

基于 Lattice LSTM 的古汉语命名实体识别

崔丹丹 刘秀磊 陈若愚 刘旭红 李臻 齐林

北京信息科技大学计算机学院 北京 100192

(13718230670@163.com)

摘要 基于《四库全书》数据集,研究古汉语的命名实体识别技术。提出了基于 Lattice LSTM 模型的古汉语命名实体识别算法,该方法将字符序列信息和词序列信息共同作为模型的输入。采用甲言(jiayan)分词工具,利用 word2vec 训练古文字、词向量并作为 Lattice LSTM 模型的输入,提升了古汉语命名实体识别的效果。基于 Lattice LSTM 模型和预训练的古文字、词向量,提高了古汉语的实体识别效果,相比传统的 BiLSTM-CRF 模型,其 F1 分数提升 3.95% 左右。

关键词: 古汉语;命名实体识别;BiLSTM-CRF;Lattice LSTM;深度学习

中图法分类号 TP312

Named Entity Recognition in Field of Ancient Chinese Based on Lattice LSTM

CUI Dan-dan, LIU Xiu-lei, CHEN Ruo-yu, LIU Xu-hong, LI Zhen and QI Lin

Computer School, Beijing Information Science and Technology University, Beijing 100192, China

Abstract Investigated the named entity recognition problem of ancient Chinese literature based on the Complete Collection of Four Treasuries dataset. Proposed an algorithm for named entity recognition of ancient Chinese literature based on the Lattice LSTM model. This method combines both character sequence information and word sequence information as input to the model. Using jiayan word segmentation tool, word2vec is used to train character and word level embedding of ancient Chinese as input to the Lattice LSTM model, which improves the performance of named entity recognition based on ancient Chinese literature. Based on the Lattice LSTM model and pre-trained character and word level embedding of ancient Chinese, the performance of named entity recognition based on ancient Chinese literature is improved. Compared with the traditional Bi-LSTM-CRF model, its F1 score is improved by about 3.95%.

Keywords Ancient Chinese literature, Named entity recognition, BiLSTM-CRF, Lattice LSTM, Deep learning

1 引言

命名实体识别(Named Entity Recognition, NER)的主要任务是识别出待处理文本中预定义好的命名实体,是知识图谱、问答系统、句法分析等自然语言处理任务中基础而关键的任务。

古汉语凝聚了古人圣贤的智慧,是了解历史、传承中华文化、弘扬民族精神的重要途径,是现代汉语学习和发展的基础。对古汉语进行研究的难度不可小觑,其难点主要在于古汉语行文简练、陌生难懂等,在语法、句子长度等方面和现代文相比都具有一定差异。探索古汉语命名实体识别研究不仅要创新模型方法,还要提高分词的准确性,把握好实体边界的划分,才能使研究取得良好效果。

目前,针对古汉语暂时无法获得已经做好标注和完成分词的语料,同时也无特定的古汉语命名实体识别模型。基于上述问题,本文提出了基于 Lattice LSTM 模型的算法,研究基于《四库全书》的命名实体识别效果,有助于组织和挖掘古汉语的相关实体信息。

本文研究的主要成果:1)调研并比较了常用分词工具和

面向古文的分词工具在古汉语命名实体识别方面的应用效果;2)利用面向古文的分词工具,训练古文词向量;3)利用训练的古文词向量,基于 Lattice LSTM 模型构建古汉语命名实体识别模型,提升了古汉语的实体识别效果。

2 相关工作

2.1 命名实体识别技术

命名实体识别技术不断发展,如图 1 所示,由早期的基于规则和字典的方法发展到基于机器学习方法,再到目前比较受欢迎的基于深度学习方法、注意力机制和迁移学习方法等^[1]。

基于规则和词典的方法,根据设定的规则和词典进行字符串匹配,该方法无法适应数据的多样性,不易扩展数据类型,同时构建规则和字典过于依赖人力。

基于机器学习的方法,将实体识别看作序列标注问题。其中 CRF 模型比较受欢迎,因为其在解决序列标注问题时充分利用了内部和上下文的信息,但是运用该方法训练模型需要大量的标注样本,特征提取需要人工参与。

