

文章编号: 1003-0077(2018)06-0098-09

面向事件抽取的深度与主动联合学习方法

邱盈盈, 洪宇, 周文瑄, 姚建民, 朱巧明

(苏州大学 江苏省计算机信息处理重点实验室, 江苏 苏州 215006)

摘要: 事件抽取旨在从非结构化的文本中抽取事件的信息,并以结构化的形式予以呈现。监督学习作为基础的事件抽取方法往往受制于训练语料规模小、类别分布不平衡和质量参差不齐的问题。同时,传统基于特征工程的事件抽取方法往往会产生错误传递的问题,且特征工程较为复杂。为此,该文提出了一种联合深度学习和主动学习的事件抽取方法。该方法将 RNN 模型对触发词分类的置信度融入在主动学习的查询函数中,以此在主动学习过程中提高语料标注效率,进而提高实验的最终性能。实验结果显示,这一联合学习方法能够辅助事件抽取性能的提升,但也显示,联合模式仍有较高的提升空间,有待进一步思考和探索。

关键词: 事件抽取;深度学习;主动学习;循环神经网络

中图分类号: TP391 **文献标识码:** A

Combining Deep Learning and Active Learning for Event Extraction

QIU Yingying, HONG Yu, ZHOU Wenxuan, YAO Jianmin, ZHU Qiaoming

(Provincial Key Laboratory of Computer Information Processing Technology,
Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: Event extraction aims at extracting event information from raw texts and representing them as a structured text. As a basic event extraction method, supervised learning often suffers from small scale, imbalanced distribution and uneven quality of training corpus. Moreover, traditional event extraction methods based on feature engineering are complicated and will always cause error propagation. To address these issues, this paper presents a method to combine deep learning and active learning by the confidence of the query function based on RNN's trigger classification, in order to improve the quality and efficiency of corpus annotation as well as the ultimate performance. The experimental results show that this joint learning method can improve the event extraction, with substantial room for further exploration.

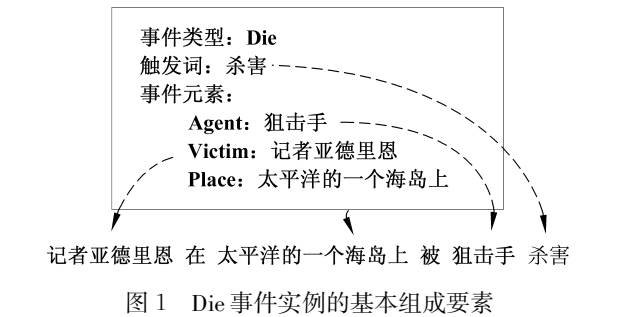
Key words: event extraction; deep learning; active learning; RNN

0 引言

根据 ACE 事件抽取的任务定义,事件抽取包含四个子任务:触发词识别(trigger identification)、事件类型分类(event type classification)、事件论元识别(argument identification)以及事件角色分类(argument role classification)。此外,也可以将这四个子任务合并为两个子任务:触发词分类(trigger classification)和事件论元分类(argument

classification)。图 1 详细表述了一个完整事件的结构。其中,“杀害”是该事件的触发词,触发的事件类型为 Die(死亡)。事件的三个论元“记者亚德里恩”“太平洋的一个海岛上”“狙击手”分别对应 Die 事件模板中的三个角色标签,即:Victim(受害者)、Place(发生地点)、Agent(主体)。简言之,事件抽取即是驱动机器对上述事件成分和类型进行判别、提取和形成结构化报表的过程。

基于监督学习的事件抽取方法存在数据稀疏和分布不平衡问题,往往导致机器学习模型训练出现



偏差,难以识别出未登录的事件信息。此外,传统的基于特征工程的事件抽取方法存在错误传递情况,且特征工程较为复杂。针对以上不足,本文提出了一种将深度学习和主动学习相结合的方法,实验表明,相较 Liao 等^[1]提出的基于 Co-Testing 主动学习的事件抽取方法,本文方法可以有效提高事件抽取的性能。

本文章节组织结构如下:第一节详细陈述本文动机;第二节介绍相关工作;第三节专注于联合深度学习与主动学习的事件抽取方法;第四节介绍实验设计,并给出实验结果和相应分析;第五节总结全文。

1 动机：主动学习与深度学习的潜在优势

和其它领域一样,面向事件抽取的监督学习也需要充分的标准实例作为机器学习的先验知识,但这类数据极为稀疏且质量难以保证。例如,现有的 ACE 2005 事件抽取语料包含 16 375 条句子,其中含有事件信息的句子仅有 3 966 条。此外,就质量而言,事件抽取语料存在标注不一致的情况,如下例所示。

例 1 *The outlines for the international stability force were decided at a conference.*^①

(译文:国际稳定军事力量的纲领在会议上得以确立)

例 2 *Campbell apologized to British Columbians in a tearful news at a conference.*^②

(译文:坎贝尔在新闻发布会上含泪向英国哥伦比亚人道歉)

其中,例 1 和例 2 均来自 ACE 2005 事件抽取语料,但只有例 1 中的“conference”被标注为触发词,其表示 Meet(举办会议)事件类型,而例 2 中的“conference”没有被标注为触发词,说明事件抽取语料存在标注不一致的情况。Hong 等^[2]对英文事件

