

基于 ELECTRA 模型与词性特征的金融事件抽取方法研究^{*}

陈星月 倪丽萍 倪志伟
(合肥工业大学管理学院 合肥 230009)
(合肥工业大学过程优化与智能决策教育部重点实验室 合肥 230009)

摘要：【目的】针对金融事件抽取中金融事件实体边界模糊、抽取不准确的问题，提出基于预训练模型 ELECTRA 和词性特征的金融事件抽取方法。【方法】为增强模型对金融关键实体的感知力，充分考虑语料原始的语义信息以及词性特征信息，将语料分别通过两个 ELECTRA 预训练模型后进行融合操作，达到增强语义的效果；将学习到的信息传入 BiGRU 中，获取上下文长距离的语义依赖，输出原始的序列标签；利用 CRF 克服标签偏差问题，通过上述步骤完成金融事件抽取。【结果】基于预训练模型 ELECTRA 和词性特征的金融事件抽取方法在金融事件数据集上 F_1 值达到 70.96%，比经典的抽取模型 BiLSTM-CRF 性能提升 20.74 个百分点。【局限】数据集中事件数较少，预训练模型体积较大，会受到 GPU/TPU 内存的限制。【结论】本文模型能够更加全面地捕捉金融事件元素之间的联系，提升金融事件抽取的效果。

关键词：ELECTRA 词性特征 金融事件抽取 预训练模型

分类号：TP183

DOI：10.11925/infotech.2096-3467.2020.1296

引用本文：陈星月，倪丽萍，倪志伟. 基于 ELECTRA 模型与词性特征的金融事件抽取方法研究[J]. 数据分析与知识发现，2021，5(7)：36-47. (Chen Xingyue, Ni Liping, Ni Zhiwei. Extracting Financial Events with ELECTRA and Part-of-Speech[J]. Data Analysis and Knowledge Discovery, 2021, 5(7): 36-47.)

1 引言

事件抽取(Event Extraction, EE)，是文本数据挖掘中一项重要又具有挑战性的任务，主要是从描述事件的相关文本中学习事件的结构化表示^[1]。

金融事件抽取是事件抽取技术在金融领域的应用。在金融领域，投资者的决策受到各种因素的影响，如公司自身新闻、政治政策以及宏观经济等因素^[2]，这些因素往往以文本的形式呈现给大众，并且大多数文本中所包含的事件信息会成为影响金融领域市场状态的主要因素。因此，金融事件抽取能够

帮助投资者获取公司主要事件，识别投资风险以及投资机会，预测股市走向，做出正确的投资决策。此外，金融事件中的关键实体可以帮助投资者更加迅速地获取事件的主题，从而对金融市场有更为全面的了解。

随着信息技术的快速发展，金融文本数量呈指数级增长，而金融领域文本有别于其他领域，具有时效性高、句法结构复杂、高维稀疏、语义模糊和无规律性等特点^[3]，需要设计有效的方法理解金融文本中深层的语义信息以及上下文对应关系。此外，金融文本中往往包含大量噪声文本和无关金融实体，

通讯作者(Corresponding author)：倪丽萍(Ni Liping)，ORCID:0000-0002-7067-302X，E-mail: niliping@hfut.edu.cn。

^{*}本文系国家自然科学基金青年项目(项目编号：71301041)、国家自然科学基金重大研究计划培育项目(项目编号：91546108)和国家自然科学基金青年项目(项目编号：71701061)的研究成果之一。

The work is supported by the National Natural Science Foundation of China (Grant No. 71301041, 91546108, 71701061).

大大增加了抽取金融事件关键实体的难度。因此,如何从金融文本中准确快速地挖掘出关键并且能为投资者所利用的事件信息,成为亟待解决的问题。

2 相关工作

近年来,越来越多的研究者开始关注金融领域事件的抽取,主要分为基于模式匹配的方式和基于机器学习的方式。

基于模式匹配的方式是根据领域专业人员构建的模板进行事件抽取。如 Feldman 等^[4]邀请领域专家构建了金融领域的情感词典、制定短语级的模式进行情感极性匹配,制定事件抽取的规则。Arendarenko 等^[5]构建了一个知识库,首先对企业商务文本进行词性标注、形态分析以及语义分析等预处理,再使用基于规则的方式,查找事件实体,进行事件抽取。Hogenboom 等^[6]构建了一套基于语义的事件检测方法,对新闻文本进行分词、词性标注、形态分析等预处理,使用知识库(如 WordNet)从语料中进行模式匹配抽取金融事件,并利用事件抽取中学到的知识进一步更新知识库。

