

面向事件检测的预训练主动学习模型

冯琳慧, 乔林波, 阚志刚

(国防科技大学并行与分布处理国家重点实验室, 湖南 长沙 410073)

[摘要] 深度学习在事件检测任务上取得了显著的成果,但模型严重依赖于大量的标注数据. 由于事件结构化的信息和丰富的标签表示,使得获取注释的成本很高,难以大量获得. 针对事件检测任务,为了提高语料标注效率,减少训练过程所需的标注样本数量,提出一种联合主动学习和预训练模型的事件检测模型. 针对主动学习模型存在的冷启动问题,设计了基于融合不确定性的特殊样本选择策略,估计样本在微调下游事件检测任务方面的潜在贡献. 一方面,结合预训练模型从原始任务中带来的丰富的语义信息,避免了重新设计网络结构或从零开始训练;另一方面,利用主动学习选择信息丰富的样本能更好地微调预训练模型,减少数据标注成本. 在 ACE 2005 语料上进行数值实验验证,结果证明了所提出的 EDPAL 算法的有效性.

[关键词] 主动学习, 事件检测, 预训练模型, 样本选择策略, 微调

[中图分类号] O643; X703 **[文献标志码]** A **[文章编号]** 1672-1292(2022)02-0041-07

Design and Implementation of a Pretraining Active Learning Model for Unstructured Event Detection

Feng Linhui, Qiao Linbo, Kan Zhigang

(National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha 410073, China)

Abstract: With the rapid growth of network information, it has become more and more important to find the key information. Event detection focuses on extracting event triggers from unstructured natural language texts. Deep learning has achieved a great success in event detection tasks, but the model relies on a large amount of labeled data which are difficult to be obtained. And the cost of obtaining annotations is very high due to the structured information of the event and the rich label representation. To address these issues, this paper proposes a joint active learning and pre-trained event detection model (EDPAL). To handle the cold start problem of the active learning, a special sample selection strategy on the basis of fusion uncertainty is designed to estimate the potential contribution of samples in fine-tuning downstream event detection tasks. On the one hand, combined with the rich semantic information brought by the pre-training model from the original task, it avoids redesigning the network structure or training from scratch. On the other hand, the pre-training model can be better fine-tuned by selecting information-rich samples and reduce the cost of data labeling at the same time. The experimental results on the ACE 2005 corpus shows the effectiveness of the proposed EDPAL.

Key words: active learning, event detection, pre-trained model, selecting strategy, fine-tuning

事件抽取是信息抽取领域的一个重要研究方向,同时也是该领域最具挑战性的课题之一. 根据事件提取基准 ACE2005 的定义,事件抽取任务可分为两个子任务,即事件检测(识别并分类事件触发器)和参数提取(识别事件触发器的参数并标记其角色). 其中,事件检测任务是事件抽取中最基本、最重要的任务. 最近的工作表明,深度学习模型很大程度上解决了事件检测问题,减轻了昂贵的特征工程的必要性. 然而,深度学习模型需要大量的训练数据. 由于事件的结构化信息和丰富的标签表示,标注语料通常需要领域专家进行人工标记. 而让专家对大量实例进行标记的过程是极不经济的,且标注效率难以适应以深度学习为代表的学习模型对标注语料的巨大需求. Hong 等^[1]表明不同的标注者对于事件抽取标注任务的一致性差异很大,难以保证高水平语料标注质量. 且事件抽取语料也极为稀疏,在 ACE2005 事件抽取

语料中共包含 16 375 条句子,其中含有事件的句子仅为 3 966 条,约占 24.2%。

针对上述问题,本文提出面向非结构化文本事件检测的预训练主动学习模型(EDPAL)。一方面,在大规模开放域中训练得到的预训练模型能够很好地应用在下游任务中,并减小标注代价,提升事件模型在训练语料不足情况下的训练效果。另一方面,将构造的预训练事件检测模型作为主动学习框架中的学习模型,进一步针对事件检测任务设计了融合不确定性的特殊样本选择策略,选择信息丰富的样本进行人工标注,提高标注效率的同时,也更好地微调了预训练模型。

