文章编号：2096-1472(2020)-01-07-05
DOI:10.19644/j.cnki.issn2096-1472.2020.01.002

基于强化学习协同训练的命名实体识别方法

程钟慧[1]，陈 珂[1,2]，陈 刚[1,2]，徐世泽[3]，傅丁莉[3]

(1.浙江大学计算机科学与技术学院，浙江 杭州 310027；
2.浙江省大数据智能计算重点实验室，浙江 杭州 310027；
3.浙江华云电力工程设计咨询有限公司，浙江 杭州 310027)

摘 要：命名实体识别是一项从非结构化大数据集中抽取有意义的实体的技术。命名实体识别技术有着非常广泛的应用，例如从轨道交通列车产生的海量运行控制日志中抽取日期、列车、站台等实体信息进行进阶数据分析。近年来，基于学习的方法成为主流，然而这些算法严重依赖人工标注，训练集较小时会出现过拟合现象，无法达到预期的泛化效果。针对以上问题，本文提出了一种基于强化学习的协同训练框架，在少量标注数据的情况下，无须人工参与，利用大量无标注数据自动提升模型性能。在两种不同领域的语料上进行实验，模型F_1值均提升10%，证明了本文方法的有效性和通用性。同时，与传统的协同训练方法进行对比，本文方法F_1值高于其他方法5%，实验结果表明本文方法更加智能。

关键词：强化学习；协同训练；命名实体识别
中图分类号：TP391.1 **文献标识码**：A

Named Entity Recognition Method Based on Co-training of Reinforcement Learning

CHENG Zhonghui[1],CHEN Ke[1,2],CHEN Gang[1,2],XU Shize[3],FU Dingli[3]

(1.*College of Computer Science and Technology,Zhejiang University,Hangzhou* 310027,*China*;
2.*Key Laboratory of Big Data Intelligent Computing of Zhejiang Province,Hangzhou* 310027,*China*;
3.*Zhejiang Huayun Electric Power Engineering Design & Consultation CO.,LTD.,Hangzhou* 310027,*China*)

Abstract: Named entity recognition(NER)is a technique for extracting meaningful entities from unstructured big datasets.NER has a wide range of applications.An example of NER is advanced data analysis which extracts date,train,platform and other entity information from a large operation logs dataset produced by rail transit trains.In recent years,the reinforcement learning based method has become the mainstream method of solving this task.However,these algorithms rely heavily on manual labeling.The over-fitting problem may occur when the training set is small,and cannot achieve the expected generalization effect.In this paper,we propose a novel method,Reinforced Co-Training.With only small amount of labeled data,the performance of the named entity recognition model can be automatically improved by using a large amount of unlabeled data.We have experimented our framework on corpus in two different fields,the results show that the F1 value of our proposed method is increased by 10%,which proves the effectiveness and generality of the method in this paper.We also compared our method with the traditional co-training methods,the F1 value of our method is 5% higher than other methods,which shows that this method is more intelligent.

Keywords:reinforcement learning;co-training;named entity recognition

1 引言(Introduction)

给定一个非结构化大数据集(如轨道交通列车控制系统产生的车辆运行控制日志)，命名实体识别(NER)技术的目的是从该数据集中提取出具有特定意义的实体，如站台名、列车号、控制指令等[1]，进而为其他大数据建模任务提供实用信息。研究者们将NER任务归约为序列标注问题[2]，基于统计机器学习的方法和深度学习的方法成为主流，例如条件随机场模型[3]，基于卷积网络的序列标注模型[4]以及基于双向LSTM网络的模型[5]等。然而，基于学习的方法严重依赖人工标注，训练集较小时会出现过拟合现象，无法达到预期的泛化效果。同时，命名实体具有极强的不确定性，在进行大规模的数据标注时需要消耗大量的人力和时间，其代价是难以接受的。与标注语料不同，无标注语料数量巨大且极易获得，因此如何发挥大量无标注语料的价值，在少量标注数据的情况下改善模型学习性能是命名实体识别进一步研究的重点。

