

基于深度学习的竞争情报命名实体识别研究*

马千程 王崑声 周晓纪

(中国航天系统科学与工程研究院 北京 100089)

摘 要: [目的/意义]旨在为命名实体识别技术在竞争情报中的应用提供参考。[方法/过程]将基于深度学习的命名实体识别方法应用在竞争情报工作中,在构建 BiLSTM-CRF 神经网络模型的基础上,加入注意力层,帮助模型更好地聚焦局部特征。[结果/结论]实验结果表明,深度学习算法在竞争情报命名实体识别中是高效的,同时与单一 CRF 模型和传统 LSTM+CRF 模型进行实验对比,证明改进后的模型在识别效果上有很大提升。

关键词: 竞争情报;命名实体识别;双向长短期记忆网络;条件随机场;注意力机制

中图分类号: TP391

文献标识码: A

doi: 10.3969/j.issn.1005-8095.2020.09.001

Named Entity Recognition of Competitive Intelligence Based on Deep Learning

Ma Qiancheng Wang Kunsheng Zhou Xiaojie

(China Academy of Space System Science and Engineering, Beijing 100089)

Abstract [Purpose/significance]The paper intends to provide references for the application of named entity recognition technology in competitive intelligence. [Method/process] The paper applies named entity recognition in competitive intelligence based on deep learning, adds an attention layer to help the model focus on local features better on the basis of BiLSTM-CRF neural network model. [Result/conclusion]The results indicate that the deep learning algorithm is efficient in the identification of named entity recognition in competitive intelligence. Compared with the single CRF model and the traditional LSTM + CRF model, the deep learning algorithm has better recognition effect.

Keywords: competitive intelligence; named entity recognition; BiLSTM; CRF; attention mechanism

0 引言

竞争情报是现代企业生存的重要因素,作为一种特殊的战略资源,其对企业竞争战略和风险的预警价值,需要通过对海量信息的实时采集和有效分析来实现^[1]。随着大数据时代的到来,我们生活以及理解世界的方式正发生深刻改变,如商业组织和社会组织的运行模式发生了巨大变化^[2]。企业面对的数据呈现出爆炸式急剧增长^[3],传统的情报分析方法和处理模式远远无法满足企业的需要,企业迫切需要引入新理论、新技术来改善现状。自然语言处理技术可实现人与计算机的有效通信,同时可对各种形式化信息进行分析与处理,是实现海量信息高效分析的必由之路^[4]。

自然语言处理是人工智能领域的重要方向之

一,人与计算机之间采用自然语言技术可实现有效通信。命名实体识别(Name Entity Recognition, NER)最早由 MUC-6 会议提出^[5],是自然语言处理的基本任务之一,旨在从文本中识别出人名、地名和组织机构名等实体。命名实体识别是问答系统^[6]、句法分析^[7]、机器翻译^[8]等研究领域的重要理论基础,其研究具有重要的意义。命名实体识别的方法目前主要有三大类:基于规则的方法、基于统计机器学习的方法和基于深度神经网络的方法。

早期学者的命名实体识别研究采用基于规则和基于统计机器学习的方法,主要利用隐马尔科夫模型^[9]、支持向量机模型^[10]、最大熵模型^[11]和条件随机场模型^[12]等来训练相关特征命名实体。程志刚^[13]采用基于规则和条件随机场相结合的命名实

收稿日期: 2020-04-16

* 本文系国家自然科学基金委 2040 工程科技项目管理与知识系统项目“情报数据平台”(项目编号: L1724033)研究成果之一。
作者简介: 马千程(1995—),男,2017 级硕士研究生,研究方向为自然语言处理;王崑声(1960—),男,硕士,研究员,博士生导师,研究方向为系统工程。

