

基于 Bi-LSTM 的医疗事件识别研究*

侯伟涛, 姬东鸿

(武汉大学 计算机学院, 武汉 430072)

摘要: 文本的信息抽取及属性识别是文本语义分析的一项重要任务。在已有工作基础上提出了双向 LSTM 与 MLP 集成的深度神经网络模型。在 2016 年 SemEval 的医疗事件抽取以及事件属性预测任务中, 该模型将医疗文本的词性以及命名实体的描述信息当做附加属性, 使用双向 LSTM 神经网络学习文本的隐藏特征, 解决了传统方法通用性不强以及无法捕捉前后文隐含信息的缺点, 再使用全连接的方式去判断候选词汇是否属于医疗事件以及识别其相关属性。实验结果表明, 提出的神经网络模型对医疗文本的抽取效果优于以往学者的方法。

关键词: 信息抽取; 医疗事件识别; 深度学习; 双向 LSTM

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1001-3695(2018)07-1974-04

doi:10.3969/j.issn.1001-3695.2018.07.011

Research on clinic event recognition based Bi-LSTM

Hou Weitao, Ji Donghong

(School of Computer, Wuhan University, Wuhan 430072, China)

Abstract: The entity recognition and information extraction is an important part of text semantic analysis. For this reason this paper proposed a deep learning model utilized by Bi-LSTM as well as MLP based in previous work. To finish the task of clinic event extraction and attribute prediction from 2016 SemEval, the model employed the context words together with their part-of-speech tags and named entities to compose features. Then it utilized the bidirectional LSTM neural network to learn the hidden feature representations. In prediction step, this paper used softMax to predict event spans and its attributes. The empirical evaluation demonstrates that the approach significantly outperforms baseline methods.

Key words: information extraction; clinical event recognition; deep learning; Bi-LSTM

本文致力于提高非结构化医疗文本中医疗事件的自动识别以及该医疗事件的分类效果。本文的目标即识别非结构化医疗文本中的事件对象, 以及该事件对象的四种属性。这四种属性分别是类型(type)、极性(polarity)、程度(degree)以及物理疗法(modality)。实例如表1所示。其中医疗文本指医生的临床文本记录; 医疗事件是指其中与疾病相关的事件, 为短语或者词汇; 事件属性指该医疗事件的相关描述信息。

表1 医疗文本以及事件示例

医疗文本	医疗事件	事件属性
April 23, 2014: The patient did not have any postoperative bleeding so we will resume chemotherapy with a larger bolus on Friday even if there is slight nausea.	bleeding	type = N/A polarity = NEG degree = N/A modality = ACTUAL
	resume	type = ASPECTUAL polarity = POS degree = N/A modality = ACTUAL
	chemother-apy	type = ASPECTUAL polarity = POS degree = N/A modality = ACTUAL
	bolus	type = ASPECTUAL polarity = POS degree = N/A modality = ACTUAL
	nausea	type = ASPECTUAL polarity = POS degree = N/A modality = HYPOTHETICAL

近年来有很多学者投入了医疗事件识别领域的研究, 取得了不错的效果。经过调研发现, 前人比较常见的流程是先利用分词工具处理原始文本, 然后利用 CRF、SVM 等结合大量人工规则来进行医疗事件的抽取以及医疗事件属性的预测工作。不过以往工作仍然存在不足之处, 如模型无法充分地获取隐含信息; CRF、SVM 规则并不能捕捉长距离上下文的隐含信息; 人工

规则的创建与具体领域有较大关系, 存在通用性不强的局限性等。因此, 本文提出了一个基于 LSTM 神经网络的端对端的模型来进行医疗事件的抽取以及属性预测, 可以更好地获取隐含关系, 并且具有更强的通用性。经过本文实验证实, 基于神经网络的模型具有比传统方法更好的事件提取效果, 事件预测的准确率和召回率都有所提升。