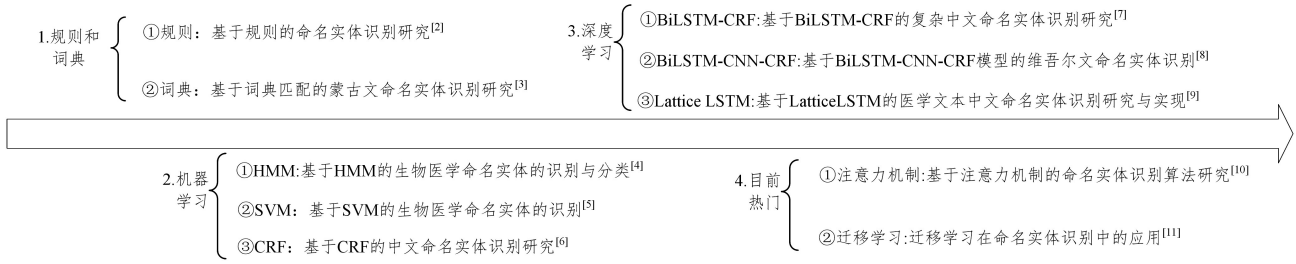


图1 实体识别的发展

Fig. 1 Development of named entity recognition

基于深度学习的方法是基于神经网络进行实体识别,不需要复杂的特征工程,降低了在领域研究中对领域知识的依赖性,这种端到端的方式具备自动学习的能力,能够自动学习输入文本中句子的抽象特征[12]。

注意力机制最早应用于视觉图像领域,随着技术的不断

发展,注意力机制和迁移学习等逐渐被应用到命名实体识别领域,成为目前较热门的方法。表1调研了通过不同的输入表示以及不同的神经网络结构完成的NER任务,主要采用CoNLL03数据集上的F1指标进行不同方法的效果评估,其中BiLSTM-CRF是目前NER中最常用的深度学习结构。

表1 NER 的不同方法

Table 1 Different methods of NER

Work	Input representation	Context encoder	Tag decoder	F1-score / %
[13]	LSTM(charcter)	LSTM	CRF	84.52%
[14]	LSTM(character), Word2vec(word)	LSTM	CRF	86.26
[15]	LSTM(character), Glove(word)	LSTM	CRF	91.07
[16]	LSTM(character), GloVe(word)	LSTM	Semi-CRF	91.38
[17]	LSTM(character), SENNA(word)	LSTM	CRF	90.94
[18]	LSTM(character), SENNA(word)	LSTM	Softmax	91.48

2.2 古文语料的研究

(1) 古代人名结构形式多样。古代人名由“名”“字”“号”组成,“字”通常为“名”的解释及补充。其中“名”与“字”有意义相同、相辅、相反等多种不同关系,例如唐代诗人白居易字乐天,“名”与“字”意义相辅,因“乐天”故能“居易”。

(2) 古代地名分布特点复杂。古代地名承载着古代文化的地域特征和民族特征,其主要有两方面:1)反映古代交通;2)反映古代军屯。在古代,“铺”“驿”“堡”“站”“台”等字多用于反映交通的地名;“所”“哨”“都”等字多用于反映古代军屯的地名。

(3) 朝代命名规则不一。朝代命名来源于4点:1)部落名称;2)创建者原有封号;3)创建者政权统治;4)宗族关系。例如:秦朝,源于古部落;魏,源于曹操的原有封号“魏公”。

(4) 古代官职级别种类繁多。国君级别为皇帝,中央级别官职有丞相、太尉、御史大夫,经过演变,后出现三省六部制度。三省:1)中书省;2)门下省;3)尚书省。六部:1)吏部;2)户部;3)礼部;4)兵部;5)刑部;6)工部。六部长官为尚书,副长官为侍郎。地方级别有太守、刺史等。品级官职有太师、太傅等。武官有将军、督统、都司、守备等。

3 Lattice LSTM 模型

Lattice LSTM 是 LSTM[19] 的一个变种。Lattice LSTM 将字序列信息和词序列信息同时编码[20],利用门结构控制信息流动。图2给出Lattice LSTM结构[20],该模型的主干是基于字符的LSTM[22],与基本LSTM的区别是该模型中LSTM的每个cell内部对信息的处理方式不同。图中,深色cell代表句子中的潜在词信息,与字符级cell互相连接。