体识别模型对《人民日报》的语料进行人名、地名、机构名识别;邱泉清等^[14]设计特征模板后,使用条件随机场模型对微博中的语料进行命名实体识别。基于规则的方法需要较强的领域知识,导致领域迁移性差,并且人工消耗较大。基于统计机器学习的方法性能严重依赖人工设计的特征数量,领域的泛化性较差,同时随着特征的增多,容易在模型训练时间过长后导致过拟合。近年来,随着人工智能技术的快速发展,基于神经网络的深度学习方法备受关注,LSTM-CRF模型成为当前命名实体识别最流行的方法。L.Liu等^[15]用BiLSTM提取字符级特征后对词向量进行预训练,再通过BiLSTM+CRF的组合模型识别出实体;车金立等^[16]将字位向量融合作为输入,在BiLSTM-CRF模型中进行训练,完成对军事领域的命名实体识别;张俊飞等^[17]利用Word2Vec训练《人民日报》语料,并通过Jieba分词获得字的信息特征向量,将组合向量输入到BiLSTM-CRF模型中进行训练,得到面向中文领域命名实体识别模型;买买提阿依甫等^[18]通过CNN来获取维吾尔单词字符特征,应用BiLSTM-CRF模型实现对维吾尔命名实体的识别,F1值达到91.9%。本文在现有研究的基础上,结合所需处理的竞争情报文本特点,在LSTM-CRF模型中加入注意力层,改进后的模型有效提高了识别效果。

1 相关技术及原理

1.1 长短期记忆网络

长短期记忆网络(LSTM)属于一种特殊的循环神经网络(RNN),RNN忽略了长距离的依赖关系,导致循环神经网络会丧失学习到远距离信息的能力^[19]。LSTM对RNN进行改进,针对需要保存长期信息这一问题,增加了长期记忆单元C,针对梯度消失问题,增加了门限来控制状态的更新和传递,一个t时刻的LSTM计算单元如图1所示。

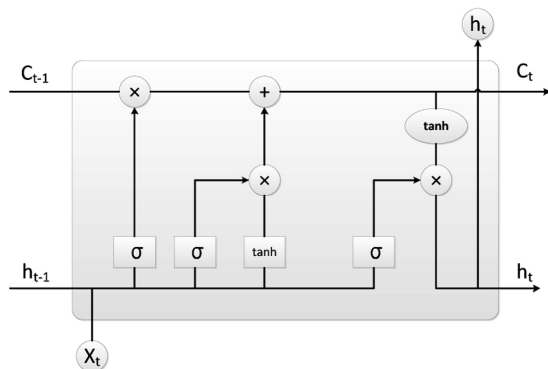


图1 LSTM计算单元

一个LSTM的计算单元包括遗忘门 f 、输入门 o 、记忆单元 c 和输出 h 。

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

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (2)$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

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

其中 x_t 是t时刻的神经网络输入向量, f_t 是t时刻 f 的输出向量, i_t 是 o 在t时刻的输出向量, ρ_t 是输出门在t时刻的输出向量, \tilde{c}_t 是t时刻记忆单元中的候选单元, c_t 是t时刻的记忆单元, h_t 是t时刻LSTM的输出向量, W_f 、 W_o 和 W_c 是不同门和记忆单元在 h_{t-1} 上的参数矩阵, U_f 、 U_i 、 U_o 和 U_c 是不同门和记忆单元在 x_t 上的参数矩阵, b_f 、 b_i 、 b_o 和 b_c 是不同门和记忆单元的偏置向量, σ 表示sigmoid激活函数, \odot 表示相乘。

记忆单元、遗忘门、输入门、输出门是LSTM的核心和关键,记忆单元将当前和历史信息进行保存,遗忘门控制记忆单元丢弃哪些信息,输入门控制记忆单元加入哪些信息,输出门控制将记忆单元中哪些信息进行输出。遗忘门、输入门和输出门通过Sigmoid激活函数,将每个门的输出向量值控制在0到1之间。表面上看,LSTM的三个门可保留长短时的记忆,让LSTM具备学习到远距离信息的能力。理论上看,使用sigmoid激活函数可有效解决三个门的梯度消失和爆炸问题,让长短期记忆网络在处理长序列时达到很好的训练效果。