1 相关工作

针对本文涉及到的医疗事件识别任务, 前人已经做了一定的工作, Caselli 等人^[1]使用了 Stanford CoreNLP tool 来分词并得到词性特征, 然后加入多个人工规则, 训练五个相互独立的 CRF 模型来对候选事件进行预测。Abdulsalan 等人^[2]使用了医疗领域的分词工具 cTAKES 来分词, 然后使用 BIO 法训练五个 CRF 模型对结果进行预测。Lee 等人^[3]则使用了更多的分词工具 (CLAMP、OpenNLP、ClearNLP、cTAKES 等) 以及更多的规则来提升预测的效果。通过之前学者的研究可以发现, 以往的工作^[4-8]已经在医疗事件抽取中取得了一定的成绩, 但他们使用的大多是基于规则机器学习的方法, 在应用于本文任务时这些方法存在以下缺点: a) 传统方法使用多个分类器, 把多个分类问题割裂开来, 费时费力; b) 基于人类感官经验依赖语言学知识加入不同的规则, 较为繁琐复杂; c) 传统机器学习方法无法刻画单词前后长距离依赖; d) 模型需要借助医学领域分词工具来提升分词效果, 通用性较差。

相对于以 CRF 为代表传统机器学习方法, 深度学习有着独特的优势。深度学习模型可以充分逼近任意复杂的非线性关系, 而且具有很强的鲁棒性、记忆能力、非线性映射能力, 以及强大的自学习能力。得益于此, 深度学习近年来在自然语言处理领域如序列标注、关系分类等得到了越来越广泛的应用。2010 年 Mikolov 等人^[9]对递归神经网络 (recursive NN) 进行了

收稿日期: 2017-02-27; 修回日期: 2017-04-26 基金项目: 国家自然科学基金面上项目 (61373108)

作者简介: 侯伟涛 (1991-), 男, 山东东营人, 硕士, 主要研究方向为自然语言处理 (houwt@whu.edu.cn); 姬东鸿 (1967-), 男, 教授, 博导, 博士, 主要研究方向为自然语言处理、语义网技术、机器学习、数据挖掘等。

开创性的改进,增强了语言模型的特征能力,对深度学习的发展影响巨大;2011年 Collobert 等人^[10]将深度学习模型应用到自然语言处理的词性标注、分词以及命名实体识别等工作中,取得了卓有成效的效果;2012年 Socher 等人^[11]把递归神经网络用于自然语言处理领域的文本语义研究。有学者对 RNN 进行改进,得到更加高效稳定的 LSTM 模型。Hammerton^[12]将 LSTM 成功应用到了命名实体识别任务中;2015年 Huang 等人^[13]利用 Bi-LSTM 来获取输入元素的隐含信息,并将 Bi-LSTM 的输出信息作为 CRF 的输入,借此来实现对文本内容的序列标注;除此之外,2016年 Ma 等人^[14]将词语包含的所有字符作为 CNN 的输入来获得单词的表示,并结合词向量一起作为 Bi-LSTM 的输入,再将 Bi-LSTM 的输入作为 CRF 的输出,最后得到序列标注的结果。这种方法适用于序列标注问题的场景,实验效果出色。通过以上众多学者的工作可以看出,LSTM 作为深度学习的代表之一,在自然语言处理问题上取得了令人信服的效果^[15,16],是切实可行的。但是具体应用到本文任务时,以上基于深度学习的方法都有相似的不足,即本文任务需要对单个元素进行五个维度的分类,这些维度可能会有这内在的关联,而以上模型都是针对单个分类问题,想要解决多维度的分类问题,必须要分别训练五个相互独立的模型,这样做不仅费时费力,而且考虑到不同维度的属性可能有着某种内部关联,粗暴地使用多个分类器可能会带来信息的丢失问题。

综上,过去学者在解决本文涉及的任务时大都停留在传统的机器学习方式,基于 Bi-LSTM 在自然语言处理领域上相对于传统机器学习方式的优越性,本文决定以 Bi-LSTM 为基础来构建基本框架,解决医疗事件的识别任务;而又基于以往工作中对于多维度分类问题的不足,将对其他学者的模型进行改进,提出一个将五种维度属性特征同时训练的方法,端到端的一次性得到待检测元素是否是医疗实体,以及当该元素是医疗实体时该元素的其他四种维度 type、polarity、degree、modality 的信息,将标注和分类同时进行,从而避免单独训练可能造成的信息丢失问题。