数值实验中,采用 ACE2005 英文语料集进行实验,实验结果表明:联合主动学习的预训练的事件监测模型效果优于传统模型;本文所提出的融合不确定性的特殊样本选择策略,优于不确定性选择策略。

1 相关工作

传统的事件抽取任务处理方法^[1-3]主要依赖于精心设计的特征和复杂的自然语言处理工具,这些方法的性能在很大程度上依赖于能否人工设计出高质量的特征信息,导致性能无法进一步提升。

近年来,基于深度学习的方法被广泛应用于事件抽取中。为了改善事件信息在神经网络层间的传递,Chen 等^[4]改进了传统 CNN 模型,添加了动态多池机制。为了改进神经网络对不定长度文本的事件抽取,Feng 等^[5]结合双向 LSTM 和卷积神经网络,通过捕获文本的序列信息和短语块信息,进行了事件抽取。Nguyen 等^[6]使用联合循环神经网络的方法,同时进行事件触发词和事件元素的识别和分类,减少了因预测事件触发词错误而导致预测事件元素出错的错误级联问题。在基于深度学习的方法中,不再使用复杂的自然语言处理工具进行特征提取,使得事件抽取任务性能得到进一步提升。

使用深度学习的方法需要大量的训练数据,即使是广泛使用的 ACE2005 语料库,少于 20 个标注的实例的事件类型也占到了 25% 左右,存在着样本总量少、类别不平衡的问题。主动学习^[7]是一种通过一定选择策略筛选样本标注训练的学习方法。已有研究表明,使用标注数据集训练可以得到一个错误率小于 ε 的分类器,传统的监督学习算法所用样本的复杂度为 $O\left(\frac{1}{\varepsilon} \ln\left(\frac{1}{\varepsilon}\right)\right)$,而使用主动学习方法^[8]能够将样本复杂度减少到 $O\left(\ln\left(\frac{1}{\varepsilon}\right)\right)$ 。因此,主动学习作为一种重要的提升模型训练效果、减少样本数量的学习方法,已被广泛应用于各项自然语言处理任务中。

在未标记数据可以大量获取而标注数据难以大量获取的事件检测任务中,早期 Liao 等^[9]基于特征的方法提出 Co-Testing 主动学习方法,邱盈盈等^[10]提出将深度学习方法与主动学习方法相结合,通过将 RNN 模型嵌入到主动学习框架中,实现了对事件检测任务的性能提升。然而,仅利用深度学习进行文本隐含特征信息的提取,在缺乏训练数据的基础上很难获得充分的学习训练。近年来,预训练词向量的方法^[11-13]得到了广泛关注,在下游任务中使用在大量语料上进行预训练的模型,可以很好地提取到文本语义信息。目前尚未有工作将其与主动学习方法结合来解决事件检测样本不足的问题。Huang 等^[14]将主动学习和基于预训练的 CNN 模型结合来降低标注任务的工作量,实验结果证明了这种方法在生物医学图像上的有效性。

2 面向事件检测的预训练主动学习框架

联合预训练模型的主动学习事件检测框架(EDPAL)的设计包括 EDPAL 的框架、样本选择策略和事件检测学习模型。

2.1 EDPAL 框架

主动学习将迭代地从未标注数据里根据设定的选择策略筛选出一部分重要的数据去标注,从而获得更多标记数据。传统的主动学习框架可抽象为以下 5 个部分:

$$A = (C, L, S, Q, U), \quad (1)$$

式中, C 表示主动学习中待训练的学习模型; L 为已标记样本集; Q 为不同的选择策略函数,旨在从未标记样例中筛选出高信息量的待标注样本; U 为整个未标记样本池; S 为标注者。在 EDPAL 中将数据集划分为未标记数据集 UL 、已标记数据集 L 和测试集 TS 。详细过程参见算法 1。

算法 1 EDPAL 算法

输入:

- L , 已标注数据集;
- UL , 未标注数据集;
- N , 每次选择标注的样本数量;
- T , 终止条件;
- M , 预训练学习模型.