1.2 条件随机场

条件随机场(Conditional Random Field,CRF)模型是一种概率无向图模型,最早由John Lafferty等人^[20]在隐马尔科夫模型和最大熵模型的基础上提出。相空间内所有位置按照某种条件随机分布,全部被赋予值后统称为随机场,当随机场中某个位置的赋值只与其相邻位置词相关时,称之为马尔科夫随机场。假设 x 和 y 都是一组随机变量, $P(y|x)$ 表示给定输入 x 集合的情况下,变量 y 的条件概率分布,同时输出变量 y 构成一个马尔科夫随机场,那 $P(y|x)$ 的条件概率分布称为条件随机场。

图2所示是“链式”条件随机场,CRF在序列标注问题中,一般就是指“链式”条件随机场。条件随机场会定义一组势函数,通常选用指数势函数,用于描述概率分布^[21]。对条件概率 $P(\vec{y}|\vec{x})$ 按图2结构

进行建模。

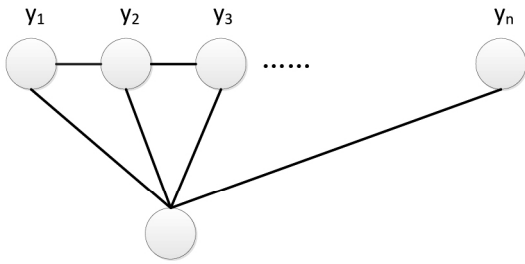


图2 链式条件随机场示意图

$$P(\vec{y}|\vec{x}) = \frac{1}{Z} \exp \left(\sum_{j=1}^{n-1} \lambda_j t_j(y_{i+1}, y_i, \vec{x}, i) + \sum_k \sum_{i=1}^n \mu_k s_k(y_{i+1}, \vec{x}, i) \right) \quad (7)$$

其中: $t_j(y_{i+1}, y_i, \vec{x}, i)$ 是转移特征函数, 相邻标记间的相关关系通过转移特征函数来描述, 同时可以观测到对相邻标记的影响; $s_k(y_{i+1}, \vec{x}, i)$ 是状态特征函数, 标记变量的变化通过状态特征函数来展现。

2 竞争情报命名实体识别模型

本文提出的基于深度学习的命名实体识别模型实际应用在企业竞争情报系统中, 竞争情报工作常常面临所需处理文本篇幅不定、文本夹带很多无关网络用词和图片等情况, 传统的NER模型针对这一特殊情况识别效果受到极大影响, 本文对传统NER模型进行优化调整。

2.1 融合 BiLSTM 的算法模型

传统NER模型使用单向LSTM结构, 单向

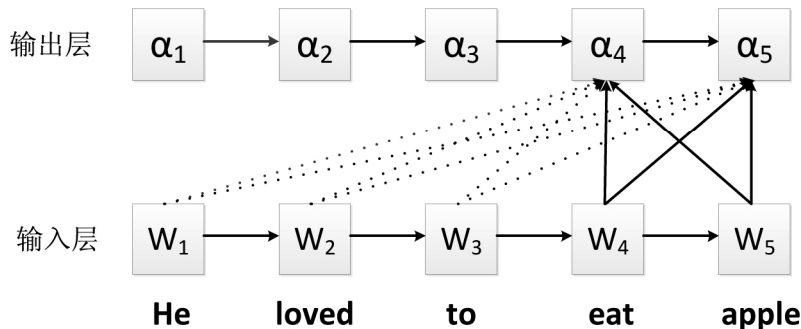


图3 注意力模型

在为“apple”打标签时, 注意力模型会体现出该单词周围的单词对其影响程度的大小, 给出下面一个概率分布值:

(He 0.1) (loved 0.15) (to 0.05) (eat 0.3) (apple 0.4)。

上面的概率分布值代表注意力模型给出的各个单词注意力值, 有效展示了注意力模型聚焦的区域。结合BiLSTM, 合成整个句子中间语义的变换函数, 如公式8所示:

LSTM结构只能记录t时间步之前的输入, 无法获取未来时间步的信息。当遇到文本篇幅短小的情况时, 模型需要更有效地抓住仅有的特征信息, 兼顾上下文语境, 才能更有效地捕捉到特征。