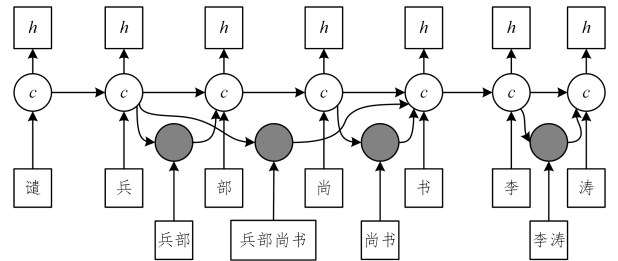


图2 Lattice LSTM 的结构

Fig. 2 Architecture of Lattice LSTM model

与基于字向量的LSTM模型相比, c_t^w 的计算不仅运用了字向量,还运用了词语序列 $w_{b,e}^d$,通常来讲,每个词向量可表示为 $x_{b,e}^w$ 。计算方法如式(1)所示:

$$x_{b,e}^w = e^{w \cdot (w_{b,e}^d)} \quad (1)$$

其中, e^w 为词向量矩阵。除此之外,定义 word cell $c_{b,e}^w$ 记录 $x_{b,e}^w$ 的递归状态。 $c_{b,e}^w$ 的计算方法如式(2)、式(3)所示:

$$\begin{bmatrix} i_{b,e}^w \\ f_{b,e}^w \\ \tilde{c}_{b,e}^w \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \tanh \end{bmatrix} \times W^{wT} \times \begin{bmatrix} x_{b,e}^w \\ h_b^c \end{bmatrix} + b^w \quad (2)$$

$$c_{b,e}^w = f_{b,e}^w \odot c_b^c + i_{b,e}^w \tilde{c}_{b,e}^w \quad (3)$$

其中, $i_{b,e}^w$ 代表输入门, $f_{b,e}^w$ 代表遗忘门。由于 $c_{b,e}^w$ 的存在,隐含层中 c_t^c 的值会被更多不同的信息流影响,例如图3中 c_t^c 由 x_t^c , $c_{3,t}^c$, $c_{1,t}^c$ 以及上一个隐含层中的输出共同影响[23]。通过构建额外的门 $i_{b,e}^w$,对 $c_{b,e}^w$ 到 $c_{b,e}^c$ 之间的信息流进行控制。计算方法如式(4)所示:

$$i_{b,e}^w = \sigma(W^{i^T} \times \begin{bmatrix} x_e^c \\ c_{b,e}^w \end{bmatrix} + b^i) \quad (4)$$

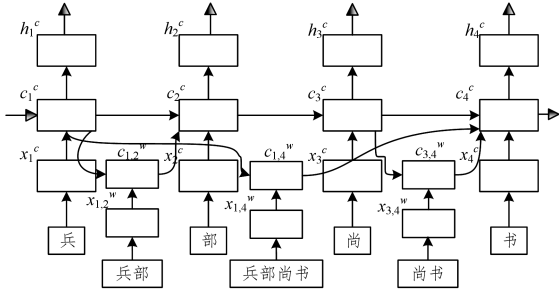


图3 Lattice结构

Fig. 3 Lattice structure

利用 $c_{b,e}^w$ 和 \tilde{c}_j^c 的值计算 c_j^c , 计算方法如式(5)所示:

$$c_j^c = \sum_{b \in (b' | w_b^d, j \in D)} \alpha_b^c \odot c_{b,j}^w + \alpha_j^c \odot \tilde{c}_j^c \quad (5)$$

对 $i_{b,j}^c$ 和 i_j^c 进行归一化, 使其 $i_{b,j}^c$ 和 i_j^c 各自和为 1。归一化操作计算方法如式(6)、式(7)所示:

$$\alpha_{b,j}^c = \frac{\exp(i_{b,j}^c)}{\exp(i_j^c) + \sum_{b' \in (b' | w_b^d, j \in D)} \exp(i_{b',j}^c)} \quad (6)$$

$$\alpha_j^c = \frac{\exp(i_j^c)}{\exp(i_j^c) + \sum_{b' \in (b' | w_b^d, j \in D)} \exp(i_{b',j}^c)} \quad (7)$$