抽取语料尝试人工标注,具体性能如表 1 所示。不同标注者分别对触发词分类任务进行人工标注,标注者 A 了解任务定义但不清楚事件抽取的标注规则,标注者 B 同时掌握任务定义和标注规则。实验结果表明,标注者对标注规则的掌握程度直接影响实验的最终性能。其次,从两者标注一致性的性能可以看出,不同标注者在进行事件抽取标注任务过程中,标注差异性很大,从而表明事件抽取任务的难度直接影响标注语料的规模和质量,进而影响事件抽取的最终性能。

表 1 人工标注事件抽取语料结果

人工标注结果	触发词分类		
	P/%	R/%	F ₁ /%
标注者 A(blind)	58.9	59.1	59.0
标注者 B(know)	74.3	76.2	75.2
两者标注一致性	48.9	42.3	45.3

解决上述问题的办法之一是引入主动学习,通过对监督学习过程中使用的样本进行主动筛选,选择信息量较大的样本进行人工标注,以提高语料的质量,进而提升学习效果。如下一节(相关工作)所述,Liao 等^[1]曾尝试利用主动学习对 Attack(表示攻击)事件类型进行学习与实际抽取。本文是对这一前瞻性研究的顺延,对所有 33 种 ACE 事件类型进行主动学习,并分析其适用性。

此外, Ji 等^[3]和 Hong 等^[2]提出的基于流水线的事件抽取方法(Pipeline)以及后期 Li 等^[4]提出的并行事件抽取模型^③,或存在错误传递的问题,或存在特征工程过于复杂的问题。比如, Pipeline 模型的一种流水作业是“先判断事件类型和触发词,后利用这一判定结果作为已知条件,对论元(Argument)和角色进行判定”。显然,如果类型判断失误,后续论元分类的过程将受到误导(即产生错误传递)。另一方面,并行模型对触发词、事件类型、论元和角色同时建模,每个模态的特征描述复杂性很高,因而对样本多样性和实例数量都要求极高,相对地,现有封闭域实践环境缺乏多样而充足的样本,这种需求与供给的反差,直接导致并行模型优势难以发挥的局

① 来自 ACE 2005 事件抽取语料中的文档 APW_ENG_20030502.0686.sgm

② 来自 ACE 2005 事件抽取语料中的文档 APW_ENG_20030502.0768.sgm

③ 并行抽取模型也称为联合模型 Joint model,为区别本文方法而采用并行模型的称谓。

面。值得注意的是,实际应用环境是一种开放域环境,对学习过程获得的模态的普适性要求较高。如果训练资源匮乏,相应监督学习方法的实用性将被大大削弱。

作为一种深度学习方法,循环神经网络(recurrent neural networks, RNN)具有较好的泛化能力,其能够通过对有限训练样例的学习,将样本所具有的抽象特征及相互关系(非线性关系)保存在权值矩阵中,从而,在实际工作阶段,即使被处理对象具有明显的“未登录”特性,也可利用上述特征关系给出间接推理,不影响系统的辨识能力。因此, RNN 能够有效抓取历史信息,获得和应用隐含在样本中的内在规律性和普适的特征。

综上,深度学习可以作为优化面向事件的监督学习的另一种途径。那么,主动学习和深度学习是否可以联用,如何联用,效果如何呢?是否可以对多种事件类型的样本公平地给予处理呢? 本文在 ACE 2005 数据集上进行相关实验,对上述问题给予解答。此外,在方法学方面,本文尝试利用 RNN 模型衡量触发词分类的不确定度,并将这一指标嵌入主动学习的查询函数中,借以高效地辅助事件抽取性能的提升。

2 相关工作

2.1 基于监督学习的事件抽取

Li 等^[4]在结合局部特征与全局特征的基础上,采用结构化感知机模型,将触发词分类和事件论元分类看成序列标注任务,从而能挖掘出事件触发词和事件角色间的内在依存关系,有效解决了传统串行结构事件抽取方法产生的错误传递问题。肖升等^[5]利用事件超图模型根据事件的属性及其结构计算相似度,并借此完成事件类型识别。Bronstein^[6]针对触发词分类任务,仅使用标注规范中每种事件类型的触发词种子集和少量的标注样例训练一个二元分类器,性能较 Li 等^[4]有所提高。Nguyen 等^[7]针对传统的基于特征工程的事件抽取方法耗时费力,以及 NLP 工具存在错误传递的情况,采用深度学习模型 CNN^[8](convolutional neural networks),仅使用实体类型和待测词及其上下文等特征,在触发词分类及领域适应性试验中,性能较传统的特征工程的抽取方法有所提高。Chen^[9]针对 Li 等^[4]在未登录词以及未登录特征的泛化能力方面的不足,

以及无法从人工构造的特征集中抽取出潜在的事件信息,提出了 DMCNN(dynamic multi-pooling convolutional neural networks)深度学习模型,此方法能够从训练集中自动抽取出更有效的特征,其性能较基于 CNN 的事件抽取方法有所提高。