双向LSTM结构(BiLSTM)可以有效解决这一问题, BiLSTM由向前和向后的两个单向LSTM组成, 两个网络结构中间用一个输出层进行连接。前向LSTM将数据通过输入层进入神经网络结构, 按照正常计算与传递方式在输出层得到训练结果。反向LSTM是指在训练过程中, 神经网络将误差逐层传递至输入层, 并根据误差对每一层的网络参数进行更新。双向LSTM模型同时考虑过去和未来时刻的序列信息, 实现完整记录每一个时间步未来与过去信息的目标, 当文本篇幅短小时, 预测的结果也能相对准确。

2.2 注意力机制

深度学习的发展, 注意力模型(Attention Model)受到更多关注, 目前广泛应用在不同的深度学习领域中。本文引入Bahdanau等人^[22]提出的注意力模型, 帮助NER模型更好地聚焦局部特征, 在极小篇幅中抓住文本重点。

引入注意力模型, 模型将重点关注打标签单词附近的其他单词, 而适当忽略距离较远或无关的单词信息, 如图3所示。

$$C_i = \sum_{j=1}^{L_x} a_{ij} h_j \quad (8)$$

注意力模型的当前状态 C_i 需要通过输入句子的长度 L_x 、注意力系数 a_{ij} 和第 j 个单词的状态值 h_j 共同决定。注意力模型的更新由注意力系数决定, 输出项分给输入项的注意力越多, 其对应的 a_{ij} 数值就越大。

2.3 模型总体设计

针对竞争情报所需处理文本的特殊性, 在LSTM与CRF结合的基础上, 提出一种基于双向长短期记

忆网络与条件随机场相结合的NER方法,整个模型架构如图4所示。

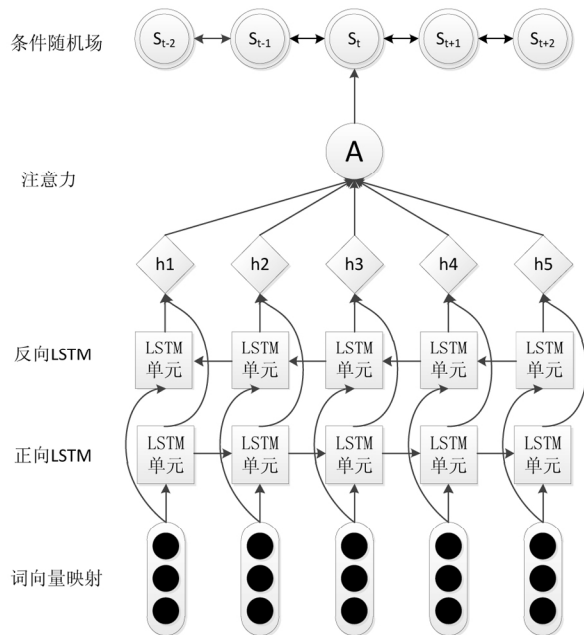


图4 本文命名实体识别模型结构图

第一层是词嵌入层,需要将输入的单词序列训练成词向量,本文采用 Word2Vec 中的 CBOW 模型进行词向量训练,CBOW 模型通过对上下文分析确定每个词的位置,输出每个单词的词向量作为下一层神经网络输入的各个时间步。

第二层,将词向量输入到 BiLSTM 层中进行训练,学习单词与输出标签的关系。双向 LSTM 网络会通过正向 LSTM 以及反向 LSTM 得到相对应的隐含输出序列,将其拼接组成每一个时刻的完整隐含序列,作为下一层的输入。

第三层,在 BiLSTM 层的输出序列上引入注意力

机制,进一步处理标签问题,使模型更好聚焦局部特征,突出关键词的重要作用。模型使用 RNN 对上一层的输出序列进行解码,注意力模型的当前状态 C_i 会影响隐藏层状态 z_i :

$$z_i = f(z_{i-1}, y_{i-1}, C_i) \quad (9)$$

其中 f 表示非线性的激活函数, C_i 表示文本中的关键词。

上文提到的注意力系数 a_{ij} 又称为感知机, BiLSTM 生成的隐藏层 h_j 值就是通过感知机 a_{ij} 来测量与输出标签的位置 i 的关系。隐藏层不仅包含了文本全局信息,还包含文本的局部关键词信息,通过加权求和得到当前时间步的输出状态。接着还需要进行线性转换,使其与标签维度相对应,再经过 softmax 算法得到最后的输出向量。为了换取较高精度,本模型中采用的注意力模型由加法模型组成。