2 模型构建

2.1 LSTM 简介

循环神经网络(recurrent neural network, RNN)通过添加跨越时间点的自连接隐藏层,可以赋予神经网络对时间进行显式建模的能力。在 RNN 中,隐藏层的反馈不仅会进入输出端,而且还会进入下一时刻的隐藏层。不过 RNN 在实际应用中却有着梯度消失和梯度爆炸的情况,因此引入了长短期记忆网络(long short term memory network, LSTM)。LSTM 是一种特殊的 RNN,能够学习到长期依赖关系,以解决梯度消失和梯度爆炸的问题。LSTM 最初由 Hochreiter 和 Schmidhuber 提出,之后许多研究者针对 LSTM 进行了一系列的优化改进工作,LSTM 得以快速发展,现已被广泛应用于自然语言处理领域的各个方面。每个 LSTM 单元都有三个门,用于控制是否将当前信息忽略或者传递给下一个单元。每个 LSTM 单元如图 1 所示。

LSTM 单元的状态更新可由以下公式得到:

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (1)$$

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (5)$$

$$h_t = o_t \odot \tan(c_t) \quad (6)$$

其中: σ 是点对的 sigmoid 函数; \odot 是点对乘积; x_t 是 t 时刻的输入向量(如词向量); h_t 是 t 时刻以及之前时刻存储了所有有用信息的隐含状态向量; $U_i, U_f, U_c, U_o, W_i, W_f, W_c, W_o$ 是隐含状态的权重矩阵; b_i, b_f, b_c, b_o 表示对应权重偏置。

LSTM 的特性是使当前单元得到本单元之前所有单元的信息,其缺点是无法获得本单元之后单元的信息,因此双向 LSTM 神经网络(bi-directional LSTM, Bi-LSTM)便应运而生,其

模型示意图如图 2 所示。双向 LSTM 神经网络是对普通 LSTM 的进一步改进,在序列标注任务上有着出色的效果。其基本思想是分别用前向和后向的 LSTM 来抓取过去和将来所含的隐藏信息,这两部分的信息组成最终的输出。

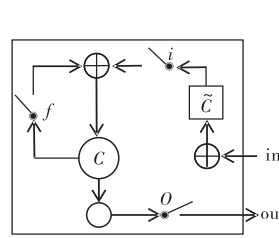


图1 LSTM单元示意图

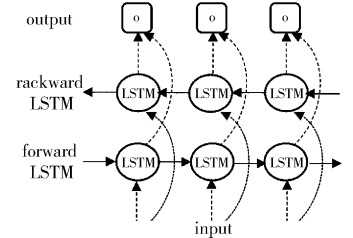


图2 Bi-LSTM单元示意图

2.2 Bi-LSTM-based model I

为了抽取句中医疗事件之间的关系,本文使用基于序列特征和双向 LSTM 模型来学习事件以及事件上下文的表示,如图 3 所示。

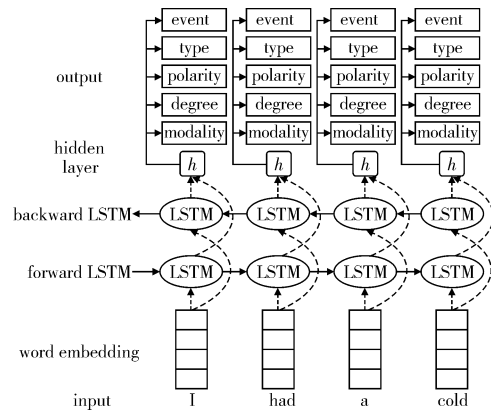


图3 Bi-LSTM-based model I示意图

模型最底层是输入层。将一段医疗文本输入模型,分词之后得到每个单词的词向量信息作为神经网络的输入。经过前向、反向两个 LSTM 深度学习模型对词向量信息学习得到隐含层 h ,其中每个 LSTM 节点对应了一个 h ;最后对 h 同时进行五个 softmax 处理,得到需要的五种分类的预测结果。

