	-0.88	-1.07

注: F_n 表示 RBM 层数为 n 时的 F 值,FI 为 F_n 与 F_{n-1} 之差,FIR 为 FI 与 F_{n-1} 比值.

第 2 组实验结果如表 4 所示.结果表明,特征抽象层的引入对总体识别效果有很好的提升,随着特征抽象层的增加,三项识别性能指标均为上升状态.但是不同的特征层对识别性能提升的贡献度并不一样,从表中可以看出依存语法特征层的加入使得识别效果提升最高.触发词位置层对识别效果提升有限.

表 4 不同特征层无监督的 DL-ERM 模型识别效果对比

模型 + 特征层	R	P	F	FI	FIR(%)
DBN4 + L1	78.13	66.32	71.74	0.00	0.00
DBN4 + L[1-2]	92.42	66.52	77.36	5.62	7.83
DBN4 + L[1-3]	83.87	77.42	80.52	3.16	4.08
DBN4 + L[1-4]	83.87	78.39	81.04	0.52	0.65
DBN4 + L[1-5]	80.01	86.32	83.05	2.01	2.48
DBN4 + L[1-6]	87.32	83.12	85.17	2.12	2.55

注: F_c 表示特征层数为 c 时的 F 值,FI 为 F_c 与 F_{c-1} 之差,FIR 为 FI 与 F_{c-1} 比值.

4.2 两种深度分类器的识别效果对比

为了比较两种深度分类器的识别效果,我们做了对比试验.测试 DBN 网络层数增加以及特征层增加两种情况下,两种分类器识别效果的差别.同时我们也对两者识别效果的稳定性以及训练时间做了对比分析.

表 5 不同 RBM 层混合监督与无监督分类器识别效果对比

模型	R	P	F	FI	FIR	CFI	CFIR
DBN1	80.41	78.52	79.45	0.00	0.00	0.00	0.00
DBN2	84.57	83.31	83.94	4.49	5.65	3.05	3.77
DBN3	89.60	84.72	87.09	3.15	3.75	3.35	4.00
DBN4	90.32	88.11	89.20	2.11	2.42	4.03	4.73
DBN5	90.30	87.21	88.73	-0.47	-0.53	4.09	4.83
DBN6	89.71	84.82	87.20	-1.53	-1.72	4.04	4.86
DBN7	89.11	84.23	86.60	-0.60	-0.69	4.38	5.33
DBN8	87.15	83.41	85.24	-1.36	-1.57	3.9	4.79

注: F_h 和 F_u 分别表示混合监督和无监督分类器的 F 值,CFI 为两者之差,CFIR 为 CFI 与 F_u 比值.

表 5 是在增加 RBM 网络层数的情况下,两种分类器识别效果比较.表中可以看出混合监督分类器三项指标的平均增量以及总的绝对增量均高于无监督的分类器, F 值最高增率达到 5.33%.由此可以得出结论增加混合监督学习过程可以在一定程度上提高识别效果.

下表 6 是在增加特征层数的情况下,两种分类器识别效果比较.从实验结果可以看出,语义特征层的增加对混合监督分类器的识别效果的提升带来了积极影响,其中依存语法特征层的效果最好,触发词位置层的效果最差这和无监督的分类器效果一致.同时从两者的对比分析可以看出混合监督的分类器识别效果好于无监督的分类器,最高对比增率能达到 6.82%.

表 6 不同特征层混合监督与无监督分类器识别效果对比

特征层	R	P	F	FI	FIR	CFI	CFIR
L1	75.11	78.21	76.63	0.00	0.00	4.89	6.82
L[1-2]	81.13	80.35	80.74	4.11	5.36	3.38	4.37
L[1-3]	83.21	80.81	81.99	1.25	1.55	1.47	1.83
L[1-4]	86.32	79.61	82.83	0.84	1.02	1.79	2.21
L[1-5]	91.76	80.28	85.64	2.81	3.39	2.59	3.12
L[1-6]	90.32	88.11	89.20	3.56	4.16	4.03	4.73

下图 5 是两种分类器识别效果稳定性的对比图,我们在确保两种分类器在相同特征层的情况下,通过不断增加 RBM 层数来分析两者的稳定性,从图中可以看出随着 RBM 层数的增加由于底层的误差不断向上传递,无监督的分类器识别性能逐渐下降,下降速度较快,幅度也较大,稳定性较差.而混合监督分类器的稳定性