第四层,在注意力机制后使用 CRF 层,通过转移矩阵输出标签之间的转移得分,考虑每个标签的转换规律以及标签语法的合理性,使用维特比解码得到最佳标签序列,输出最佳的解决方案。

3 实验

本文将命名实体识别技术应用于面向商用飞机制造领域的竞争情报工作中,对商用飞机竞争情报信息中机型、技术两个重点关注的文本进行命名实体识别,帮助企业高效挖掘有价值的信息。

3.1 实验语料

本文抓取的竞争情报源主要是针对互联网上与该行业相关的网页信息,设定的竞争情报网页源见表1。

表1 竞争情报网页源

编号	名称	网址
1	波音官网新闻	http://boeing.mediaroom.com/news-releases-statements
2	空客官网新闻	http://www.airbus.com/search.news.html
3	庞巴迪官网新闻	https://www.bombardier.com/en/media/newsList.html?filter-bu=all&f-year=all&f-month=all&f-type=all&show-by-page=100&page=1&f-min-year=2002
4	巴航工业官网新闻	https://embraer.com/global/en/news
5	三菱客机官网新闻	http://www.flythemrj.com/news/index.php#changes_top
6	MRO Network	https://www.mro-network.com
7	AeroTime	https://www.aerotime.aero/
8	Aerospace Manufacturing Design	http://www.aerospacemanufacturinganddesign.com/
9	Aerospace Manufacturing	https://www.aero-mag.com/
10	AIN	https://www.ainonline.com/

本文收集其中 800 篇相关新闻,作为待标注语料,并对其中的机型、技术进行了人工标注,选取其中 600 篇作为训练语料,另外 200 篇作为测试语料。

在对 800 篇新闻完成标注后,本文抽取其中所包含的机型名、技术名,并分别对机型、技术进行简要描述,如表2所示。

表 2 类别描述及举例

类别	代号	示例
机型全称	Mod. ful	Boeing 707 - 220 , A319 neo , SpaceJet M200
机型简称	Mod. abb	B707 , A320 , SSJ100
机型别称	Mod. alt	Beluga , Hornet , DreamLiner
技术全称	tec. ful	AM Additive Manufacturing , urban air mobility , Blended Wing Body
技术简称	tec. abb	3D printing , UAM , BWB

本文对模型采用 BIO 标注法 ,B 表示命名实体

Aeroflot	O	Maveric	Mod. alt	ALIAS	B-tec. abb
has	O	seamlessly	O	showcased	O
taken	O	blends	O	its	O
delivery	O	wings	O	ability	O
of	O	cockpit	O	to	O
a	O	and	O	use	O
brand	O	body	O	the	O
new	O	into	O	existing	O
Airbus	B-Mod. ful	a	O	737	B-Mod. abb
A350-900	I-Mod. ful	Blended	B-tec. ful	auto	B-tec. ful
		Wing	I-tec. ful	landing	I-tec. ful
		Body	I-tec. ful	system	I-tec. ful
		improving	O		
		its	O		
		aerodynamics	O		

图 5 标注语料展示

3.2 评价指标

实验基于 Python3 版本完成 ,采用 TensorFlow1.1 深度学习框架 ,并通过 ubuntu 系统运行。

对机型、技术命名实体识别模型的性能评价使用三个指标 ,准确率(Precision ,P) 、召回率(Recall ,R) 与 F_1 值。三个指标是衡量实体识别重要的指标 ,各个指标公式如下:

$$\text{正确率 } P = \frac{\alpha}{\beta} \times 100\% . \quad (10)$$

$$\text{召回率 } R = \frac{\alpha}{\gamma} \times 100\% . \quad (11)$$

$$F_1 \text{ 值} = \frac{2 \times P \times R}{P + R} \times 100\% . \quad (12)$$

其中: α 表示系统正确的实体个数 , β 表示系统识别出的所有实体个数; γ 表示文档中的全部实体个数; F_1 表示 P 和 R 的调和平均值 ,实现对模型性能的综合考量。