2.2 基于直推式方法的事件抽取

直推式学习方法的核心思想在于利用已知的未标记事件信息,推测相同类型或类似事件中的未知事件信息。现有工作主要集中于跨句子(cross-sentence)、跨事件(cross-event)、跨文档(cross-document)和跨实体(cross-entity)事件抽取。Ji 等^[3]沿用 Yarowsky^[10]的“单篇章单语义”假设,将事件抽取的范围从单文档引申到话题相关的文档集中,通过使用基于规则的方法,将全局特征和局部特征相结合,在触发词分类和事件论元分类任务中,较句子级事件抽取方法获得了很大的性能提升。Gupta 等^[11]基于同类事件具有相同属性的假设,提出了跨事件推理方法。其借助语义特征,在同类事件间实现事件属性推理,有效解决了事件类型的事件论元缺失的问题。Liao 等^[12]提出了基于文档级别的跨事件推理方法,与前人工作不同的是, Liao 等^[12]并没有局限于事件所处的上下文环境,而是利用相关事件的信息及其事件类型的一致性对事件要素进行预测。该方法充分利用事件类型一致性这一关键特征,在事件抽取实验中性能较 Ji 等^[3]有所提高。Hong 等^[2]充分利用实体类型一致性特征,提出利用跨实体推理进行事件抽取,进一步提升了事件抽取的性能。

2.3 基于半监督和主动学习的事件抽取

Liao 等^[13]提出一种基于 Self-training 的半监督学习方法:对未标注语料进行人工标注,并在选择样例标注的过程中采用基于全局推理的方法考虑所选样例的多样性,有效解决了事件抽取语料规模较小且人工标注语料耗时费力的问题,实验结果较 Grishman 等^[14]有所提高。Liao 等^[1]针对事件抽取语料质量较低且人工标注耗时耗力问题,训练两个基于单词级和句子级粒度的分类器,并用基于 Co-Testing 的样例选择策略对所选样例进行人工标注,此外还考虑了样例的代表性和多样性。Li 等^[15]在 Liao 等^[16]提出的 WordNet 相似度和文档相似度扩充模板的基础上,充分考虑中文触发词的形态结构、事件角色的实体类型以及共指事件和相关事件

的关联性,有效提升事件抽取的性能。徐霞等^[17]针对中文本身在表述中存在的固有特点提出一种双视图的事件抽取自举学习方法,该方法以少量种子为基础,从文档相关度与语义相似度两个视图出发,进行交互过滤筛选,不断抽取新的有效事件模板,为事件抽取服务。

3 联合学习策略(深度学习+主动学习)

本文工作集中于事件抽取的前两个任务:触发词识别和事件类型分类。其中,本文提出的联合学习方法是指在主动学习的查询函数中,使用基于 Jordan-RNN^[18]模型的不确定度,通过 Jordan-RNN 在特征表示层面的优势,可以考察其在触发词分类任务上的不确定度对事件抽取性能的影响。以下对本文涉及到的一些模型和主动学习方法分别予以介绍。

3.1 基于特征工程的事件抽取方法

本文采用 MALLET^① 开源工具的最大熵(maximum entropy)模型分类器结合主动学习作为基准系统,分类器高斯参数设为 1.0。触发词分类任务不仅要求系统识别出触发事件的词或者短语,并且要求能正确地判断所触发的事件类型。表 2 描述了最大熵模型采用的各层面特征,主要包括词汇层面、依存句法层面以及实体层面特征,下面介绍各层面特征的用法。

表 2 触发词分类的特征

特征层面	触发词分类特征
词汇层面	(1) 待测词的词形
	(2) 待测词的词性
	(3) 待测词左/右 1 个词的词形
	(4) 待测词左/右 1 个词的词性
	(5) 待测词的同义词(WordNet)
依存句法层面	(6) 待测词的支配词和依赖词
	(7) 待测词的依存关系类型
实体层面	(8) 待测词的实体类型
	(9) 待测词左/右两个词的实体类型
	(10) 待测词的支配词和依赖词的实体类型
	(11) 与待测词在句子上最近的实体词及类型
	(12) 与待测词在依存句法上最近的实体词及类型

表 2 中的词汇层面特征主要考察待测词的词形、词性、同义词以及上下文特征。其中,待测词的同义词集指的是该词在 WordNet^[19]中具有最大概率的同义词集合,该特征能够弱化由于语料规模较小导致的特征稀疏问题,例如“move”的最常见意思为“移动”,相对应的同义词集为“travel”“go”“move”和“locomote”。依存句法层面特征主要考察待测词的依存类型、与待测词具有依存关系的词及词性等特征,以便在句法层面更好地表征触发词。实体层面主要考察待测词的实体类型以及待测词的上下文和依存关系词的实体类型。

3.2 基于 Co-Testing 主动学习的事件抽取方法

基于 Co-Testing 的主动学习方法目的是:在样例选择过程中,利用触发词级和句子级事件类型分类不确定度的差异性,选择信息量较大的句子进行事件信息标注。下面首先介绍基于触发词级事件类型分类不确定度的主动学习方法,然后介绍基于句子级事件类型分类不确定度的主动学习方法,最后介绍融合上述两种方法的基于 Co-Testing 的主动学习方法,并以此作为实验的基准系统。