半监督学习[6]方法通常利用大量的无标注数据来辅助少量的有标注数据进行学习，从而提高模型学习性能。协同训练(Co-training)[7]是广泛使用的半监督学习方法之一，它利用两个学习器的“相容互补性”来互相标记样本扩大训练集，从而达到借助无标注数据提升学习性能的目的。协同训练的关键在于挑选高质量的无标注数据添加到训练集中，目前通常使用启发式的样本选择策略。然而，现有的协同训练算法存

基金项目：国家重点研发计划课题(2017YFB1201001).

在一些缺陷。首先，在训练过程中，每次添加两个弱分类器的伪标注数据到训练集中，会造成噪声累积。其次，由于少量标注数据和大量无标注数据在分布上具有一定差异，在训练一段时间后，会导致采样偏差向无标注数据方向偏移，进而导致训练模型性能降低。此外，传统的协同训练方法为了减少噪声的引入，每次添加模型置信度高的预测结果到训练集中，容易造成局部采样限制，会限制模型泛化能力[8]。

因此，一种理想的协同训练算法应该具备两个特性，一是扩充训练集带来的噪声应尽可能小，二是能对数据空间进行充分探索，以获得更好的泛化学习性能。基于以上，本文利用深度Q网络(Deep Q-network)[9]自动学习选择策略替代传统的启发式样本选择策略，进而提高协同训练效果。

本文的主要贡献如下：

(1)提出了一种基于强化学习的协同训练框架，在少量标注数据情况下，无须人工参与，利用大量无标注数据自动提升命名实体识别模型的性能。

(2)提出了一种基于实体级置信度的模型集成方法，减少协同训练过程中噪声的引入，进一步提高添加样本的质量。

(3)在人民日报和金融新闻语料上进行重复实验，证明了本文方法的有效性、通用性和鲁棒性。同时，与传统的协同训练方法进行对比实验，本文方法F1值高于其他方法5%。

2 相关工作(Related work)

针对如何在少量标注数据的情况下，使用半监督学习方法进行命名实体识别任务，已有学者做了相关研究。Liao[10]等人提出了一种基于CRF单一分类器的半监督命名实体识别方法，需要人工分析数据，提取有效规则，难度较大且规则的领域移植性较差。Aryoyudanta[11]等人使用SVM单一分类器，基于上下文和实体两种不同的属性视图构建两个学习器进行协同训练。Xiao[12]等人提出了一种基于CRF和SVM协同训练的中文机构实体识别算法，定义了一种启发式样本选择策略。然而，这些半监督学习方法都是基于人工预先设定的样本选择策略，无法对数据空间进行充分准确的学习。协同训练算法的核心在于样本选择策略，Zhang[13]等人提出了一种性能驱动的样本选择策略，选择有助于提高分类精度的无标注数据进行半监督学习。同时，Chawla[14]等人论证了随机挑选样本的方法会导致训练模型向无标注数据分布方向发生采样偏移。

与上述半监督命名实体识别方法相比，本文使用深度强化学习模型自动学习样本选择策略。深度强化学习(DRL)[15,16]是人工智能领域新的研究热点，它将深度学习(DL)[17]在特征表示方面较强的抽象感知能力和强化学习(RL)[18]的推理决策能力相结合。Lange[19]等人最先将深度学习模型和强化学习方法结合，提出了一种深度自动编码器，但是只适用于状态空间维度较小的问题。Mnih[9]等人结合深度学习中的卷积神经网络和传统强化学习中求解最优动作值函数的Q学习算法，提出了深度Q网络模型(DQN)来近似表示动作值函数。近年来，深度强化学习在自然语言处理领域获得了越来越多的关注，在会话生成、文本摘要等任务中均有应用。但由于语言是离散的，句子空间是无穷的，所以在将NLP任务转化为DRL问题时存在诸多挑战。