3.3 结果及分析

表 3 给出实验结果 ,展示了本文改进的 NER 模型在上述新闻文本下的实验结果 ,可以看出 ,在测试集的测试数据中 ,相关实体的命名实体有着不错的

的首字 ,I 表示命名实体中非首字部分 ,O 表示不是命名实体。对话料进行标注后的文本内容如图 5 所示。

在训练之前还需要对数据进行预处理 ,训练模型使用 CBOW 模型 ,其他参数使用 Word2Vec 默认的参数形式。训练得到的词向量维度为 100 维 ,最终获得 25 380 个词向量。

成绩 , F_1 值达到了 73.6。分析改进后的模型能取得较好的实验结果的原因 ,机型名字的构成模式有明显规律 ,技术名字的构成成分虽相比机型名字更加复杂 ,但模型引入注意力机制 ,帮助模型在获得句子的上下文特征时 ,能合理地对实体部分进行聚焦。

表 3 模型性能对比表

模型	准确率/%	召回率/%	F_1 值/%
CRF	69.5	53.3	60.2
单向 LSTM+CRF	76.9	60.1	67.6
本文改进的模型	82.3	66.6	73.6

为了更全面地说明本文实现的 NER 模型的性能 ,将本文实现的 NER 模型与传统的 LSTM+CRF 模型和 CRF 模型进行对比实验。选择 CRF 模型进行对比实验是因为本文模型就是基于 CRF 模型来实现的;选择 LSTM+CRF 模型进行对比是因为本文基于此模型进行改进 ,想通过对比来验证引入双向 LSTM 和注意力机制是否更有利于对于本文的命名实体的识别。

通过表 3 对比可以发现 ,改进后的模型在识别命名实体上有较大的性能提升 ,比单向 LSTM+CRF 模型提升大约 6% ,比单一的 CRF 模型效果提升更

加明显。可以看出,改进的双向隐含层结构,帮助模型更高效地捕捉上下文信息,注意力机制的融入对文本语义特征提取的效果有了明显提高,尤其面对文本篇幅不定的情况,对文本特征的捕捉发挥不可替代的作用。

3.4 错误分析

从识别有误的样本中随机选取 100 个,通过人

工方式复查这些错误项,其中 57 个是对标签的标注错误。主要是实体标注错误、实体类型区分错误,还有一些是将短语错误地标注成命名实体,如图 6 所示。这类错误占抽样总错误数的 57%。针对这一点,下一步工作需要重点关注实验语料的修复、优化和补充。

an	O	CFM	B-tec. ful
official	O	International	I-tec. abb
at	O	CFM56-5Bs	I-tec. ful
the	O	and	O
Ministry	B-tec. ful	International	B-Mod. ful
of	I-tec. ful	V2500s	I-Mod. ful
Land	I-tec. ful	powering	O
Infrastructure	I-tec. ful	a	O
Transport	I-tec. ful	fleet	O
(MLIT)	B-tec. abb		

图 6 错误标注示意图

出现的第二类错误是机型技术识别的泛化能力过强,导致将不是机型技术的词识别为机型技术,如 Boeing Q2 实际上是指波音的第二财季,被错误识别为机型。针对泛化问题,目前还没有合适的解决方案,下一步将其作为研究重点来继续提高模型识别能力。

4 结语

本文将基于深度学习的命名实体识别方法应用在竞争情报工作中,并针对竞争情报文本的特殊性对命名实体识别模型进行相应改进,在 LSTM-CRF 命名实体识别模型的基础上,将单向 LSTM 改成双向 LSTM,并加入注意力模型。改进后的命名实体识别模型应用于面向商用飞机制造领域的竞争情报工作中,对机型、技术两个重点关注的文本进行识别,实验结果 F_1 值为 73.6,证明将基于深度学习的命名实体识别方法用于竞争情报工作中是高效的。将改进后的新模型与单一 CRF 模型和单向 LSTM+CRF 模型进行对比,改进后的模型在识别效果上有明显优势,证明双向 LSTM 帮助模型更好地捕捉了上下文信息,注意力机制更好地解决了文本篇幅短小等问题。后续工作中,将进一步对实验语料进行修复和补充,同时继续优化命名实体识别模型来解决泛化问题。综上所述,基于深度学习的命名实体识别技术应用到竞争情报工作中,对海量开源信息的特征捕捉取得了良好效果。