输出:

- M' , 微调后的预训练学习模型.

迭代: 满足终止条件 T 为止

1. 利用已标注数据集 L 微调预训练学习模型 M ;
2. 利用微调后的模型 M 进行测试评估;
3. 根据评估函数 S , 在未标记数据集 UL 中选择 N 个样本;
4. 得到 N 个样本的正确分类类型;
5. 从 UL 中删除所选的 N 个样本, 已标注数据集 L 中添加标注后的 N 个样本.

2.2 事件检测学习模型

不同于传统的主动学习框架, EDPAL 将经过大量语料预训练的模型作为学习模型, 对新构建的训练数据集进行迭代学习.

2.2.1 事件检测任务形式化

定义事件句 $x = \{x_1, x_2, \dots, x_n\}$ 对应的标签序列为 $y = \{y_1, y_2, \dots, y_n\}$, 其中, x_1 为事件提及句子中的单词. 本文使用 BIO 的标签格式, 事件检测模型预测出最佳的标签序列 \hat{y} .

2.2.2 基于 BERT 的事件检测

针对事件检测任务, 考虑到词性对于触发词的影响, 本文在 BERT^[1] 的初始输入上改进词向量的初始编码, 对于每个词添加了其词性向量编码, 即词向量 $E(x)$ 为由 BERT 本身的编码 E_{BERT} 和词性编码 E_{POS} 拼接而成:

$$E(x) = [E_{\text{BERT}}; E_{\text{POS}}]. \quad (2)$$

利用一个简单线性的 CRF 分类器进行分类:

$$P(y|x; \mathbf{A}) = \frac{1}{Z(x)} \exp \left(P(y_1; x_1) + \sum_{k=1}^{n-1} P(y_{k+1}; x_{k+1}) + \mathbf{A}_{y_k, y_{k+1}} \right), \quad (3)$$

式中, $Z(x)$ 是 x 的所有可能标记的归一化因子; $P(y_{k+1}; x_{k+1})$ 表示第 $k+1$ 处插入 y_{k+1} 标记的概率; \mathbf{A} 是转移矩阵参数.

最可能的标签序列 \hat{y} 可表达为:

$$\hat{y} = \arg \max_y P(y|x). \quad (4)$$

2.3 融合不确定性的特殊样本选择策略

EDPAL 不断选择出对模型训练分类有帮助的样例交给标注者进行标注, 主动地迭代式微调预训练模型. 本文提出一种融合不确定性的特殊样本选择策略, 从不确定性和特殊性两个方面衡量样本的价值, 并针对不同时期模型分配不同的注意权重. 定义事件句 x 的价值函数为:

$$\text{Score}(x) = (1 - \alpha \cdot t) D(x) + (\alpha \cdot t) U(x), \quad (5)$$

式中, t 为模型迭代次数; α 为权重; $D(x)$ 表示基于特殊性的选择策略得分; $U(x)$ 代表基于不确定性的选择策略得分.

2.3.1 基于特殊性的事件句选择算法

诸如 BERT 的预训练模型, 在大规模数据上学习通用的语义信息. 当运用在下游任务时, 针对特定的任务微调, 就能获取到良好的性能. 文献[15]表明, BERT 的前几层 transformer 主要捕获语言的通用语义信息, 微调主要影响了顶层的 transformer. 且对于跨领域的不同任务, 微调的效果仍有很大的提升空间.