3 基于强化学习的协同训练框架(Reinforced Co-training)

3.1 未标注数据子集的划分

由于无标注数据数量巨大，如果在每次迭代过程中只选择一个样本添加到训练集中，并重新训练两个学习器，那将十分低效。所以，我们首先将大量的无标注数据样本切分成句子，并根据句子间的相似度大小，将其划分成子集。这样每次算法挑选一个无标注数据子集作为候选样本添加到训练集中，更新两个学习器，能极大提高计算效率，节约时间成本。

本文使用Jaccard相似度算法计算句子间的相似度：

$$sim(S_1,S_2)=\frac{|S_1\cap S_2|}{|S_1\cup S_2|} \tag{1}$$

式中，$S_1,S_2\in R^{|V|}$为采用独热编码表示的无标注数据集中的句子向量，V表示由无标注数据集中出现过的字符构成的字表。然后，使用局部敏感哈希算法[20]将无标注数据集U划分成K个子集$\{U_1,U_2,...,U_K\}$。

3.2 基于实体级置信度的模型集成方法

由于协同训练的学习器都是弱学习器，准确率较低，导致伪标注数据中存在大量错误标注，在迭代过程中特别是在训练初期噪声不断累积。此外，由于实体是短语级的，所以不能简单地对标注序列进行加权融合。因此，本文提出了一种基于实体级置信度的模型集成方法，有效地降低了噪声影响。对于两个模型同时识别出的实体且类型一致时，直接加入结果集合；对于单个模型识别出的实体且不冲突时，若大于阈值则添加否则丢弃；对于两个模型识别不一致的冲突实体，比较置信度大小。

对于一个无标注样本，使用两个学习器C_1、C_2分别进行预测，可以得到两个预测实体集合E_1、E_2，以及每个实体e的置信度：

$$conf_e=(\sum_{i=f}^{b-1}A_{y_i,y_{i+1}}+\sum_{i=f}^{b}P_{i,y_i})/n \tag{2}$$

其中，f、b为实体e在输入序列中起始，终止位置的下标，A为序列的转移矩阵，$A_{y_i,y_{i+1}}$表示从标签y_i转移至标签y_{i+1}的转移分数，P_{i,y_i}表示序列中第i个字符被预测为标签y_i的分数，n为实体长度。

3.3 强化学习模型

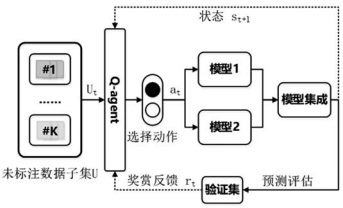

图1 基于强化学习的协同训练框架

Fig.1 The reinforced co-training framework

如图1所示，本文将协同训练的样本选择过程看作时序决策问题，使用基于流的数据组织形式。在两个命名实体识别模型C_1、C_2迭代过程中t时刻，智能体观察当前所处环境状态s_t，包括候选样本U_t的状态和两个学习模型C_1、C_2的状态，之后根据当前策略π选择执行的动作a_t。若$a_t=1$表示将当前候选样本加入训练集中，之后重新训练两个模型；若$a_t=0$则表示不添加，对数据流中的下一个子集继续进行判断。同时设置一个验证集，将模型在验证集上F_1值的变化作为奖赏反馈r_t，返回给强化学习智能体。重复这一过程，直到无标注数据被耗尽或不再有可添加的无标注数据子集时终止。强化学习模型的核心是智能体，其功能是在每一时刻，选择执行使未来累积折扣奖赏最大化的动作，从t时刻开始到T时刻终止的累积折扣奖赏函数定义如下：

$$R_t=\sum_{t'=t}^{T}\gamma^{t'-t}r_{t'} \tag{3}$$

其中，$\gamma\in[0,1]$为折扣因子，用来权衡未来奖赏对累积奖赏的影响。

3.3.1 状态表示

状态表示的目的是将协同训练的两个学习器，以及当前

候选样本的状态进行集成，传递给智能体，使其可以充分感知当前环境，从而在不断与之交互的过程中学习到良好的样本选择策略。由于语言是离散的，句子空间是无穷的，所以本文采用卷积神经网络，将高维的输入数据编码成低维抽象的特征向量。