参考文献

- [1] TUAN L T. Leading to Learning and Competitive Intelligence [J]. Learning Organization 2013 20(3): 216-239.
- [2] 吴金红,王翠波. 大数据时代企业竞争情报系统的评价指标体系研究 [J]. 现代情报 2016 36(2): 3-7.
- [3] EZIGBO C A, JUDUJI J I. Manage Competitive Intelligence for Strategic Advantage [J]. European Journal of Business & Management 2013 5(3): 17-36.
- [4] AHEARNE M, LAM S K, HAYATI B, et al. Intra-functional Competitive Intelligence and Sales Performance: A Social Network Perspective [J]. Journal of Marketing 2013 77(5): 37-56.
- [5] 刘浏,王东波. 命名实体识别研究综述 [J]. 情报学报 2018 37(3): 329-340.
- [6] 于根,李晓戈,刘睿,等. 基于信息抽取技术的问答系统 [J]. 计算机工程与设计 2017 38(4): 1051-1055.
- [7] 田家源,杨东华,王宏志. 面向互联网资源的医学命名实体识别研究 [J]. 计算机科学与探索 2018 12(6): 898-907.
- [8] 杨锦锋,于秋滨,关毅,等. 电子病历命名实体识别和实体关系抽取研究综述 [J]. 自动化学报 2014 40(8): 1537-1562.
- [9] BIKEL D M, MILLER S, SCHWARTZ R, et al. Nymble: a high-performance learning name-finder [J]. Conference on Applied Natural Language Processing 1997 20(3): 194-201.
- [10] 张华平,刘群. 基于角色标注的中国人名自动识别研究 [J]. 计算机学报 2004 27(1): 85-91.
- [11] 黄德根,岳广玲,杨元生. 基于统计的中文地名识别 [J]. 中文信息学报 2003 17(2): 36-41.

- [12] 邱泉清,苗夺谦,李浩然,等. 基于条件随机场的中文地名识别方法[J]. 武汉大学学报(信息科学版), 2017, 42(2): 150-156.
- [13] 程志刚. 基于规则和条件随机场的中文命名实体识别方法研究[D]. 武汉: 华中师范大学, 2015.
- [14] 邱泉清,苗夺谦,张志飞. 中文微博命名实体识别[J]. 计算机科学, 2013, 40(6): 196-198.
- [15] LIU L, SHANG J, XU F F, et al. Empower Sequence Labeling with Task-Aware Neural Language Model[J]. Proceedings of the Association for the Advancement of Artificial Intelligence, 2017, 36(13): 5253-5260.
- [16] 车金立,唐力伟,邓士杰,等. 融合词位字向量的军事领域命名实体识别[J]. 火炮发射与控制学报, 2019, 40(3): 50-55, 65.
- [17] 张俊飞,毕志升,王静,等. 基于 BiLSTM-CRF 中文领域命名实体识别框架设计[J]. 计算技术与自动化, 2019, 38(3): 117-121.
- [18] 买买提阿依甫,吾守尔·斯拉木,帕丽旦·木合塔尔,等. 基于 BiLSTM-CNN-CRF 模型的维吾尔文命名实体识别[J]. 计算机工程, 2018, 44(8): 230-236.
- [19] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE transactions on neural networks, 1994, 5(2): 157-166.
- [20] LAFFERTY J, MCCALLUM A, PEREIRA F, et al. Probabilistic Models for Segmenting and Labeling Sequence Data[J]. Proc. international Conf. on Machine Learning, 2002, 53(2): 282-289.
- [21] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [22] 单义栋,王衡军,黄河. 基于注意力机制的命名实体识别模型研究: 以军事文本为例[J]. 计算机科学, 2019, 46(增 1): 111-114, 119.