该网络模型在 i 时刻的输入为

$$X_i = V_{\text{word}}$$

其中: V_{word} 是词向量。模型在 i 时刻模型的输出(output)为

$$t_{e_i} = f(W_e h_i + b_e) \quad (7)$$

$$t_{i_i} = f(W_i h_i + b_i) \quad (8)$$

$$t_{p_i} = f(W_p h_i + b_p) \quad (9)$$

$$t_{d_i} = f(W_d h_i + b_d) \quad (10)$$

$$t_{m_i} = f(W_m h_i + b_m) \quad (11)$$

其中: f 是 softmax 函数; t_{e_i} 是指在 i 时刻模型预测事件(event)的值; W_e 以及 b_e 是隐含状态对应于 event 的映射矩阵和偏置; t_{i_i} 是指在 i 时刻模型预测类型(type)的值; W_i 以及 b_i 是隐含状态对应于 type 的映射矩阵和偏置; t_{p_i} 是指在 i 时刻模型预测极性(polarity)的值; W_p 以及 b_p 是隐含状态对应于 polarity 的映射矩阵和偏置; t_{d_i} 是指在 i 时刻模型预测病情程度(degree)的值; W_d 以及 b_d 是隐含状态对应于 degree 的映射矩阵和偏置; t_{m_i} 是指在 i 时刻模型预测物理疗法(modality)的值; W_m 以及 b_m 是隐含状态对应于 modality 的映射矩阵和偏置。

这样,对于每个候选词,模型都会对应生成五个预测结果,分别对应了 event、type、polarity、degree、modality 五种属性值。

本模型对应的损失函数为

$$\text{cost} = L(e_i, t_{e_i}) + L(t_i, t_{i_i}) + L(p_i, t_{p_i}) + L(d_i, t_{d_i}) + L(m_i, t_{m_i}) \quad (12)$$

其中: L 是交叉熵函数; e_i 是指 event 在 i 时刻的真实值; t_{e_i} 是指

event 在 i 时刻的预测值; t_i 是指 type 在 i 时刻的真实值; t_{t_i} 是指 type 在 i 时刻的预测值; p_i 是指 polarity 在 i 时刻的真实值; t_{p_i} 是指 polarity 在 i 时刻的预测值; d_i 是指 degree 在 i 时刻的真实值; t_{d_i} 是指 degree 在 i 时刻的预测值; m_i 是指 modality 在 i 时刻的真实值; t_{m_i} 是指 modality 在 i 时刻的预测值。

该损失函数综合了 event、type、polarity、degree、modality 五种属性,因此训练后的模型可反映出这五种属性的内在关联信息,以便更好地作出预测

2.3 Bi-LSTM-based model II

第一个模型用 GloVe200 维度的词向量来表示每个单词,并不能很好地表示出单词本身的其他属性信息。因此,本文在第一个模型的基础上又添加了单词的属性信息,如词性以及命名实体特征,以完善单词的表示。这两个特征是通过 Stanford CoreNLP 来获取的。词性特征 POS 和命名实体特征 NE 的加入,结合词向量一起,可以更好地刻画出深层次的语义信息,使得预测结果更加准确。引入单词属性特征后的新模型如图4所示。

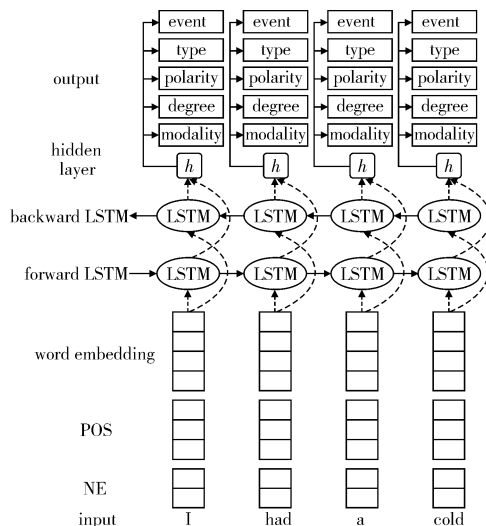


图4 Bi-LSTM-based model II 示意图

模型最底层是输入层。将一段医疗文本输入模型,与 Bi-LSTM-based model I 不同的是分词之后得到每个单词的词向量以及该单词的词性特征 POS 和命名实体 NE 特征结合在一起作为本模型的输入。这个模型下 i 时刻的输入。