金融事件抽取研究早期大多都是基于模式匹配的方式,但是这种方式需要花费大量时间、精力维护构建的模板,需要根据金融领域信息的更新及时更新规则,否则会抽取无用的信息。因而机器学习方式越来越受到研究者的关注,采用机器学习的方式即让模型自动地从金融语料的文本表示中学到特征,进而实现事件抽取。如 Ronnqvist 等^[7]从新闻中抽取金融风险事件,根据新闻文本的语义信息使用一层前馈拓扑网络进行预训练获得语义神经网络,再结合语义神经网络以及一层前馈神经网络对银行危机事件进行预测。Jacobs 等^[8]定义了 10 种经济事件,将多分类问题转化为二分类问题,使用 one-vs-rest 的方法,利用 SVM 和 LSTM 对经济事件进行检测。Yang 等^[9]通过远程监督生成金融文本数据扩充训练数据集,使用 BiLSTM-CRF 模型进行句子级的事件抽取后,再进行篇章级的事件抽取整合。Zheng 等^[10]提出一个端到端的模型 Doc2EDAG,使用三个 Transformer 模型^[11]将金融事件表转换为基于实体的有向无环图,以此解决篇章级事件抽取的难题。由于机器学习在事件抽取中取得的效果更好,本文采

用基于机器学习的方式进行金融领域的事件抽取。

预训练模型利用迁移学习的思想,在大规模的数据集上进行训练生成模型,再根据下游任务,使用较少的数据对预训练模型进行微调,有效提升了模型性能,加快模型训练收敛速度。预训练模型可以分为静态预训练模型如 Word2Vec^[12]、GloVe^[13]和动态预训练模型如 BERT^[14]、RoBERTa^[15]。由于动态预训练模型可以解决静态预训练模型无法解决的一词多义问题,学习到文本中的深层语义信息,因此近年来被广泛应用于事件抽取,并取得了较好的效果。如 Yang 等^[16]将事件抽取分为两个阶段,先使用预训练模型 BERT 进行触发词的抽取,再根据事件元素重要性重新加权损失函数,使用预训练模型 BERT 进行事件元素的抽取。Du 等^[17]将事件抽取看作阅读理解问题,使用两个基于预训练模型 BERT 的问答模型分别进行触发词抽取以及元素抽取。Liu 等^[18]使用逻辑回归模型抽取触发词,根据触发事件类型利用问题模板实现无监督问题生成,最后使用基于预训练模型 BERT 的问答模型进行事件元素抽取。Ein-Dor 等^[19]基于维基百科,将事件抽取任务转化为二分类问题,即判断新闻中是否存在与特定公司有关的“重要事件”:使用预训练模型 BERT-base 在新闻语料上进行训练,对生成的模型检测常见的金融事件,并在检测能力以及获取事件的多样性上进行对比,从而得到一个较好的金融事件检测模型。Li 等^[20]根据金融文本使用一个预训练模型 BERT 生成文档表示,进而将其运用到下游预测事件类型、文档聚类以及文档摘要的任务中。Zhao 等^[21]使用预训练模型 RoBERTa 进行情感分析、粗粒度的实体选择、细粒度的关键实体识别,从而实现在线金融文本事件抽取。

综上,前期的金融事件抽取大多利用知识库、情感极性构建模板进行事件抽取。近年来,运用预训练模型特别是基于 BERT 及其改进的模型进行金融事件抽取取得了较好效果,逐渐成为金融事件抽取的一种主流方法。但基于预训练模型的金融事件抽取方法目前仅利用预训练模型的自学能力,没有将语料中存在的知识融合到模型中,因此仍存在模型语义提取能力不强的问题。本文在使用预训练模型的基础上融入词性特征,以增强模型对关键实体的

抽取能力。词性特征是自然语言处理中比较重要的研究内容,不同词性的词汇在句子中发挥着不同的功能。在金融事件抽取中,名词、动词往往在句子中有重要的作用,对抽取事件元素有很大帮助,此外,词性自动标注的精度较高,噪声数据输入到模型的可能性较小。因此,利用预训练模型与词性特征,模型可以更好地感知上下文,提升抽取金融事件的性能。

3 模型介绍

3.1 问题定义

金融事件抽取问题可以描述为:从文本语料中识别是否存在相关的金融事件,再识别出相关事件的涉及元素,最后确定每个元素所扮演的角色^[22]。如下所示,将例句输入到事件抽取模型中,需要将E1、A1、A2、A3和A4抽取出来。其中E1称为触发词,A1、A2、A3和A4称为事件元素。

例句:美锦集团(A1)持有的上市公司28.37亿股股份(A2)中,已有27.8亿(A3)股处于质押(E1)状态,占其持股的97.98%(A4)。

例句中的触发词为“质押”,说明包含着一个质押事件,而抽取的元素A1、A2、A3和A4在质押事件中的元素角色分别为质押公司、质押物、质押数量、质押比例。

基于机器学习的事件抽取主要有两种方法,分别是管道式方法(The Pipelined Approach)和联合学习方法(The Joint Approach)。管道式方法在第一阶段进行触发词识别、事件类型确定,第二阶段进行事件元素的识别,即先将例句中的E1抽取出来,判断属于哪类事件,之后根据E1事件框架进行A1、A2、A3和A4的抽取。联合学习方法则同时提取触发词和事件元素,即将例句中E1、A1、A2、A3和A4同时抽取出来。由于管道式方法存在误差传播的现象,如果在第一阶段事件类型判断错误,会导致第二阶段事件元素抽取错误,因此本文采用联合学习方法,利用触发词和事件元素之间的联系,让模型更有效地学习到语料的局部特征,提高事件抽取的精度。本文将事件抽取看作序列标注任务,标注策略采用BIO策略,B表示事件元素的开始,I表示事件元素的中间或结尾词,O表示无关词。