Liao 等^[9]提出的基于 Co-testing 的主动学习方法, 其选择策略建立在传统的特征工程设计模型, 而传

统的基于特征模型很难捕获历史信息,且泛化能力弱。邱盈盈等^[10]提出的 JRNN_AL 主动学习方法,其选择策略建立在 Jordan-RNN 模型上,由于深度学习模型在初期可能过度自信,使得基于不确定性的选择方法错误地选择冗余样本,误导模型。本文设计了针对预训练模型的基于特殊性选择策略,利用预训练模型不同学习层在大规模语料中学习到的语义信息,基于特殊性的选择算法利用样本在不同预训练层的输出表示的差异性 $D(x)$,

$$D(x) = \text{dist}(E_k(x), E_{k+n}(x)), \quad (6)$$

来衡量待选择样本的价值,而不受模型初期预测准确率低的影响。其中, $E_k(x)$ 表示事件句 x 在第 k 层的输出, $E_{k+n}(x)$ 表示事件句 x 在第 $k+n$ 层的输出。dist 为距离衡量函数,距离越大,表明预训练 A 层和预训练 B 层对该样本的表示差异越大。

2.3.2 基于不确定性的事件句选择算法

基于最大不确定性事件句选择算法,选择模型预测的最不确定^[16]样本(LC)进行人工标注,并加入训练集来加强模型对该类样本的训练程度。根据最可能的标签序列进行建模:

$$U^{\text{LC}}(x) = 1 - P(\hat{y}|x; \mathbf{A}), \quad (7)$$

式中, \hat{y} 表示模型对事件句 x 预测概率最大的标签序列。LC 算法选择模型预测最不确定的样本 x_{LC}^* 进行人工标注,尽管这种方法存在着更容易选择长句的缺点,但现有许多工作已证明了该方法的有效性。

基于最大不确定性选择算法只考虑了模型对样本预测最不确定的一类,而忽略了属于其他类别的比重^[17],若在考虑样本最可能属于的序列的同时,还考虑其第二可能属于的序列的情况,则可定义如下:

$$U^{\text{MC}}(x) = -(P(\hat{y}_1|x; \mathbf{A}) - P(\hat{y}_2|x; \mathbf{A})), \quad (8)$$

式中, \hat{y}_1 和 \hat{y}_2 分别表示模型预测样本 x 最大和次大可能所属于某序列的概率。该差值越大,说明模型不容易将其分错;该差值越小,证明模型很难判断其所属的真正序列。

与现有策略不同,考虑序列信息和单词信息都具有特有的优势,两者可以互补,本文使用最低令牌(LTP)不确定性的选择策略^[18],该策略选择在最可能的标签序列下概率最低的单词:

$$U^{\text{LTP}}(x) = 1 - \min_{\hat{y}_i \in \hat{y}} P(\hat{y}_i|x_i; \mathbf{A}). \quad (9)$$

3 实验与结果

3.1 数据集与度量标准

实验选用在事件抽取任务中被广泛使用的 ACE2005 英文语料作为实验数据集,数据集包含 599 篇文档,其中预先定义了 8 大类、33 个子类的事件。与已有的事件抽取研究一致,在进行触发词抽取分类任务时,将不考虑大类与子类的结构关系,并设置非触发词类型为 None 类型,共 34 个类别,直接对 34 个子类进行分类。与文献[4-6, 19]的实验设置相同,本文选取 599 篇文档中固定的 40 篇作为测试集,剩下 559 篇(15 703 句)作为未标注集进行实验,每次迭代选取的标注样例数量为 100。设置迭代次数为 60 次, α 取 0.01。选择倒数第 4 层的特征输出和最后一层进行比较,利用余弦距离函数衡量相似度,并进行归一化处理。

触发词抽取的性能评价标准与文献[4-6, 19]相同:(1)触发词识别:若一个触发词实例的位置和标注文档一致,则认为该触发词实例识别正确。(2)触发词分类:若一个触发词实例的位置和类别都与标注文档一致,则认为该触发词分类正确。本文采用通用的评价指标:准确率(P)、召回率(R)和 F 值($F1$)对事件触发词抽取性能进行评定, P 、 R 、 $F1$ 的数值越高意味着模型效果越好。