$$X_i = [V_{\text{word}}; V_{\text{pos}}; V_{\text{ne}}]$$

其中: V_{word} 是词向量; V_{pos} 表示特征向量化后的词性特征; V_{ne} 表示特征向量化后的命名实体特征; $[\]$ 表示将其中的向量级联。这样输入 X_i 不仅包含了词语本身的信息,也将词性特征 POS 和命名实体特征 NE 融合进去,使得输入的语义信息更加丰富,组合后的词向量可以更准确、全面地表达单词本身,进而对模型有优化作用。

此外,由于不同属性对目标损失的影响不同,本模型在第一个模型的基础上对损失函数进行了改进,给损失函数中的五个部分分别添加了损失权重,其他部分与模型 I 保持一致。

$$\text{cost} = \alpha_e \times L(e_i, t_{e_i}) + \alpha_t \times L(t_i, t_{t_i}) + \alpha_p \times L(p_i, t_{p_i}) + \alpha_d \times L(d_i, t_{d_i}) + \alpha_m \times L(m_i, t_{m_i}) \quad (13)$$

3 实验与分析

3.1 数据

2016 医疗信息抽取任务 (<http://alt.qcri.org/semeval2016/task12/>) 致力于从医疗笔记中抽取医疗事件信息,其中本任务包含了六个子任务:a) 时间抽取 (spans of time expressions); b) 事件抽取 (spans of event mentions); c) 时间属性抽取 (attributes of times); d) 事件属性抽取 (attribute of events); e) 事件与文档创建时间的时序性预测 (events' temporal relations to the docu-

ment creation times); f) 时间事件关系 (narrative container relations among events and times)。本文完成了其中第二个和第四个子任务,专注于医疗事件相关的任务。

本文用到的数据是 Mayo clinic (<http://www.mayoclinic.org/>) 的医疗笔记、癌症病例报到以及该文件的人工标注数据,共计 600 个文件。数据集分为训练集、验证集、测试集三部分,分别包含 297、150、153 个文件。训练集、验证集、测试集三者的统计信息如表 2 所示,具体到医疗事件属性的数目分布统计则如表 3~6 所示。

表2 数据集医疗事件统计

数据集	候选词	医疗事件	非重复候选词	非重复医疗事件
训练集	203 090	32 421	8 970	3 647
验证集	108 499	17 757	6 569	2 673
测试集	114 975	18 459	6 751	2 762
总计	426 564	68 637	13 205	5 242

表 2 中,候选词是指医疗文本分词之后每个词语单元的数目;医疗事件是指医疗文本中出现的医疗事件的数目;非重复候选词是指候选词去充值后的候选词种类的数目;非重复医疗事件是指医疗事件去充值后的医疗事件种类。

表3 医疗事件的属性

type 分布统计			
type	N/A	ASPECTUAL	EVIDENTIAL
训练集	30 239	459	1 723
验证集	16 489	183	1 085
测试集	17 276	209	974

表4 医疗事件的属性

degree 分布统计			
degree	N/A	LITTLE	MOST
训练集	32 239	111	71
验证集	17 675	41	41
测试集	18 373	53	33

表5 医疗事件的属性 polarity 分布统计

polarity	POS	NEG
训练集	29 430	2 991
验证集	16 150	1 607
测试集	16 835	1 624

表6 医疗事件的属性 modality 分布统计

modality	HEDGED	GENERIC	HYPOTHETICAL	ACTUAL
训练集	695	466	1 377	29 883
验证集	339	275	662	16 481
测试集	422	254	864	16 919