对隐含层的计算方法如式(8)所示:

$$h_j^c = o_j^c \odot \tanh(c_j^c) \quad (8)$$

得出隐含层的所有输出 h_1, h_2, \dots, h_l , 其中 l 为句子中字的数量, 标签序列的生成概率 $y = l_1, l_2, \dots, l_l$, 计算方法如式(9)所示:

$$P(y|s) = \frac{\exp(\sum_i (W_{CRF}^{l_i} h_i + b_{CRF}^{(l_{i-1}, l_i)}))}{\sum_{y'} \exp(\sum_i (W_{CRF}^{l_i'} h_i + b_{CRF}^{(l_{i-1}', l_i')}))} \quad (9)$$

其中, y' 表示任意可能的 label 序列, $W_{CRF}^{l_i}$ 代表 l_i 的模型参数, $b_{CRF}^{(l_{i-1}, l_i)}$ 代表 l_{i-1} 到 l_i 的具体偏差。最后, 利用 Viterbi 算法找到得分最高的标签序列, 将其作为最后的结果输出^[22]。

4 实验

4.1 古文分词及字、词向量的预训练

目前针对现代汉语的分词工具较多, 研究也很成熟, 如 jieba 分词, 但现代汉语分词方法并不适用于古汉语分词。通过对古文分词方法的调研, 选取 jiayan 分词对古文本数据进行分词, 其主要技术是利用字符级隐马尔可夫模型对古汉语自动分词。分词效果如表 2 所列。

表 2 jiayan 分词结果

Table 2 jiayan word segmentation result

四库全书原文数据选取	jiayan 分词结果
改郡为州, 改太守为刺史, 立建成为太子, 封世民为秦王秦王, 以裴寂为右仆射, 窦威现为右将军之职。	改 郡 为 州, 改 太守 为 刺史, 立 建 成 为 太 子, 封 世 民 为 秦 王, 以 裴 寂 为 右 仆 射, 窦 威 现 为 右 将 军 之 职。

由于 word2vec 模型训练速度快、使用广泛、效果较好, 因此本次实验采用 word2vec 模型训练古文数据的字、词向量。训练字、词向量的语料要与研究领域数据紧密结合, 要求丰富且全面的数据量。本实验中训练字、词向量的语料包括《四库全书》中集部、经部、史部和子部 4 个方面, 数据量较丰富。

训练字、词向量的训练步骤包括: 1) 分词, 采用专注于古汉语的 jiayan 分词技术, 其分词效果较好且符合语境;

2) 字、词向量模型训练。

4.2 语料获取

本实验数据集的来源为《四库全书》原文, 命名实体的类别主要有人名、地名、朝代名和官职名 4 类。针对以上 4 类实体, 《史部》相比《经部》《集部》《子部》中的实体数据量更丰富, 所以选取《四库全书》中《史部》数据作为本实验的数据集, 其数据格式为 csv 文件。通过 python 对《四库全书》内容进行批量读取, 结合人名、地名、朝代和官职 4 类词典对数据进行标注。数据集采用 BIO 标注体系 (B-begin, I-inside, O-outside), 数据集格式如图 4 所示。标注好的语料规模共 258 万字, 由于机器标注存在一定的问题, 因此在机器标注的基础上, 人工对 10 万字进行修正, 减小了基于词典方法的误差。

其 O	北 B-DYN	拜 O
大 B-POS	齐 B-DYN	济 B-LOC
夫 I-POS	. O	阴 I-LOC
以 O	魏 O	郡 B-POS
下 O	收 O	丞 I-POS
皆 O	孝 O	, O
自 O	武 O	大 B-POS
置 O	置 B-POS	守 I-POS
之 O	帝 I-POS	刘 B-PER
。 O	六 O	育 I-PER
胡 B-PER	男 O	甚 O
广 I-PER	。 O	重 O
曰 O	卫 O	之 O
: O	皇 B-POS	任 O
“ O	后 I-POS	以 O
后 B-DYN	生 O	郡 B-POS
汉 I-DYN	庚 O	职 I-POS

图 4 数据集展示

Fig. 4 Sample of dataset

实验中训练集、验证集和测试集占比分别约 70%, 15% 和 15%, 其实体类别的数量如表 3 所列。

表 3 实体类别的数量

Table 3 Data volume of entity categories

实体种类	训练集	验证集	测试集
人名	20 042	4 259	3 375
地名	26 300	6 035	6 985
朝代名	325	30	54
官职名	7 116	1 834	1 184
实体总数	53 783	12 158	11 598