要高于无监督的分类器。

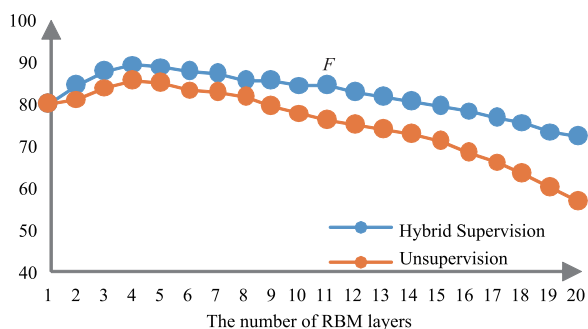


图5 两种分类器性能稳定性对比

表 7 两种分类器训练时间对比

测试条件	周期	无监督 (ms)	混合监督 (ms)	增长率 (%)
特征层保持不变(L[1-6]), 将 RBM 层数由 1 增加到 8	一个训练周期	1187	1493	25.78%
	总时间	9351	11709	25.22%
RBM 层数保持不变(4 层), 将特征层由 1 增加到 6 层	一个训练周期	1011	1289	27.50%
	总时间	6015	7665	27.43%

表 7 中给出了两种深度分类器实验消耗时间的对比。从表中可以看出混合监督的分类器训练时间较长,这是因为反向微调的过程会增加网络的训练时间,但是混合监督分类器带来的时间增长低于 30%,于该平台带来的识别效果增加相比,仍然在可以接受的范围内。无监督的分类器训练时间短,对平台的性能要求也较低。在实际应用中可以根据具体的要求以及平台性能来决定使用何种监督方式,如果对训练效率要求高可以选择无监督,如果对识别效果要求高可以选择混合监督分类器。

4.3 综合对比分析

为了验证 DL-ERM 模型的识别效果,我们将模型与已有的事件识别方法做了综合对比,见下表 8。从表中可以看出,本文提出的混合监督模型有一定的优势。与最好的基于规则的方法相比,三项评测指标的分别增加了 7.22%, 4.11%, 5.7%。本文提出的方法具有以下两点优势。第一:基于规则的方法需要人工定义规则,特别是针对不同领域的文本中的事件需要定义不同的规则。本文的方法并不需要预先定义规则,所以在处理不同领域文本的事件识别时,系统的识别效果变化不大。第二:基于规则的方法对规则的质量要求很高,往往需要领域专家来制定相关规则,但无法保证所有专家都能够提供一致性很高的规则,因此客观性和准确性难以保证。本文的方法主要是通过特征本身的统计规律

来进行分析识别,所以可以最大程度的保证系统的客观性和准确性。

表 8 DL-ERM 识别效果与现有事件识别方法效果对比

模型	方法分类	监督方式	R	P	F
统计 + 语义特征 ^[24]	基于规则	—	78	70	73
基于事件框架 ^[15]	基于规则	—	74.05	80.64	77.2
基于角色匹配 ^[25]	基于规则	—	83.1	84.0	83.5
依存语法 + SVM ^[18]	浅层机器学习	有	67.2	71.6	69.3
最大熵模型 ^[26]	浅层机器学习	有	62.6	80.7	70
频繁子树 + SVM ^[27]	浅层机器学习	有	84.2	78.5	81.3
DBN4 + L[1-6]	深度学习	无	87.32	83.12	85.17
DBN4 + L[1-6]	深度学习	混合	90.32	88.11	89.2

本文的混合监督模型与最好的基于浅层机器学习方法相比,三项评测指标分别增加 6.12%, 9.61%, 7.9%。其原因在于以下三个方面:(1)现有的浅层机器学习在处理小规模的样本数据时,能够取得很好的识别效果,但是随着数据量的增加,这些方法的运算时间以及识别能力都会受到极大的影响。而深度学习能够用来处理大量高维度的特征信息,因此系统的识别效果不受样本数据容量的影响。(2)现有的浅层机器学习,以 SVM 为例,是一种典型的二分类方法,在解决多分类问题时存在困难,虽然可以通过例如一对多组合模式、一对一组合模式和 SVM 决策树;或者构造多个分类器的组合来解决多分类问题,但是实际处理时非常困难,而且识别的准确率低同时系统可扩展性差。本文的混合监督深度信念网络是一种多分类器,可以轻松的完成多分类任务。(3)现有的浅层机器学习方法大多采用有监督的训练模式,这种模式能够提高识别效果,但是训练时间较长,本文的混合监督模型则融合无监督和有监督两种方法的共同优点,在实现识别性能提升的同时,系统的训练时间增长的幅度确有限。