3.2 实验设置

本文首先使用 Stanford CoreNLP tool (<http://stanfordnlp.github.io/CoreNLP/>) 去处理原始文本,获取原始文本分词后的单词以及单词的词性信息、命名实体信息;然后借助 GloVe (<http://nlp.stanford.edu/projects/glove/>) 得到每个单词 200 维度的词向量信息;再对词性以及命名实体信息词袋模型的方法进行编码,分别得到 44 维和 13 维的向量,将这两者与词向量拼接到一起,组成新的词向量;最后使用 Tensorflow (<http://www.tensorflow.org/>) 来建模以及处理数据,模型的学习率为 0.02。经过 LSTM 训练后的模型将每一个单词映射到一个 256 维的矩阵中,即隐含层神经元的数量为 256。对隐含层进行五次运算得到该单词的五个属性特征。考虑到对候选词是否是 event 的判断最为重要(本文采用的策略是如果该候选词不是 event,则舍弃之后四种属性的预测结果),因此提高 event 的权重。模型 2 中各个权重依次为 0.4、0.15、0.15、0.15、0.15。最后采用随机梯度下降算法对神经网络进行训练。

特别说明的是,模型中每个 LSTM 单元都生成一个隐含层,而本文实验都是以文档为单位进行学习和预测,这就带来一个问题,即各个文档的长度(文档所含的单词数目)是不同的。经过统计,最长的文档单词个数不超过 4 000,因此本文以 4 000 为固定的训练步长,不足 4 000 的部分使用循环填充的方法来补全,这样便解决了文档长度不统一的问题。

本次实验以 CRF 作为基准线。CRF 作为传统机器学习工具,广泛地应用于分词、词性标注和命名实体识别等序列标注任务中,具有出色的效果。将本文模型与 CRF 对比可以很好地体现模型的优劣。表 7 列出了本文使用的全部特征组合。

表7 CRF 模型使用的特征组合

特征组合	特征组合
U00: % x[-2,0]	U05: % x[-2,0]/% x[-1,0]
U01: % x[-1,0]	U06: % x[-1,0]/% x[0,0]
U02: % x[0,0]	U07: % x[0,0]/% x[1,0]
U03: % x[1,0]	U08: % x[1,0]/% x[2,0]
U04: % x[2,0]	U09: % x[-1,0]/% x[1,0]

3.3 实验结果以及结果分析

本文采用准确率和召回率来评价模型的优劣。准确率和召回率是广泛用于信息检索和统计学分类领域的两个度量值,用来评价结果的质量。其中精度是指检索出的相关文档数与文档总数的比率,衡量的是检索系统的查准率;召回率是指检索出的相关文档数和文档库中所有的相关文档数的比率,衡量的是检索系统的查全率。

实验结果如图 5、6 所示。从图 5、6 中可以看出,对于本任务,Bi-LSTM-model-I 相对于以 CRF 为代表的传统机器学习方法在每个医疗事件属性的判断上都有一定的提升,而加入了更多特征的 Bi-LSTM-model-II 相对于未改进的方法在事件属性的预测中有了进一步的性能提升。与传统方法相比,本文模型可以更好地学习长距离文本的语义特征,因此与传统方法相比有较大优势,而文本附加特征的加入可以强化模型的预测性能。

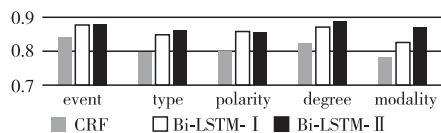


图5 各个模型准确率统计

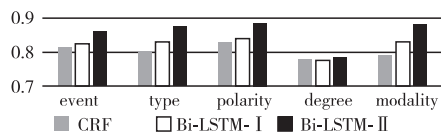


图6 各个模型的召回率统计

3.4 误差分析

本文所用的数据集中,有一小部分的医疗事件是由多个单词构成。据统计,总计有 86 种医疗事件是由多个单词构成,占所有医疗事件的 1.64%。本文针对的是单个单词医疗事件的预测,所以这部分内容是本文模型无法作出预测的。除此之外,数据集的标注文件是 Mayo Clinic 的工作人员人工标注,因此难免会有疏漏错误的情况出现。根据本文统计,梅奥医院给出的标注中出错的标注有 38 种(标注的医疗事件末尾包含空格或者标点符号),占所有医疗事件的 0.72%。这部分内容也一定程度上影响了模型的学习。剩下的误差都是模型还不够完美,无法模拟真实情况带来的误差。为了增强模型的通用性,下一步的工作会解决多词语事件无法预测的缺陷。眼下笔者认为在当前模型的基础之上,以当前端对端模型学习到的隐含层作为 CRF 的输入,利用 BIO 法,这样理论上既可以同时抓取到不同维度间属性的关联,又可以实现对多词语医疗事件的预测。