传统的协同训练样本选择策略具有一定的局限性，本文提出的强化学习智能体权衡候选样本的多样性、学习器对该样本预测结果的置信度以及不确定性三方面因素。因此，用于状态表示的特征向量由内容表征、不确定性表征以及置信度表征三部分构成，下面将详细展开介绍。

(1)内容表示

对候选样本的内容进行抽象表征，使用卷积神经网络对输入的字向量矩阵进行编码，得到一个固定长度的向量表示。对于一个含有n个字符的序列，我们可以得到表示这个序列的字向量矩阵$X_i=\{x_{i,1},x_{i,2},\ldots,x_{i,n}\}$。我们的候选样本为一个包含$N$个序列的无标注数据子集$U$，我们将这$N$个序列的字向量矩阵进行叠加得到矩阵$U_s=\{X_1,X_2,\ldots,X_N\}$，通过三个大小不同的卷积核，进行卷积操作，之后使用ReLU激活函数对输出的隐藏层进行非线性变换，最后通过最大池化层，保留关键特征，过滤其余特征，得到三个隐藏状态向量h_{s1}、h_{s2}、h_{s3}，串联起来组合成内容表征向量h_s。

(2)不确定性表示

分类的不确定性通常采用“信息熵”来度量，为了使模型学习到更多不同的、潜在的特征规律，本文直接对模型输出的预测概率向量$p_C(y|x_i)$进行学习，其中，$p_C(y|x_i)$表示学习器C对序列中的字符x_i的预测为y中各标签的相应概率值。对于一个序列我们使用$P_i=\{p_{i,1},p_{i,2},\ldots,p_{i,n}\}$表示两个协同训练模型$C_1$，$C_2$对其标注的不确定性度量，其中：

$$p_{i,j}=(p_{c1}(y|x_{i,j})+p_{c2}(y|x_{i,j}))/2 \tag{4}$$

表示两个模型对字符$x_{i,j}$进行预测得到的各标签类别概率分布的均值。对于含有N个句子的候选样本U，我们将每个句子的概率分布矩阵进行卷积操作，之后使用平均池化层进行降维，从而得到每个卷积核的不确定性度量。最后的输出向量作为学习模型对候选样本预测结果不确定性的表示h_e。

(3)置信度表示

首先，使用如下公式计算命名实体识别模型C_1，C_2对序列X_i预测结果的置信度：

$$Conf_C=\sqrt[n]{\max_Y p_C(Y|X_i)} \tag{5}$$

$$Conf=(conf_{C_1}+conf_{C_2})/2 \tag{6}$$

其中，Y表示标注结果序列，$n=|X_i|$为输入序列长度。所以，对于候选样本U，依次对其中的所有句子计算置信度得到长度为N的数组h_c作为置信度表示。

最后，将上面得到的内容特征向量h_s，不确定性特征向量h_e和置信度特征向量h_c，一起作为环境状态特征表示，输入到深度Q网络中。

3.3.2 深度Q网络

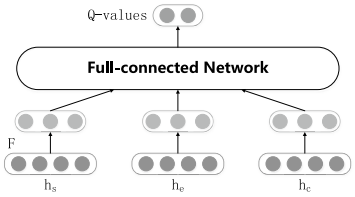

图2 深度Q网络结构图

Fig.2 The structure of Q-network

在每个时间步中，强化学习智能体在完成对输入数据关键特征的感知后，需要依据当前状态特征做出动作的选择，如图2所示，使用深度神经网络表示状态动作值函数$Q(s,a)$：

$$z_a=\varphi(\{F(h_s),F(h_e),F(h_c)\};\theta) \tag{7}$$

其中，F表示全连接层的映射操作，将状态特征向量转换成固定输出维度，$\varphi(\cdot)$为一层全连接神经网络，输出对应动作的价值估计。之后进行softmax归一化得到选择动作：