3.2.1 基于触发词级事件类型分类不确定度的主动学习方法

基于触发词分类不确定度^[20]的主动学习事件抽取方法,目的是在样例选择标注过程中,挑选不确定度较大的样例进行人工标注。该方法认为:在包含事件信息的句子中,如果候选触发词的置信度越接近 0.5,则该句的不确定度越大。其中,候选触发词的置信度指的是该词在触发词分类任务中所触发最有可能事件类型的概率。例如,候选触发词“leave”的事件类型概率分布为 End-Position(表示离职)0.62、Transport(表示移动)0.35 以及 N/A(表示没有触发事件)0.03,则候选触发词“leave”的置信度为 0.62。样例选择标注过程中的查询函数(Query Function)如式(1)所示。

$$t_Info(S_i) = 1 - \min_{w_j \in S_i} |0.5 - prob_e(w_j)| \quad (1)$$

其中,prob_e(w_j)表示句子 S_i 中待测词 w_j 在触发词分类任务中的置信度。

3.2.2 基于句子级事件类型分类不确定度的主动学习方法

基于句子级事件类型分类不确定度的主动学习

① <http://mallet.cs.umass.edu/>

方法,考虑句子级事件类型分类不确定度。句子级事件类型分类使用的特征是句子去除停用词后剩余词的原形。训练过程中,一个句子可能包含多个事件,例如,既包含 Injure(表示受伤)也包含 Die(表示死亡)事件,这里采用多标签多标记学习^[21] (multi-label learning)方法,即在训练过程中将该句视为两个训练样例,并拥有两个事件类型。句子级的样例选择查询函数如式(2)所示。

$$s_Info(S_i) = 1 - |0.5 - prob_s(S_i)| \quad (2)$$

其中, $prob_s(S_i)$ 表示句子 S_i 在句子级事件类型分类任务中的置信度。

3.2.3 基于 Co-Testing 的主动学习方法

基于 Co-Testing 的主动学习方法在句子选择过程中,利用触发词级和句子级事件类型分类置信度的差异进行联合判定,以选择两个分类器结果不一致且不确定度大的句子进行标注。基于 Co-Testing 的主动学习方法的查询函数如式(3)所示。

$$co_Info(S_i) = \max_{w_j \in S_i \& isCP} |prob_e(w_j) - prob_s(S_i)| \quad (3)$$

其中, $prob_e(w_j)$ 表示句子 S_i 中待测词 w_j 在触发词分类任务中的置信度, $prob_s(S_i)$ 表示句子 S_i 在句子级事件类型分类任务中的置信度, isCP 指 w_j 在触发词分类任务中的事件类型与 S_i 在句子级事件类型分类任务的结果不一致。

3.3 联合深度学习与主动学习的事件抽取方法

3.3.1 Jordan-RNN

本文使用 Jordan-RNN 辅助事件抽取任务。Jordan-RNN 通过使用前一个状态的输出以及当前状态的单词的词向量作为输入,经过隐藏层计算得到当前状态的输出。模型仅采用单词的词向量作为特征,避免了特征抽取过程中出现的错误传递等问题,同时,模型能够通过前一状态的输出保存前文的信息,从而能够学习到隐含文本中的深层语义信息。过程如图 2 所示。

Jordan-RNN 模型使用 Google 的开源工具 Word2vec^① 训练得到词向量,将句子中的每一个单词 x_i 转化为一个 n 维的词向量作为 RNN 的输入层,通过网络计算出隐藏层状态 h_i (m 维)以及输出层状态 o_i (k 维)如式(4)~式(7)所示。

万方数据

$$h_i = W_{hx} x_i + W_{hh} o_{i-1} + b_h \quad (4)$$

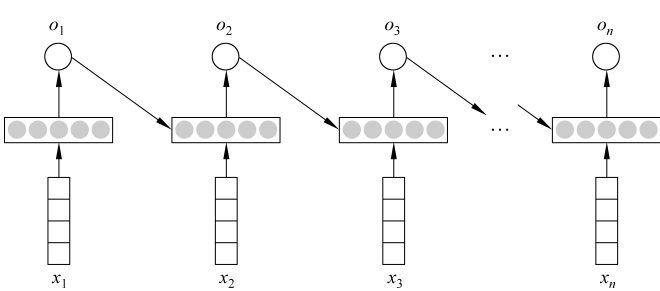


图 2 RNN 模型示意图

$$h_i = e(t_i) \quad (5)$$

$$s_i = W_{hy} h_i + b_k \quad (6)$$

$$o_i = g(s_i) \quad (7)$$

其中, W_{hx} 、 W_{hh} 、 W_{hy} 为权值矩阵, b_h 、 b_k 为偏置, t_i 和 s_i 分别为隐藏单元和输出单元的输入(k 维)。 e 和 g 是预定义的激活函数。此外, Jordan-RNN 还具备一特殊的初始偏置向量 $b_k^{init} \in R^m$ 表示第一次迭代时 $W_{hh} o_0$ 的值。