5 结论

本文结合语义信息讨论了事件识别问题。通过统计分析 CEC2.0 语料库中事件触发词的语义特征,抽象出分层泛化的语义特征表示集,将事件触发词的识别问题转化为语义特征信息的分类问题;同时,针对现有的事件识别方法的局限性,设计了基于深度信念网络的事件识别模型(DL-ERM)。DL-ERM 模型以深度学习框架为基础,可以获得文本特征数据的深度特征,通过在突发事件语料库中的事件识别进行实验验证,证明了本模型能够提高识别效果。此外本文为 DL-ERM 模型设计了两种监督类型不同的深度分类器:无监督的深度分类器学习效率高,但是识别效果较差;混合监督的

深度分类器在改善识别效果的同时,也在一定程度上控制了训练时间。

后续研究中将会进一步提高 DL-ERM 模型的性能以及可扩展性。同时尝试通过该模型识别事件的其他要素,最终实现基于深度学习的事件及要素的自动化标注。此外,现有的深度学习中受限 RBM 的结构无法精确表达文本的向量结构,因此设计出符合文本特性的深度学习模型来更好的表达文本向量结构是下一步的理论研究重点。

参考文献

- [1] Yuan L. A part-of-speech tagging model employing word clustering and syntactic parsing[J]. Chinese Journal of Electronics, 2014, 23(1): 109 - 114.
- [2] Jing W, Liu Z, Hui Z. Micro-blogs entity recognition based on DSTCRF[J]. Chinese Journal of Electronics, 2014, 23(1): 147 - 150.
- [3] Liang M, Junping D U, Zhou Y. Cross-media hot topic auto-tracking model based on semantics and temporal context [J]. Chinese Journal of Electronics, 2015, 24(3): 529 - 534.
- [4] Yang H, Cao X. Linguistic steganalysis based on meta features and immune mechanism[J]. Chinese Journal of Electronics, 2010, 19(4): 661 - 666.
- [5] Muyun Yang, Sheng Li, Leilei Kong, et al. A hybrid model for microblog real-time filtering[J]. Chinese Journal of Electronics, 2016, 25(3): 432 - 440.
- [6] Wang X, Zha T, Chunming W U, et al. Text semantics based automatic summarization for chinese videos[J]. Chinese Journal of Electronics, 2015, 24(3): 462 - 467.
- [7] 胡舜耕, 刘晓宇, 钟义信. 基于多 Agent 技术的自动文摘系统的研究和设计[J]. 电子学报, 2001, 29(2): 247 - 249.
Hu Shungeng, Liu Xiaoyu, Zhong Yixin. Research and design of automatic abstracting systems based on multi agent technologies[J]. Acta Electronica Sinica, 2001, 29(2): 247 - 249. (in Chinese)
- [8] 杨思春, 戴新宇, 陈家骏. 面向开放域问答的问题分类技术研究进展[J]. 电子学报, 2015, 43(8): 1627 - 1636.
Yang Sichun, Dai Xinyu, Chen Jiajun. Advances in question classification for open-domain question answering[J]. Acta Electronica Sinica, 2015, 43(8): 1627 - 1636. (in Chinese)
- [9] 张志昌, 张宇, 刘挺, 等. 开放域问答技术研究进展[J]. 电子学报, 2009, 37(5): 1058 - 1069.
Zhang Zhichang, Zhang Yu, Liu Ting et al. Advances in open-domain question answering[J]. Acta Electronica Sinica, 2009, 37(5): 1058 - 1069. (in Chinese)
- [10] 韩立新, 陆桑璐, 谢立. 一个面向 Internet 的分布式信息检索系统模型[J]. 电子学报, 2002, 30(8): 1130 - 1133.
Han Lixin, Lu Sanglu, Xie Li. A model of distributed information retrieval systems for internet applications[J]. Acta Electronica Sinica, 2002, 30(8): 1130 - 1133. (in Chinese)
- [11] 林亚平, 刘云中, 周顺先, 等. 基于最大熵的隐马尔可夫模型文本信息抽取[J]. 电子学报, 2005, 33(2): 236 - 240.
Lin Yaping, Liu Yunzhong, Zhou Shunxian. Using hidden Markov model for text information extraction based on maximum entropy[J]. Acta Electronica Sinica, 2005, 33(2): 236 - 240. (in Chinese)
- [12] Hou J, Dongmei LI, Qiu C, et al. A semantic retrieval model based on domain ontology of orchard disease and pests[J]. Chinese Journal of Electronics, 2016, 25(3): 460 - 466.
- [13] Huang Q, Zhaofeng M A, Yang Y, et al. Secure data sharing and retrieval using attribute-based encryption in cloud-based OSNs [J]. Chinese Journal of Electronics, 2014, 23(3): 557 - 563.
- [14] 姜吉发. 一种事件信息抽取模式获取方法[J]. 计算机工程, 2005, 31(15): 96 - 98.
Jiang Jifa, An event IE pattern acquisition method[J]. Computer Engineering [J]. 2005, 31(15): 96 - 98. (in Chinese)
- [15] 梁晗, 陈群秀, 吴平博. 基于事件框架的信息抽取系统[J]. 中文信息学报, 2006, 20(2): 40 - 46.
Liang Han, Chen Qunxiu, Wu Pingbo. Information Extraction system based on event frame[J]. Journal of Chinese Information Processing, 2006, 20(2): 40 - 46. (in Chinese)
- [16] 冯礼. 基于事件框架的突发事件信息抽取[D]. 上海: 上海交通大学, 2008.
- [17] Chieu H L, Ng H T. A maximum entropy approach to information extraction from semi-structured and free text [A]. Eighteenth National Conference on Artificial Intelligence [C]. Edmonton: AAAI Press, 2002. 786 - 791.
- [18] 付剑锋, 刘宗田, 付雪峰, 等. 基于依存分析的事件识别[J]. 计算机科学, 2009, 36(11): 217 - 219.
Fu Jianfeng, Liu Zongtian, Fu Xuefeng, et al. Dependency parsing based event recognition [J]. Computer Science, 2009, 36(11): 217 - 219. (in Chinese)
- [19] Ahn D. The stages of event extraction[A]. Proceedings of the Workshop on Annotating and Reasoning about Time and Events [C]. Sydney: Association for Computational Linguistics, 2006. 1 - 8.
- [20] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527 - 1554.