$$Q(s,a)=\text{softmax}(z_a) \tag{8}$$

$$a=\max_a Q(s,a) \tag{9}$$

3.3.3 动作表示

本文采用基于流的数据组织形式，智能体的决策动作简化为$a_t\in\{0,1\}$。对于候选样本U_t，智能体进行选择决策a_t：当$a_t=1$时，使用两个学习模型对U_t进行标注，之后将结果进行集成，并把集成后的标注样本加入训练集中，在更新后的训练集上重新训练两个模型；当$a_t=0$时，不对U_t采取任何动作，对数据流中的下一个无标注数据子集U_{t+1}继续进行判断，直到全部无标注数据耗尽，或数据流中无可添加的样本时终止。

3.3.4 奖赏函数

我们希望通过训练强化学习智能体使其学习到高质量的样本选择策略，最终目的是提升协同训练模型的性能，所以本文采用性能驱动的奖赏函数。在第t次迭代时，奖赏反馈值r_t定义为在执行动作a_t后，将协同训练模型C_1、C_2进行集成得到的模型E在验证集上的准确率Acc(F_1值)的变化：

$$r_t=Acc_t(\mathrm{E})-Acc_{t-1}(E) \tag{10}$$

这样做一方面合理有效地综合考虑了两个协同训练模型的性能变化，同时也解决了深度学习模型的性能波动问题。

3.3.5 网络的训练

本文使用Q学习(Q-learning)算法[21]训练智能体学习样本选择策略，通过优化基于贝尔曼公式计算的目标值函数$V(\theta_{i-1})$和当前值函数$Q(s,a;\theta_i)$之间的均方误差函数$L_i(\theta_i)$来更新Q网络参数θ。公式如下：

$$L_i(\theta_i)=E_{s,a}[(V(\theta_{i-1})-Q(s,a;\theta_i))^2] \tag{11}$$

$$V(\theta_{i-1})=E_{s'}[r+\gamma\max_{a'}Q(s',a';\theta_{i-1})|s,a] \tag{12}$$

使用自适应学习率的随机梯度下降算法对网络参数进行端到端的更新。

4 实验与分析(Experiment and analysis)

4.1 实验数据

本文选用人民日报(1998年)和金融新闻两种不同领域的语料库对前文提出的基于强化学习协同训练模型进行评估实验，其中人民日报为通用领域公开数据集，是中文命名实体识别任务常用的语料；金融新闻是从金融网站上利用爬虫技术获取的1000篇经人工标注的新闻语料，具有一定的领域特性。其中人民日报语料共有19484个句子、52735个实体，包括人名、地名、机构名三类；金融新闻语料含有26233个句子、56813个实体，包括人名、地名、机构名、日期、货币、百分比、时间七类。我们将原始的带标注语料划分成四个数据集：训练集、验证集、测试集和无标注数据集，首先随机选取500个句子作为少量标注训练集，之后从剩余的数据中选取10%作为验证集，10%作为测试集，其余80%去除标注结果作为协同训练过程中待添加的无标注数据集。

4.2 实验配置

4.2.1 实验环境

本文的实验是在一台小型服务器上运行的，CPU处理器为Intel(R) Xeon(R) Silver 4114 CPU @2.2GHz，GPU为

GeForce GTX 1080Ti，内存为100GB，操作系统为Ubuntu 18.04.1 LTS Server。使用的编程语言为Python，版本为3.6.7，使用深度学习框架TensorFlow 1.12.0。

4.2.2 实验设置

(1)协同训练模型

本文选用了两个主流的命名实体识别模型进行协同训练，其一是CRF序列标注模型[3]，另一个则是基于深度学习的BiLSTM-CRF模型[5]。前者属于传统的概率图模型，对条件分布进行建模，后者旨在通过一个深度神经网络直接学习从输入文本到标注序列的映射关系，是一个端到端的过程。两种模型从学习和训练的原理上具有显著的差异性和互补性。