Jordan-RNN 关于训练样例 (x, y) 的目标函数定义为 $f(\theta) = L(\hat{y}; y)$, 其中, L 是衡量系统输出和目标输出之间偏差的损失函数, 具体形式为 $-\sum_i \sum_j \log(\hat{y}_{ij})$ 。

3.3.2 基于 Jordan-RNN 主动学习的事件抽取方法

由于 Liao 等^[1]提出的基于 Co-Testing 的主动学习方法建立在基于特征工程的事件抽取模型之上,而传统的基于特征工程的事件抽取方法无法从人工构造的特征集中有效抓取历史信息,且在特征抽取过程中, NLP 工具会存在错误传递的情况。因此,本节提出了基于 Jordan-RNN 模型的主动学习方法,目的是在样例选择标注过程中,将 Jordan-RNN 模型对触发词分类的置信度融入到主动学习的查询函数中,根据查询函数挑选不确定度较大的句子进行人工标注。基于 Jordan-RNN 主动学习方法的查询函数如式(8)所示。

$$J_Info(S_i) = 1 - \min_{w_j \in S_i} |0.5 - prob_J(w_j)| \quad (8)$$

其中, $prob_J(w_j)$ 表示句子 S_i 中待测词 w_j 在 Jordan-RNN 模型中的置信度。下面介绍基于 Jordan-RNN 模型主动学习中样例选择的算法流程:

① <https://code.google.com/p/word2vec>

算法 1 基于 Jordan-RNN 模型的主动学习方法

输入：

- S_i , 已标注数据集

- S_u , 未标注数据集

- n , 每次迭代向已标注数据集加入的样例数量

- T , 迭代次数

输出：

- J_{trg} , Jordan-RNN 触发词分类模型

迭代: $t = 1, \dots, T$

1. 根据已标注数据集 S_i 训练触发词分类模型 J_{trg}

2. 利用 J_{trg} 对未标注数据集 S_u 进行分类

3. 根据查询函数 $J_Info(S_i)$ 从 S_u 中选择 n 个不确定度最大的句子进行标注, 并记为 S_s

4. 将标注好的 S_s 加入到已标注数据集 S_i

5. 从未标注数据集 S_u 中去除标注好的 S_s

直到: 已经达到所需要标注的样例数量

算法 1 描述了基于 Jordan-RNN 模型的主动学习过程,该算法主体部分与基于最大熵模型的主动学习方法相同,不同之处在于 Jordan-RNN 模型采用基于词向量的特征表示方法,其能够充分利用历史数据的信息,从而利用基于 Jordan-RNN 的触发词分类置信度,能在同等语料规模的条件下,提高触发词分类任务的性能。

4 实验

4.1 实验数据与评价方法

本文针对事件抽取中的触发词识别和事件类型分类任务,分别选取 ACE 2005 的 599 篇英文文档作为实验语料,包含 16 375 个句子。为防止训练过拟合,实验采用 5 倍交叉验证,即每次随机选择 200 句作为主动学习的种子集、2 000 句为测试集,剩余的 14 175 句作为待标注数据集,主动学习每次迭代选取的标注样例数设为 200。此外,主动学习中的人工标注参考 ACE 2005 英文事件抽取语料以已标注好的结果,实验的基准系统基于 MALLET 的最大熵模型实现,深度学习模型采用 theano^① 工具实现的 Jordan-RNN。

触发词识别任务仅考察待测词是否触发事件,而不考虑待测词所触发的事件类型,即系统性能主要取决于被正确识别出的触发词的个数。事件类型分类任务要求触发词不仅要被正确识别,同时要求触发词被赋予正确的事件类型。系统采用在主动学习样例标注过程中不同阶段的触发词识别和事件类型分类的 F_1 值(F1-Measure)曲线来表示系统的性

能。此外,为便于用数值化形式表示主动学习方法的性能,系统也采用 Schein^[22] 提出的差异值(deficiency)用于区分不同模型的主动学习方法的性能,如式(9)所示。

$$\text{def}_n(A, B) = \frac{\sum_{i=1}^n (F_n(B) - F_i(A))}{\sum_{i=1}^n (F_n(B) - F_i(B))} \quad (9)$$

其中, n 表示主动学习的总迭代次数; F_i 表示触发词识别或事件类型分类在第 i 次迭代的 F_1 值; B 表示基准系统的主动学习方法, A 表示改进后的主动学习方法; 差异值(deficiency)越小,表示主动学习方法 A 相应于 B 的性能提高越大。此外,当 A 也表示基于基准系统的主动学习方法时, $\text{def}_n(B, B)$ 表示基准系统的主动学习方法相较于其自身的差异值,差异值为 1。

4.2 实验系统设置

为比较本文提出的联合深度学习与主动学习事件抽取方法以及 Liao 等^[1] 提出的基于 Co-Testing 主动学习的事件抽取方法性能,本文提出以下实验对比系统:

- Max_Ran: 使用最大熵模型,主动学习的样例选择策略采用随机选取;
- MaxEnt_T: 使用最大熵模型,主动学习的样例选择策略采用 3.2.1 节考虑触发词分类不确定度的主动学习方法;
- MaxEnt_S: 使用最大熵模型,采用 3.2.2 节仅考虑句子级事件类型分类不确定度的主动学习方法;
- MaxEnt_C: 使用最大熵模型,采用 3.2.3 节 Liao 等^[1] 提出的基于 Co-Testing 的主动学习方法;
- RNN_Ran: 使用 Jordan-RNN 模型,主动学习的样例选择策略采用随机选取;
- RNN_AL: 使用 Jordan-RNN 模型,主动学习的样例选择策略采用基于 RNN 的触发词分类不确定度。

其中,实验系统 Max_Ran 和 RNN_Ran 采用的样例选取策略为随机游走。具体的,样例选择过程中,两者都没有采取主动学习的策略,而是在每轮迭代过程中,随机选取固定迭代的样例数目进行人工标注,两者的不同之处在于实现事件抽取方法的模

① <http://www.deeplearning.net/software/theano/>

型不同,旨在考查不同的事件抽取模型对主动学习性能的影响。实验系统 MaxEnt_T、MaxEnt_S 以及 MaxEnt_C 旨在考查基于 Co-Testing 主动学习方法中考虑触发词分类不确定度时的性能。实验系统 RNN_AL 考察基于 RNN 的触发词分类不确定度对主动学习性能的影响。

4.3 实验结果与分析

图 3 给出实验系统 Max_Ran、MaxEnt_S、MaxEnt_T 以及 MaxEnt_C 主动学习中触发词识别和事件类型分类任务的性能曲线,图中的横轴表示主动学习中样例选择的迭代次数,纵轴表示触发词识别和事件类型分类任务在不同迭代次数中的 F_1 值。此外,为更清楚地以数值化的形式表示各实验系统的主动学习性能,表 3 给出了以系统 Max_Ran 作为基准系统的各实验系统的差异值,差异值越小说明实验系统较 Max_Ran 的性能越好。

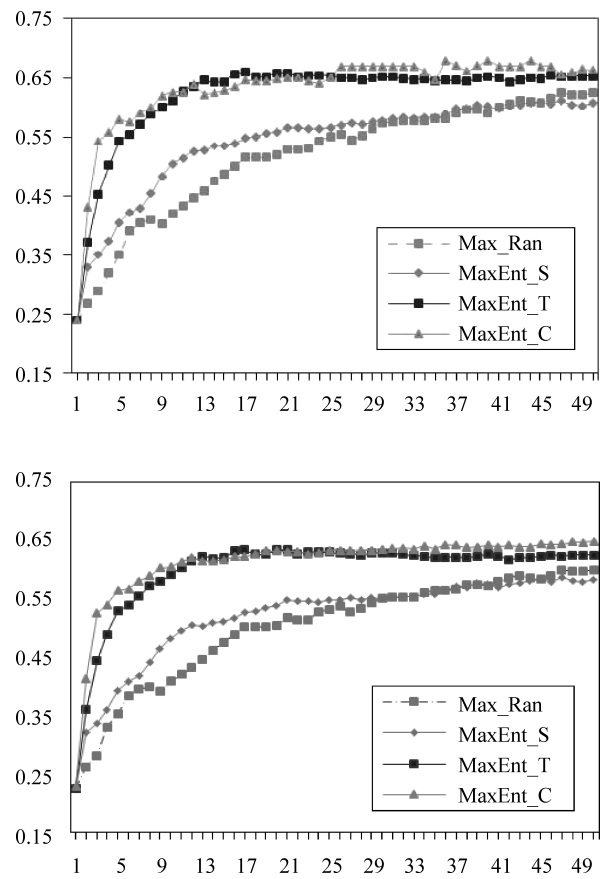


图 3 Max_Ran、MaxEnt_S、MaxEnt_T 以及 MaxEnt_C 的触发词识别和事件类型分类曲线

表 3 显示了各系统在全语料上相较于 Max_Ran 的差异值,实验系统 MaxEnt_T 优于 Max_Ran,表明基于最大熵的主动学习方法可以提高样

例选择标注的效率。原因在于,实验系统 Max_Ran 是随机选择样例进行标注,而实验系统 MaxEnt_T 通过主动学习方法选择触发词分类不确定度大的样例进行人工标注,从实验结果发现不确定度大的样例更可能包含事件信息,进而实验系统 MaxEnt_T 的差异值优于 Max_Ran。

表 3 实验系统差异值

实验系统	触发词识别	事件类型分类
Max_Ran	1.000	1.000
MaxEnt_S	0.769	0.812
MaxEnt_T	0.008	0.032
MaxEnt_C	-0.124	-0.105
RNN_Ran	-0.320	-0.247
RNN_AL	-0.547	-0.494

此外,实验系统 MaxEnt_S 劣于 MaxEnt_T,因为实验系统 MaxEnt_S 仅采用去除停用词后的一句话中的词包模型作为特征,然后判断该句的事件类型,而实验系统 MaxEnt_T 利用了待测词的词法、句法以及语义特征,可以较准确地识别出待测词的事件类型,进而导致实验系统 MaxEnt_T 的样例选择效率优于 MaxEnt_S;实验系统 Co-Testing 优于 MaxEnt_T 和 MaxEnt_S,说明样例选择过程中融合触发词级和句子级事件类型划分不确定度的 Co-Testing 主动学习方法,在 33 种事件类型分类任务上仍具有良好的效果。原因在于实验系统 MaxEnt_T 和 MaxEnt_S 采用不同粒度的特征进行事件类型分类,通过利用两个模型的差异性以及特征的冗余性,提高基于 Co-Testing 主动学习方法的样例选择效率。