主动学习方法的性能评价标准采用 Schein 等^[20]提出的差异值区分不同模型的主动学习方法的性能,表示为:

$$D(A, B) = \frac{\sum_{i=1}^n (F_n(B) - F_i(A))}{\sum_{i=1}^n (F_n(B) - F_i(B))}, \quad (10)$$

式中, n 表示主动学习总迭代次数; $F_i(B)$ 表示在第 i 次迭代中,基准系统 B 的 $F1$ 值; A 为做比较的主动学习系统。差异值越小,说明主动学习系统 A 相较于 B 性能提升越大。

3.2 基线实验设置

为了对比本文提出的 EDPAL 模型的性能,设置以下对比系统:

- (1) BERT_Random: 使用基于 BERT 的事件检测模型, 主动学习策略选择随机选取方法;
- (2) BERT_LC: 使用基于 BERT 的事件检测模型, 主动学习策略选择基于最大不确定性的选择策略;
- (3) BERT_LTP: 使用基于 BERT 的事件检测模型, 主动学习策略使用基于最低令牌概率的选择策略;
- (4) BERT_Dist: 使用基于 BERT 的事件检测模型, 主动学习策略选择基于特殊性的选择策略;
- (5) BERT_DLTP: 使用基于 BERT 的事件检测模型, 主动学习策略选择融合最低令牌不确定性的特殊样本选择策略。

3.3 实验结果与分析

首先在 ACE2005 英文语料上对学习模型进行单独的触发词分类数值, 实验结果如表 1 所示。DiBERT 是 BERT 预训练模型的简化版本。在触发词识别任务上, 基于 BERT 的事件检测模型与 Liao 等^[9]、Li 等^[2]的研究模型和 JRNN-AL^[10]相比, $F1$ 值分别提高了 12.2%、10.0% 和 9.1%; 在触发词分类任务上, $F1$ 值分别提高了 11.1%、8.8% 和 7.9%。

在触发词分类任务上, 联合主动学习的 BERT 预训练事件检测模型 (BERT_DLTP) 在 BERT 的基础上又提高了 0.9%。说明联合预训练模型和主动学习的方法能够寻找到有利于提升分类效果的样本。实验迭代 60 次, 仅选出 6 000 个样本进行标注, 相较于使用全部的 15 703 句样本而言, 提升了样例的标注效率, 在较小规模的事件检测模型上获得了更好的性能。在不同的预训练模型上, 加入融合不确定性的特殊性选择策略后, 模型性能分别提高。因此, 本文提出的主动学习方法是有效的。

表 1 事件检测任务实验结果

Table 1 The Results of event detection

| 模型 | 触发词识别 | | | 触发词分类 | | |
|-------------|--------|--------|--------|--------|--------|--------|
| | $P/\%$ | $R/\%$ | $F/\%$ | $P/\%$ | $R/\%$ | $F/\%$ |
| Liao's | 76.2 | 61.7 | 68.2 | 74.3 | 58.1 | 65.2 |
| Li's | 76.9 | 65.0 | 70.4 | 73.7 | 62.3 | 67.5 |
| JRNN | 79.7 | 63.7 | 70.8 | 75.3 | 61.7 | 67.8 |
| JRNN_AL | 82.3 | 62.9 | 71.3 | 77.4 | 61.3 | 68.4 |
| DiBERT | 78.7 | 80.6 | 79.6 | 74.5 | 76.3 | 75.4 |
| DiBERT_DLTP | 79.2 | 77.5 | 78.3 | 76.8 | 75.1 | 75.9 |
| BERT | 81.2 | 79.6 | 80.4 | 77.1 | 75.6 | 76.3 |
| BERT_DLTP | 82.0 | 78.0 | 80.0 | 79.8 | 76.0 | 77.2 |