- [21] Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: A deep learning approach [A]. Proceedings of the 28th International Conference on Machine Learning (ICML-11) [C]. Washington: ICML, 2011. 513 – 520.
- [22] 奚雪峰, 周国栋. 基于 Deep Learning 的代词指代消解 [J]. 北京大学学报: 自然科学版, 2014, 50(1): 100 – 110.
XI Xuefeng, Zhou Guodong. Pronoun resolution based on deep learning [J]. Acta Scientiarum Naturalium Universitatis, 2014, 50(1): 100 – 110. (in Chinese)
- [23] Wanxiang Che, Zhenghua Li, Ting Liu. LTP: a Chinese language technology platform [A]. Proceedings of the 23rd International Conference on Computational Linguistics; Demonstrations [C]. Beijing: CIPS, 2010. 13 – 16.
- [24] 李婷玉. 基于语义的文本事件信息抽取方法的研究与实现 [D]. 上海: 上海交通大学, 2012.
- [25] 吴刚. 基于主题的中文事件抽取技术研究及应用 [D]. 苏州: 苏州大学, 2009.
- [26] 谭红叶. 中文事件抽取关键技术研究 [D]. 哈尔滨: 哈尔滨工业大学, 2008.
- [27] 孟环建. 突发事件领域事件抽取技术的研究 [D]. 上海: 上海大学, 2015.

作者简介



张亚军 男, 1985 年 1 月生于安徽合肥, 博士研究生, 2014 年入学就读于上海大学计算机科学与工程学院, 主要研究领域为语义智能, 机器学习, 模式识别.

E-mail: zyj1985email@163.com



刘宗田 男, 1946 年 6 月生于山东莒南, 研究员, 博士生导师. 主要研究领域为程序证明和程序变换, 高级语言反编译, 软件质量管理, 数据挖掘、粗糙集合、概念格等.

E-mail: ztliu@shu.edu.cn



周文 女, 1979 年 2 月生于安徽合肥, 博士, 副教授. 主要研究方向为数据挖掘、知识流动与创新网络、事件本体和形式概念分析等.

E-mail: zhouwen@shu.edu.cn