实验使用python-crfsuite库实现CRF模型。为避免分词的影响，采用基于字符特征的CRF模型。特征方面选择大小为2的上下文窗口，考虑前后各两个字符对当前字符的影响，特征包括1-Gram、2-Gram、字符所在词的词性以及在词中的位置特征。BiLSTM模型的batch_size为16，学习率取0.001，采用Adam梯度下降优化算法，为防止过拟合，实验中采用early_stop准则，使用验证集，如果评价指标在验证集上连续三个epoch没有变化，就停止训练。

(2)网络参数设置

在本文的协同训练框架中，我们将无标注数据子集的数量K设置为100。在候选无标注样本的内容表示部分，分别通过128个大小为3、4、5的卷积核，步长为1进行卷积，使用ReLU激活函数。在不确定性表示部分，使用20个大小为3的卷积核，步长为1进行卷积。全连接层F输出向量维度为256。设置折扣因子γ为0.99，batch_size为32。回放记忆单元M最大容量为1000个转移样本，学习率α和行为策略ε-greedy的参数ε都设置为从开始到1000个转移样本区间内线性递减的形式，即α从0.005降到0.00025，ε从1.0降到0.0001。

(3)对比实验设置

我们将本文提出的基于强化学习的协同训练方法RL Co-Training与两种经典方法对比：

①Standard Co-Training：协同训练的两个模型各自随机选择伪标注样本进行协同训练[7]。

②CoTrade Co-Training：协同训练的两个模型各自挑选置信度高的伪标注样本，添加到对方的训练集中[22]。

实验中采用F1指标衡量命名实体识别模型的性能。F_1值的计算方式如下：

$$F_1=\frac{2\cdot P\cdot R}{P+R} \tag{13}$$

其中，P代表精确率，R代表召回率。

4.3 实验结果分析

本文在人民日报和金融新闻两种不同领域的语料上分别进行实验，并与两种经典的协同训练算法Standard Co-Training，CoTrade Co-Training进行对比。实验中，首先使用从语料中随机选择的500个句子作为少量的带有标注的训练数据对两个学习模型进行初始化，得到两个弱学习器，之后分别使用三种不同的协同训练算法，每次根据各自不同的样本选择策略不断添加100句伪标注数据，扩增训练集，迭代训练学习模型，最后利用测试集计算模型对所有实体识别的F_1值，具体结果如表1和表2所示。

表1 人民日报语料实验结果

Tab.1 The experimental results on People's Daily

F1值	CRF	BiLSTM	模型集成
初始化	0.741	0.759	0.784
Standard Co-Training	0.779	0.785	0.790
CoTrade Co-Training	0.810	0.818	0.824
RL Co-Training	0.857	0.870	0.879

表2 金融新闻语料实验结果

Tab.2 The experimental results on financial news

F1值	CRF	BiLSTM	模型集成
初始化	0.737	0.730	0.758
Standard Co-Training	0.774	0.776	0.778
CoTrade Co-Training	0.790	0.800	0.802
RL Co-Training	0.831	0.840	0.849

从中可以看出，(1)本文提出的基于强化学习的协同训练方法RL Co-Training在只有少量标注数据初始化模型的情况下，无须人工参与，通过利用大量无标注语料迭代训练，可以有效提升模型性能，在两种测试语料上模型的F_1值均可有效提升10%左右，证明了本文方法的有效性和通用性；(2)本文的RL Co-Training方法表现效果要好于传统的协同训练方法，其F_1值高于其他方法5%左右。

为了进一步比较分析三种协同训练算法样本选择策略的性能，图3和图4分别给出了在两种语料上，各方法在验证集上F_1值随迭代次数的变化情况。在每种语料上分别展示了在协同训练迭代过程中，两种协同训练模型CRF和BiLSTM，以及对两个模型进行集成后的性能变化，横轴表示迭代过程中添加进训练集中的伪标注数据的句子数量，纵轴表示模型在验证集上的F_1值。