4.3 评价指标

为了评估 Lattice LSTM 模型对古汉语实体识别的效果, 本实验采用精确率 (precision)、召回率 (recall) 和 F1 score 作为评估标准。

精确率表示全部预测为正中正确预测为正确的比例, 精确率越高, 代表模型对负样本的区分程度越高; 召回率是正确预测为正的占全部实际为正确的比例, 召回率越高, 代表模型对正样本的识别程度越好; F1 score 结合了精确率和召回率, F1-score 越高, 代表模型越稳定。

4.4 实验分析

本节通过实验主要进行了数据规模影响的验证、模型超参数的调整、两种分词工具的对比以及传统机器学习模型和深度学习模型的效果对比。

为了验证数据规模对实验效果的影响, 将 BiLSTM-CRF 模型中训练集、验证集和测试集的数据规模均以 20% 的增长率增长, 且保证每次实验中训练集、验证集和测试集之间的比例均为 7:1.5:1.5, 对比实验结果, 验证数据规模对模型效果

影响。根据表 4 中的数据结果分析,随着数据量的逐渐增大,F1 分数逐渐增大,因为模型训练需要海量数据,数据量越多,BiLSTM-CRF 模型的学习效果越好。因此在合适的范围内,应该尽可能地增加数据量,进而改善模型的效果。

表 4 数据规模的影响

Table 4 Influence of datasize on performance of model

(单位:%)			
Data size	Precision	Recall	F1-score
20	75.45	71.96	73.66
40	74.16	76.00	75.07
60	82.85	83.53	83.19
80	84.19	82.94	83.56
100	85.34	83.64	84.48

经过实验,BiLSTM-CRF 和 Lattice LSTM 的最优超参数组合如表 5,传统机器学习 CRF 模型采用 L-BFGS 算法的梯度下降,其 L_1 、 L_2 正则化系数均设为 0.1,优化算法的最大迭代次数设为 100,测试模型准确性采用五折交叉验证方法。

表 5 两种模型的最优超参数

Table 5 Optimal hyperparameters for deep learning models

超参数	含义	BiLSTM-CRF	Lattice LSTM
char emb size	字向量维度	—	300
wordembsize	词向量维度	300	300
LSTM layer	LSTM 层数	2	1
LSTM hidden	隐藏单元数	200	200
learning rate	学习率	0.001	0.015
dropout	丢弃率	0.5	0.5

为了对比 jiayan 分词和 jieba 分词对古文实体识别效果的影响,采用 jieba 与 jiayan 两种分词方法分别训练古文词向量,在 BiLSTM-CRF 模型中固定数据集、超参数,通过对比两种词向量下得出的实验结果,验证两种分词方法对古文实体识别的效果。根据表 6 中的数据结果分析,基于 jiayan 分词的词向量相比基于 jieba 分词训练的词向量,在 BiLSTM-CRF 模型的 F1 值上约提升 0.97%,因此针对古文数据,jiayan 分词效果要优于 jieba 分词效果,jiayan 分词对古文数据更准确,是古汉语分词中一个较好的工具。

表 6 模型效果对比

Table 6 Comparison of model performance

(单位:%)				
模型	词向量	Precision	Recall	F1-score
CRF	—	99.17	79.99	87.98
BiLSTM-CRF	古文词向量 (jieba 分词)	82.25	84.81	83.51
BiLSTM-CRF	古文词向量 (jiayan 分词)	85.34	83.64	84.48
Lattice LSTM	古文词向量 (jiayan 分词)	90.20	86.73	88.43

为了验证深度学习方法与传统机器学习方法在《四库全书》中《史部》数据集上的效果,本实验选取传统机器学习中的 CRF 模型和基于深度学习的 BiLSTM-CRF 模型作为基线模型,并与深度学习 Lattice LSTM 模型的效果做对比。表 6 中的数据结果表明,Lattice LSTM 模型相比 CRF 模型在召回率和 F1 分数两个指标上均得到了有效提升。Lattice LSTM 模型相比 BiLSTM-CRF 模型在精确率、召回率和 F1 分数上分别提升了约 4.86%,3.09%和 3.95%,其原因是 Lattice LSTM 模型相比 BiLSTM-CRF 模型的结构是以字向量和词