图 4 描述实验系统 Max_Ran、RNN_Ran、MaxEnt_C 以及 RNN_AL 的性能曲线。其中实验系统 RNN_Ran 优于 Max_Ran,体现了 RNN 模型较于最大熵模型对于事件抽取性能的影响。原因在于 RNN 模型采用词向量的特征表示方法,进而减小了传统的基于特征工程的事件抽取方法中存在的错误传递的情况;实验系统 RNN_AL 优于 RNN_Ran,显示主动学习方法在 RNN 模型中可以提高样例选择的效率。原因在于,实验系统 RNN_Ran 随机选取样例进行标注,导致许多样例不包含事件信息,而实验系统 RNN_AL 通过主动学习方法选择基于 Jordan-RNN 模型的触发词分类不确定度大的样例进行标注,实验结果表明不确定度大的样例更有可

能包含事件信息,因而实验系统 RNN_AL 优于 RNN_Ran。最后,实验系统 RNN_AL 优于 Max-Ent_C,说明本文提出的基于 RNN 的主动学习方法较于基于 Co-Testing 主动学习方法,更适用于事件抽取任务。

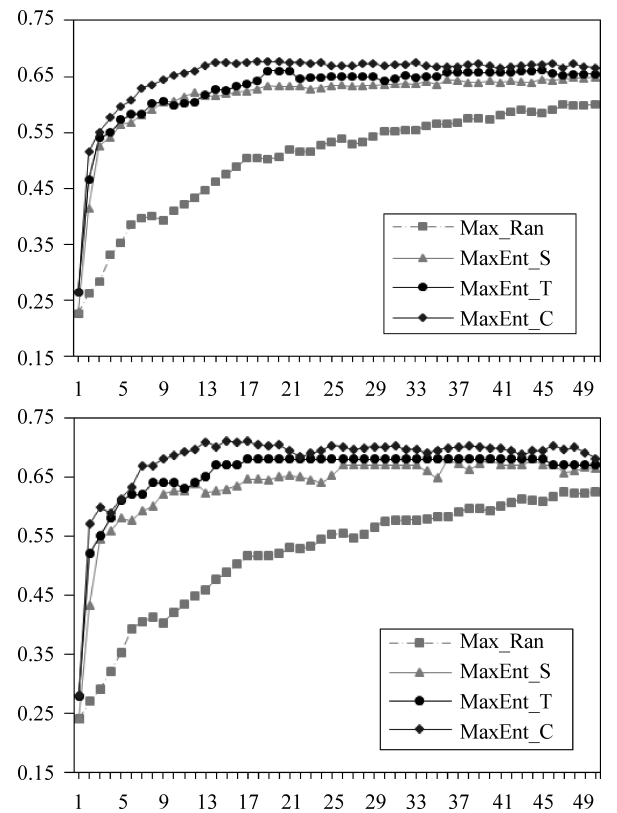


图4 Max_Ran、MaxEnt_C、RNN_Ran以及RNN_AL的触发词识别和事件类型分类曲线

本文针对触发词识别和事件类型识别的实验结果如表4所示。

实验系统	触发词识别/%			事件类型分类/%		
	P	R	F	P	R	F
Liao's(2011)	76.2	61.7	68.2	74.3	58.1	65.2
Li's(2013)	76.9	65.0	70.4	73.7	62.3	67.5
RNN	79.7	63.7	70.8	75.3	61.7	67.8
RNN_AL	82.3	62.9	71.3	77.4	61.3	68.4

在触发词识别任务上,实验系统 RNN 比 Liao 等^[1]和 Li 等^[4]的性能(F_1 值)分别高 2.6% 和 0.4%;在事件类型分类任务上,实验系统 RNN 比 Liao 等^[1]和 Li 等^[4]的性能分别高(F_1 值)2.6% 和 0.3%。验证了 Jordan-RNN 模型的有效性。原因在于 Liao 等^[1]提出的基于 Co-Testing 的主动学习

方法和 Li 等^[4]提出的并行抽取模型,都是建立在基于特征工程的事件抽取模型之上,而传统的基于特征工程的事件抽取方法,使用了丰富的人工设计的特征,且在特征抽取过程中,NLP 工具会存在错误传递的情况,同时普通的特征很难利用待测词的语义特征。而 Jordan-RNN 模型仅采用当前状态的词向量作为输入,且没有使用复杂的特征抽取工具,避免了传统方法非常耗时的特征选取过程,节省人力和时间。同时 Jordan-RNN 模型能够通过前一状态的输出保存前文的信息,从而能够学习到隐含文本中的深层语义信息。两种任务上,实验系统 RNN_AL 比 RNN 的性能(F_1 值)分别高 0.5% 和 0.6%,说明联合深度学习与主动学习的方法能够寻找有利于提升分类效果的样本,从而能提升样例的标注效率,使得能在较小规模的有效训练集上获得更好的性能。因此,本文提出的基于 Jordan-RNN 的主动学习的方法是有效的。