图 1 给出了在不同选择策略下预训练模型的事件检测性能曲线。其中, 横轴表示主动学习中选择样本的迭代次数, 纵轴表示事件检测任务的 $F1$ 值 (触发词分类结果)。从图 1 可以看出, 基于不确定性的一类选择策略在模型初期表现差于特殊样本选择。这是因为依赖初期的模型衡量不确定性, 模型本身的准确性无法保证, 选择的样本与随机挑选的样本差异不大。

图 2 给出了在不同预训练模型下的事件检测性能曲线。其中, 横轴表示主动学习中选择样本的迭代次数, 纵轴表示事件检测任务的 $F1$ 值 (触发词分类结果)。实验表明, 主动学习选择策略在不同的预训练模型上是有效的。

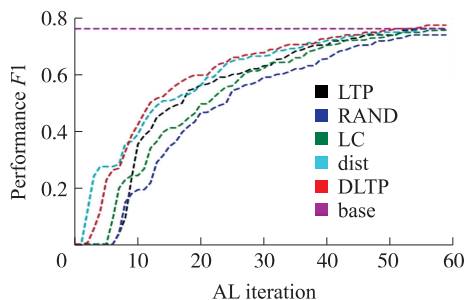


图 1 模型 $F1$ 值对比结果图

Fig. 1 Performance comparison on different selection strategies

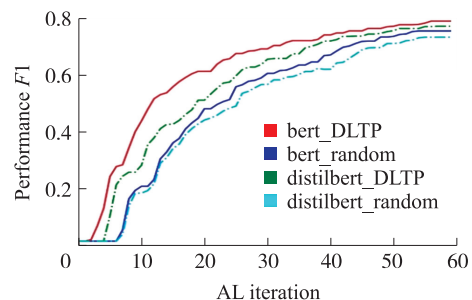


图 2 不同预训练模型 $F1$ 值对比结果图

Fig. 2 Performance comparison on different pre-training models

此外,本文计算了不同系统的差异值来衡量不同系统的性能差异,如表 2 所示. 设置基线系统为 BERT_Ran,差异值越小说明该系统性能更好. 从实验系统差异值分析,在同一预训练模型下,对比随机选择样本,融合不确定性的特殊样本选择都能得到较好的系统性能.

表 2 实验系统差异值
Table 2 Difference values on the experimental system

| 实验系统 | 触发词识别 | 触发词分类 | 实验系统 | 触发词识别 | 触发词分类 |
|-----------|-------|-------|-------------|-------|-------|
| BERT_Ran | 1.000 | 1.000 | BERT_DLTP | 0.563 | 0.586 |
| BERT_LC | 0.785 | 0.856 | DiBERT_Ran | 1.102 | 1.099 |
| BERT_LTP | 0.762 | 0.783 | DiBERT_DLTP | 0.803 | 0.806 |
| BERT_Dist | 0.602 | 0.603 | | | |

4 结论

本文面向非结构化文本事件检测任务,提出 EDPAL 主动学习框架并得到了良好的模型效果. 分析了多种样本选择策略,设计了融合不确定性的特殊样本选择. 实验结果显示,联合主动学习的方法能够提高模型性能. 在达到相同的模型性能时,本文所设计的选择策略使用的训练集样本更少. 对于数据类别分布不均的 ACE 语料数据集,模型性能也更趋于稳定. 由于 ACE 数据的标注质量不高及规模较小,在很大程度上影响了事件抽取任务的效果,在未来的工作中将进一步构建数据集,将模型推广到实际运用中.

[参考文献] (References)

[1] HONG Y,ZHANG J F,MA B,et al. Using cross-entity inference to improve event extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, USA: Association for Computational Linguistics,2011.

[2] LI Q,JI H,HUANG L. Joint event extraction via structured prediction with global features[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia,Bulgaria:ACL,2013.