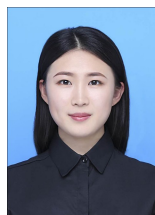
向量同时作为模型的输入,实现了字词级的特征融合。由于 Lattice LSTM 模型效果显著优于 CRF 和 BiLSTM-CRF 模型效果,本实验最终采用 Lattice LSTM 模型完成对古汉语命名实体的识别任务。

结束语 本文从古汉语中人名、地名和官职名的构成特点考虑,采用了 jiayan 分词工具,预训练针对古文的词向量,完成了基于 Lattice LSTM 的古汉语命名实体识别任务。与传统的 CRF, BiLSTM-CRF 模型相比, Lattice LSTM 模型在 NER 任务中效果更优。在未来的研究中,将继续探索古汉语实体识别在其他下游任务中的应用。

参考文献

- [1] CHEN S D, OUYANG X Y. Summary of Named Entity Recognition Technology[J]. Radio Communication Technology, 2020, 46(3): 251-260.
- [2] ZHOU K. Research on Rule-based Named Entity Recognition [D]. Hefei: Hefei University of Technology, 2010.
- [3] BAO M N, SI L, GE L. Research on Mongolian Named Entity Recognition Based on Dictionary Matching[J]. Journal of Minzu University of China, 2017, 44(3): 165-169.
- [4] CHEN J, CHANG Z Q, XU J. Recognition and Classification of Biomedical Named Entities Based on HMM[J]. Computer Age, 2006(10): 40-42.
- [5] WANG H C, ZHAO T J. Recognition of Named Biomedical Entities Based on SVM[J]. Journal of Harbin Engineering University, 2006, 27(z1): 570-574.
- [6] SHI H F. Research on Chinese Named Entity Recognition Based on CRF[D]. Suzhou: Soochow University, 2010.
- [7] GU Y. Research on Recognition of Complex Chinese Named Entity Based on BiLSTM-CRF[D]. Nanjing: Nanjing University, 2019.
- [8] MAI M T, FU A Y, WU S E, et al. Uyghur Named Entity Recognition Based on BiLSTM-CNN-CRF Model[J]. Computer Engineering, 2018, 44(8): 230-236.
- [9] ZHANG X T. Research and Implementation of Medical Text Chinese Named Entity Recognition Based on Lattice LSTM[D]. University of Electronic Science and Technology of China, 2019.
- [10] LI Y Y. Research on Named Entity Recognition Algorithm Based on Attention Mechanism[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [11] SHENG J. Application of transfer learning in named entity recognition[D]. Harbin: Harbin Institute of Technology, 2019.
- [12] KHAN W, DAUD A, ALOTAIBI F, et al. Deep recurrent neural networks with word embeddings for Urdu named entity recognition[J]. ETRI Journal, 2020, 42(1).
- [13] KURU O, CAN O A, YURET D. Character-level named entity recognition[C]//COLING. 2016: 911-921.
- [14] REI M. Semi-supervised multitask learning for sequence labeling [C]//ACL. 2017: 2121-2130.
- [15] TRAN Q, MACKINLAY A, YEPES A J. Named entity recognition with stack residual lstm and trainable bias decoding[C]//IJCNLP. 2017: 566-575.
- [16] YE Z X, LING Z H. Hybrid semi-markovcrf for neural sequence labeling[C]//ACL. 2018: 235-240.

- [17] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition [C] // NAACL. 2016:260-270.
- [18] GREGORIC A Z, BACHRACH Y, COOPE S. Named entity recognition with parallel recurrent neural networks [C] // ACL. 2018:69-74.
- [19] ZHANG C P, FANG T, LIU Y L. Research and Application of Named Entity Recognition Technology Based on LSTM-CRF [J]. Computer Technology and Development, 2019, 29(2):106-108.
- [20] Biotechnology-Biomedical Informatics; Data on Biomedical Informatics Described by Researchers at National University of Defence Science and Technology (Adversarial training based lattice LSTM for Chinese clinical named entity recognition) [J/OL]. <https://schlr.cnki.net/Detail/index/SPQDLAST/SPQDC4D58041D94430FE02664BB977E1BBDB>.
- [21] WANG B R, LIN X, ZHU X D, et al. Research on Named Entity Recognition Model of Chinese Medical Text by Lattice LSTM Neural Network [J]. Chinese Journal of Health Informatics and Management, 2019, 16(1):84-88.
- [22] ZHAO S, CAI Z P, CHEN H W, et al. Adversarial training based lattice LSTM for Chinese clinical named entity recognition [J]. Journal of Biomedical Informatics, 2019, 99.
- [23] ZHANG W J, ZHANG H M, YANG L, et al. Multi-granular Chinese Word Segmentation Based on Lattice-LSTM [J]. Journal of Chinese Information Processing, 2019, 33(1):18-24.