5 总结与展望

现有的基于监督学习的事件抽取方法受限于数据稀疏问题,导致学习算法无法抽取出来未登录事件信息。同时,使用监督学习方法进行事件抽取存在语料标注效率低下的问题。针对上述问题,本文引入了联合深度学习与主动学习的事件抽取方法。实验结果显示,基于 RNN 模型的主动学习方法优于 Liao 提出的基于 Co-Testing 的主动学习方法。

在未来工作中,本文将不再局限于基于不确定度的主动学习方法,而在样例选择过程中尝试基于多分类器投票的主动学习方法。此外,本文将尝试其他模型集成的方法,例如 Stacking 学习方法以及随机森林等,并对每个模型使用关联性较小的特征集合,以提高融合结果的准确率。

参考文献

[1] Shasha Liao, Ralph Grishman. Using prediction from sentential scope to build a Pseudo co-testing learner for event extraction[C]//Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), Chiang Mai, Thailand, 2011: 714-722.

[2] Yu Hong, Jianfeng Zhang, Bin Ma, et al. Using cross-entity inference to improve event extraction [C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL), Portland, USA, 2010: 1127-1136.

[3] Heng Ji, Ralph Grishman. Refining event extraction

- through cross-document inference[C]//Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics(ACL), Columbus, USA, 2008; 254-262.
- [4] Qi Li, Heng Ji, Liang Huang. Joint event extraction via structured prediction with global features [C]//Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics(ACL). Sofia, Bulgaria, 2013; 73-82.
- [5] 肖升,何炎祥. 事件超图模型及类型识别[J]. 中文信息学报, 2013, 27(01): 30-38.
- [6] Ofer Bronstein, Ido Dagan, Qi Li, et al. Seed-based event trigger labeling: How far can event descriptions get us? [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP), Beijing, China, 2015; 372-376.
- [7] Thien Huu Nguyen, Ralph Grishman. Event detection and domain adaptation with convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP), Beijing, China, 2015; 365-371.
- [8] Hu B, Lu Z, et al. Convolutional neural network architectures for matching natural language sentences [C]//Proceedings of the Advances in Neural Information Processing Systems (NIPS), Quebec, Canada, 2014, 2042-2050.
- [9] Yubo Chen, Liheng Xu, Kang Liu, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP), Beijing, China, 2015; 167-176.
- [10] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods[C]//Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL). Cambridge, US, 1995; 189-196.
- [11] Prashant Gupta, Heng Ji. Predicting unknown time arguments based on cross-event propagation [C]//Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Suntec, Singapore, 2009; 369-372.
- [12] Shasha Liao, Ralph Grishman. Using document level cross-event inference to improve event extraction [C]//Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL), Uppsala, Sweden, 2010; 789-797.
- [13] Shasha Liao, Ralph Grishman. Can document selection help semi-supervised learning? A case study on event extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics(ACL), Oregon, Portland, 2011; 260-265.
- [14] Ralph Grishman, David Westbrook, Adam Meyers. NYU's English ACE 2005 system description [C]//Proceedings of ACE 2005 Evaluation Workshop, Gaithersburg, USA, 2005; 5-19.
- [15] Peifeng Li, Qiaoming Zhu, Guodong Zhou. Employing event inference to improve semi-supervised chinese event extraction [C]//Proceedings of COLING 2014 and the 25th International Conference on Computational Linguistics (Coling), Dublin, Ireland, 2014; 2161-2171.
- [16] Shasha Liao, Ralph Grishman. Filtered ranking for bootstrapping in event extraction[C]//Proceedings of the 23rd International Conference on Computational Linguistics(Coling), Beijing, China, 2010; 680-688.
- [17] 徐霞,李培峰,朱巧明. 一个半监督的中文事件抽取方法[J]. 中文信息学报, 2016, 30(02): 168-174.
- [18] Mesnil G, He X, Deng L, et al. Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding [C]//Proceedings of Interspeech (Interspeech), Lyon, France, 2013; 3771-3775.
- [19] George A M. WordNet: A lexical database for English [J]. Communications of the ACM, 1995, 38(11); 39-41.
- [20] Lewis, J Catlett. Heterogeneous uncertainty sampling for supervised learning[C]//Proceedings of the International Conference on Machine Learning (ICML). Morgan Kaufmann, 1994, 148-156.
- [21] Minling Zhang, Zhihua Zhou. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [22] Schein A I, Ungar L H. Active learning for logistic regression: an evaluation [J]. Machine Learning, 2007, 68(3): 235-265.



邱盈盈(1993—),硕士研究生,主要研究领域为自然语言处理、信息抽取。

E-mail: yinyinqiu@gmail.com



洪宇(1978—),博士,副教授,主要研究领域为话题检测、信息检索和信息抽取。

E-mail: tianxianet@gmail.com



周文瑄(1991—),硕士研究生,主要研究领域为自然语言处理、信息抽取。

E-mail: chrisnotkris7@gmail.com