[3] 吴家皋,周凡坤,张雪英. HMM 模型和句法分析相结合的事件属性信息抽取[J]. 南京师大学报(自然科学版),2014, 37(1):30-34.

[4] CHEN Y B,XU L H,LIU K,et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Beijing,China:ACL,2015.

[5] FENG X C,QIN B,LIU T. A language-independent neural network for event detection[J]. Science China(Information Science),2018,61(9):81-92.

[6] NGUYEN T H,CHO K,GRISHMAN R. Joint event extraction via recurrent neural networks[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association Computation Linguistics: Human Language Technologies. San Diego,USA:NAACL,2016.

[7] REN P Z,XIAO Y,CHANG X J,et al. A Survey of Deep Active Learning[J]. arXiv preprint arXiv:2009.00236,2020.

[8] SEUNG H S,OPPER M,SOMPOLINSKY H. Query by committee[C]//Proceedings of the fifth Annual Workshop on Computational Learning Theory. Pittsburgh,USA:ACM,1992.

[9] LIAO S S,GRISHMAN R. 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. Chiang Mai,Thailand:ACL,2011.

[10] 邱盈盈,洪宇,周文瑄,等. 面向事件抽取的深度与主动联合学习方法[J]. 中文信息学报,2018,32(6):98-106.

[11] DEVLIN J,CHANG M W,LEE K,et al. Bert:pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis,USA:ACL,2019.

[12] PETERS M,NEUMANN M,IYYER M,et al. Deep contextualized word representations[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, USA:ACL,2018.

[13] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]//Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach,USA:CAI,2017.

- [14] HUANG S J,ZHAO J W,LIU Z Y. Cost-effective training of deep CNNs with active model adaptation[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London,UK:ACM,2018.
- [15] MERCHANT A,RAHIMTOROGHI E,PAVLICK E,et al. What happens To BERT embeddings during fine-tuning? [J]. arXiv preprint arXiv:2014.14448,2020
- [16] MARTINEZ-CANTIN R,DE FREITAS N,DOUCET A,et al. Active policy learning for robot planning and exploration under uncertainty[C]//Proceedings of Robotics:Science and Systems III. Atlanta,USA:MIT Press,2007.
- [17] SCHEFFER T,DECOMAIN C,WROBEL S. Active hidden Markov models for information extraction[C]//Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis. Berlin Germany:Springer,2001.
- [18] LIU M Y,TU Z Y,ZHANG T,et al. LTP:A New Active Learning Strategy for CRF-Based Named Entity Recognition[J]. arXiv preprint arXiv:2001.02524,2020.
- [19] LIU J,CHEN Y B,LIU K. Exploiting the ground-truth:an adversarial imitation based knowledge distillation approach for event detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2019,33(1):6754-6761.
- [20] SCHEIN A I,UNGAR L H. Active learning for logistic regression:an evaluation[J]. Machine Learning,2007,68(3):235-265.

[责任编辑:严海琳]

(上接第34页)

- [15] 葛尧,陈松灿. 面向推荐系统的图卷积网络[J]. 软件学报,2020,31(4):1101-1112.
- [16] TANG J X,WANG K. Personalized top-n sequential recommendation via convolutional sequence embedding[C]//Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey,USA:ACM,2018.
- [17] 燕存,吉根林. Item-Based 并行协同过滤推荐算法的设计与实现[J]. 南京师大学报(自然科学版),2014,37(1):71-75.
- [18] LIU Q,YU F,WU S,et al. A convolutional click prediction model[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne,Australia:ACM,2015.
- [19] 王俊淑,张国明,胡斌. 基于深度学习的推荐算法研究综述[J]. 南京师范大学学报(工程技术版),2018,18(4):33-43.
- [20] GUO C,LI B,TIAN X M. Flickr group recommendation using rich social media information[J]. Neurocomputing,2016,204:8-16.

[责任编辑:严海琳]