CUI Dan-dan, born in 1997, postgraduate. Her main research interests include natural language processing and knowledge graph.



CHEN Ruo-yu, born in 1982, Ph.D, lecturer, is a member of China Computer Federation. His main research interests include natural language processing, data mining and semantic Web.

(上接第 5 页)

- [9] MA H L, ZHOUS B. Human machine integration is the development direction of artificial intelligence [J]. China's Information Industry, 2017(6):44-46.
- [10] MAM, LIN W L, PAN D S, et al. Data and Decision Intelligence for Human-in-the-Loop Cyber-Physical Systems [J]. Journal of Signal Processing Systems, 2017, 90(11):1167-1178.
- [11] ADAMS S, AREL I, BACH J. Mapping the landscape of human-level artificial general intelligence [J]. AI Magazine, 2015, 33(1):25-42.
- [12] WU Z, ZHOU Y, SHI Z, et al. intelligence: Recent progress and future directions [J]. IEEE Intelligent Systems, 2016, 31(6):44-50.
- [13] ZHANG J J, WANG F Y, WANG X. Cyber-Physical-Social Systems: The State of the Art and Perspectives [J]. IEEE Transactions on Computational Social Systems, 2018, 5(3):829-840.
- [14] HAQUE S A, AZIZ S M, RAHMAN M. Review of cyber-physical system in healthcare [J]. Int. J. Distrib Sensor Netw, 2014, 10(4):217-225.
- [15] LI G J, XU Z W. Looking at the new economy from the development trend of information technology [J]. Journal of Chinese Academy of Sciences, 2017, 32(3):233-238.
- [16] CHEN A, WANG H. Research on real-time control of intelligent walker for physical human-computer interaction [J]. Computer Application Research, 2017, 34(5):1362-1366.
- [17] Chinese Academy of Science. Roadmap of information technology development in China to 2050 [M]. Beijing: Science Press, 2009.
- [18] Chinese Academy of Science. The new trend of science and technology development and the strategic choice for 2020 [M]. Beijing: Science Press, 2013.
- [19] State Council. New generation artificial intelligence development plan [J]. China Informatization, 2017, 14(8):12-13.
- [20] LI J H. Comparison and Enlightenment between German Industry 4.0 and made in China Construction 2025 [J]. Journal of China University of Geoscience, 2015, 61(5):4-8.
- [21] GIUSEPPE A, VALERIO P, ANTONIO P. A Survey on Information and Communication Technologies for Industry 4.0: State-of-the-Art [J]. IEEE Communication Surveys & Tutorials, 2019, 21(4):3467-3498.
- [22] YANG L. Cyber Physical System (CPS)-Based Industry 4.0: A Survey [J]. Journal of Industrial Integration and Management, 2017, 2(3):175-189.
- [23] SUPARN D, ZHOUY C, IKER L. Cyber-Physical-Social Frameworks for Urban Big Data Systems: A Survey [J]. Applied Science, 2017, 7(10):1017-1039.
- [24] ANDREA C, CHIARA F D. Smart innovative cities: The impact of Smart City policies on urban Innovation [J]. Technological Forecasting and Social Change, 2019, 142(5):373-383.
- [25] NICOS K S, CHARALAMPOS B, CHRISTINA K. Smart city ontologies: Improving the effectiveness of smart city applications [J]. Journal of Smart Cities, 2016, 1(1):114-127.
- [26] JIANG Z. Research on the path from smart city to smart society [J]. Communication Enterprise Management, 2019(7):65-67.



WANG Hai-tao, born in 1976, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include network architecture, wireless self-organizing network, big data and artificial intelligence.



SONG Li-hua, born in 1976, Ph.D, professor. Her main research interests include network architecture, network security, network measurement and artificial intelligence.