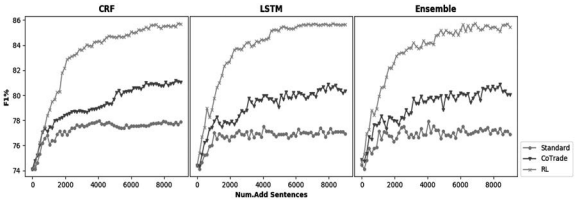

图3 人民日报语料上三种协同训练算法的学习曲线

Fig.3 Learning curves of three co-training algorithms on People's Daily Corpus

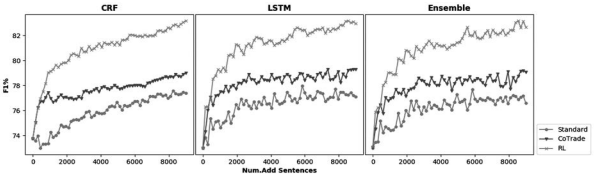

图4 金融新闻语料上三种协同训练算法的学习曲线

Fig.4 Learning curves of three co-training algorithms on financial news

从中可以看出：(1)在添加句子数量相同的情况下，本文的RL Co-Training方法获得了最好的效果，模型性能的提升最大，说明本文提出的协同训练算法学习效率最高；(2)Standard Co-training随机选择添加样本的方法造成了模型极强的不稳定性，CoTrade Co-training可以有效提升协同训练效果，但是每次迭代只添加置信度高的样本限制了模型

的泛化能力。本文方法与两者相比有显著提升，证明了本文使用强化学习智能体来自动学习一种样本选择策略，替代传统的预先定义的启发式样本选择策略方法的有效性。强化学习智能体可以对样本空间进行充分探索，选取更高质量的无标注数据，不仅可以帮助改善Standard Co-Training算法在随机挑选样本时发生的采样偏移问题，还可以改善CoTrade Co-Training算法由于局部采样造成的对模型泛化能力的限制。

此外，为了验证本文方法的鲁棒性，我们设计了如下实验：首先，使用原始数据划分训练强化学习智能体Q-agent。在测试时，随机生成另外五个训练集，并将剩余数据按原比例划分为测试集和无标注数据集，使用Q-agent已经学到的样本选择策略对两个模型重新进行协同训练，将两个模型集成后在测试集上进行测试，结果如表3所示。结果表明，本文方法对不同的初始化训练集具有鲁棒性，我们模型中的强化学习智能体Q-agent可以学习到一个良好的鲁棒的样本选择策略来选择高质量的无标注子集，以帮助协同训练过程。

表3 鲁棒性分析实验结果

Tab.3 Experimental results of robustness analysis

F_1值	最优值	最差值	平均值	标准差
人民日报	0.879	0.863	0.870	0.0068
金融新闻	0.861	0.838	0.845	0.0073

5 结论(Conclusion)

本文提出了一种基于强化学习的协同训练框架，在少量标注数据的情况下，无须人工参与，利用大量无标注数据自动提升模型性能。框架中的强化学习智能体可以学习一种良好的样本选择策略，选择高质量的无标注数据进行协同训练。我们在两种不同领域的语料上对模型进行了评估，实验结果表明本文方法性能优于其他的协同训练算法。我们还对强化学习智能体进行了测试，证明了学习到的样本选择策略对不同的初始化训练集和数据划分具有鲁棒性。在未来的研究中，我们计划将本文提出的框架扩展应用到其他不同类型的任务中去。

参考文献(References)

[1] Grishman R,Sundheim B.Message Understanding conference-6:a brief history[C].Proceedings of the 16th International Conference on Computational Linguistics,1996:466-471.