4 结束语

本文提出了一种基于 Bi-LSTM 的医疗事件抽取以及属性识别的端到端模型,这个模型不仅避免了传统机器学习方法通用性不够的缺点,而且能够避免对多属性问题单独分类造成的信息丢失问题。经过实验,本文模型在 Mayo Clinic 提供的语料上取得了很好的效果。与传统机器学习方法相比,本文的模型方法有以下优势:不依赖特定领域的自然语言处理工具,可以推广到其他领域,有较强的通用性;把医疗事件以及事件属

性的识别问题整体化,使得模型可以站在全局角度对事件进行预测;LSTM 模型的引入解决了传统方法无法刻画单词前后长距离依赖的问题;更重要的是本文的模型方法可以更好地实现对医疗事件的抽取识别工作。实验表明,本文模型对事件预测的准确率和召回率都超过了传统方法。

本文特征的选取针对的是英文医疗文本,对于中文医疗文本的医疗事件抽取则需要选择针对中文的特征。此外,特征设计方面或许可以不局限于单词级别的特征,如寻找语句级别的语义。为了达到更好的医疗事件识别效果和实际应用,以上是本文的下一步努力方向。

参考文献:

- [1] Caselli T, Morante R. VUACLTL at SemEval 2016 task 12: a CRF pipeline to clinical TempEval [C]//Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1241-1247.
- [2] Abdulsalam A A, Velupillai S, Meystre S. UtahBMI at SemEval-2016 task 12: extracting temporal information from clinical text [C]//Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1256-1262.
- [3] Lee H J, Xu Hua, Wang Jingqi, et al. UTHHealth at SemEval-2016 task 12: an end-to-end system for temporal information extraction from clinical notes [C]//Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1292-1297.
- [4] Cohan A, Meurer K, Goharian N. GUIR at SemEval-2016 task 12: temporal information processing for clinical narratives [C]//Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1248-1255.
- [5] Sarath P R, Manikandan R, Niwa Y. Hitachi at SemEval-2016 task 12: a hybrid approach for temporal information extraction from clinical notes [C]//Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1231-1236.
- [6] Li Ping, Huang Heng. UTA DLNLP at SemEval-2016 task 12: deep learning based natural language processing system for clinical information identification from clinical notes and pathology reports [C]//Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1268-1273.
- [7] Barros M A, Lamurias A, Figueiro G, et al. ULISBOA at SemEval-2016 task 12: extraction of temporal expressions, clinical events and relations using IBEnt [C]//Proc of the 9th International Workshop on Semantic Evaluation. 2016: 1263-1267.
- [8] Tourille J, Ferret O, Névél A, et al. LIMSI-COT at SemEval-2016 task 12: temporal relation identification using a pipeline of classifiers [C]//Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1136-1142.
- [9] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model [C]//Proc of the 11th Annual Conference of the International Speech Communication Association. 2010: 1045-1048.
- [10] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch [J]. *Journal of Machine Learning Research*, 2011, 12(1): 2493-2537.
- [11] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces [C]//Proc of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012: 1201-1211.
- [12] Hammerton J. Named entity recognition with long short-term memory [C]//Proc of the 7th Conference on Natural Language Learning at HltNAACL. 2003: 172-175.
- [13] Huang Zhiheng, Xu Wei, Yu Kai. Bidirectional LSTM-CRF models for sequence tagging [J]. *arXiv:1508.01991*, 2015.
- [14] Ma Xuezhe, Hovy E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [J]. *Eprint arXiv, 1603.01354*, 2016.
- [15] Graves A. Supervised sequence labelling with recurrent neural networks [M]. Berlin: Springer, 2012.
- [16] 胡新辰. 基于 LSTM 的语义关系分类研究 [D]. 哈尔滨: 哈尔滨工业大学, 2015.