[2] Finkel J R,Grenager T,Manning C.Incorporating non-local information into information extraction systems by Gibbs sampling[C].Proceedings of the 2005,43rd Annual Meeting of the Association for Computational Linguistics,2005:363-370.

[3] Lafferty J,McCallum A,Pereira F.Conditional random fields:probabilistic models for segmenting and labeling sequence data[C].Proceedings of the 18th International Conference on Machine Learning,2001:282-289.

[4] Collobert R,Weston J,Bottou L.Natural language processing(almost) from scratch[J].Journal of Machine Learning Research,2011,12:2493-2537.

[5] Chiu J P C,Nichols E.Named entity recognition with bidirectional LSTM-CNNs[J].Transactions of the Association for Computational Linguistics,2016(4):357-370.

[6] Chapelle O,Scholkopf B,Zien A.Semi-supervised learning (chapelle,o.et al.,eds.;2006)[book reviews][J].IEEE Transactions on Neural Networks,2009,20(3):542-542.

[7] Blum A,Mitchell T.Combining labeled and unlabeled data with co-training[J].Proceedings of the eleventh annual conference on Computational learning theory,1998:92-100.

[8] Zhang R,Rudnicky A I.A new data selection principle for semi-supervised incremental learning[C].IEEE,2006(2):780-783.

[9] Minh V,Kavukcuoglu K,Silver D.Human-level control through deep reinforcement learning[J].Nature,2015,518(7540):529-533.

[10] LIAO W,Veeramachaneni S.A simple semi-supervised algorithm for named entity recognition[C].Naacl Hlt Workshop on Semi-supervised Learning for Natural Language Processing,2009.

[11] Aryoyudanta B,Adji T B,Hidayah I.Semi-supervised learning approach for Indonesian named entity recognition (NER) using co-training algorithm[C].International Seminar on Intelligent Technology & Its Applications IEEE,2017.

[12] XIAO K.Chinese organization name recognition based on co-training algorithm[C].International Conference on Intelligent System & Knowledge Engineering IEEE,2008.

[13] ZHANG R,Rudnicky A I.A new data selection principle for semi-supervised incremental learning[C].18th International Conference on Pattern Recognition,IEEE Computer Society,2006:780-783.

[14] Chawla N V,Karakoulas G.Learning from labeled and unlabeled data:An empirical study across techniques and domains[J].Journal of Artificial Intelligence Research,2005,23:331-366.

[15] Mnih V,Kavukcuoglu K,Silver D,et al.Human-level control through deep reinforcement learning[J].Nature,2015,518(7540):529.

[16] Silver D,Huang A,Maddison C J.Mastering the game of Go with deep neural networks and tree search[J].Nature,2016,529(7587):484-489.

[17] Yu K,Jia L,Chen Y.Deep learning:yesterday,today,and tomorrow[J].Journal of computer Research and Developme nt,2013,50(9):1799-1804.

[18] Sutton R,Barto A.Reinforcement learning:An Introduction[M].MIT Press,1998.

[19] Lange S,Riedmiller M.Deep auto-encoder neural networks in reinforcement learning[C].The 2010 International Joint Conference on Neural Networks,2010,1-8.

[20] Rajaraman A,Ullman J D.Finding similar items[J].Mining of Massive Datasets,2010,77:73-80.

[21] Watkins C J C H,Dayan P.Q-learning[J].Machine learning,1992,8(3-4):279-292.

[22] Zhang M L,Zhou Z H.CoTrade:confident co-training with data editing[J].IEEE Transactions on Systems,2011,41(6):1612-1626.

作者简介：

程钟慧(1995-)，女，硕士生.研究领域：自然语言处理.

陈　珂(1977-)，女，博士，副教授.研究领域：时空数据库，数据挖掘，数据隐私保护.

陈　刚(1973-)，男，博士，教授.研究领域：大数据管理.

徐世泽(1973-)，男，本科，高级工程师.研究领域：电力系统及自动化.本文通讯作者.

傅丁莉(1988-)，女，本科，工程师.研究领域：通信技术.