3.2 整体流程框架

本文提出一种基于ELECTRA(Efficiently Learning an Encoder that Classifies Token Replacements Accurately)预训练模型^[23]和词性特征的金融事件抽取模型ELECTRA-POS-BiGRU-CRF模型,对金融事件中的相关要素进行抽取。首先对原始语料进行预处理,输入不同的ELECTRA预训练模型中,分别获取语料最初的语义信息以及通过外部工具包所输入的词性特征;然后将隐藏层向量点乘融合,进行语义增强,经过BiGRU获取ELECTRA输出信息中的上下文语义依赖,输出最初的序列标签;最后利用条件随机场(CRF)克服标签偏差。

单一的循环神经网络不能完全提取到语料中的事件信息,会丢失一些重要信息,而预训练模型在大量的数据上进行预训练,具有强大的提取特征能力。ELECTRA是基于预训练模型BERT的改进,可以学习到更好的文本表示。此外,通常的事件抽取模型没有充分利用语料中自带的词性特征信息,一般来说,词性为动词、名词的词语在事件描述中会使得后续模型获得的语义信息更加丰富。因此本文引入预训练模型ELECTRA和词性特征,通过一个预训练模型ELECTRA获取原始语料中的语义信息特征,通过另一个预训练模型ELECTRA获取带有外部知识的词性特征,利用预训练模型强大的迁移能力,不仅获取了语料中不同层次的语义信息,并且提取各种词性下的实体联系,增强模型对关键实体的抽取能力。

本文模型整体架构如图1所示。主要包括输入层、两个ELECTRA预训练模型层、BiGRU层、CRF层以及输出层。

3.3 输入层

设计了两个ELECTRA预训练模型,其中一个用于提取文本语义,另一个用于提取词性特征,针对两者的不同功能,预处理操作包括如下流程,如图2所示。

(1)文本分词:利用分词方法将文本分为一个字或者短语等词汇单位。由于需要将原始语料输入到两个不同的预训练模型中,因此进行两种不同的分词。

对于给定句子 $x = \{x_1, x_2, \dots, x_n\}$,其中 x_i 表示输

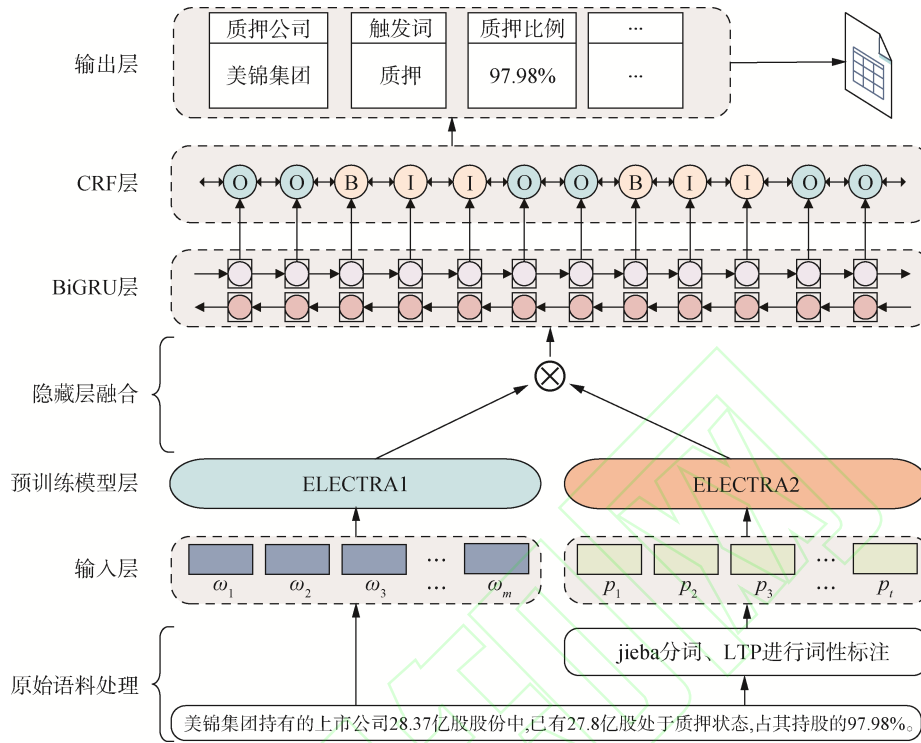


图1 模型整体架构

Fig.1 The Overall Architecture of Model

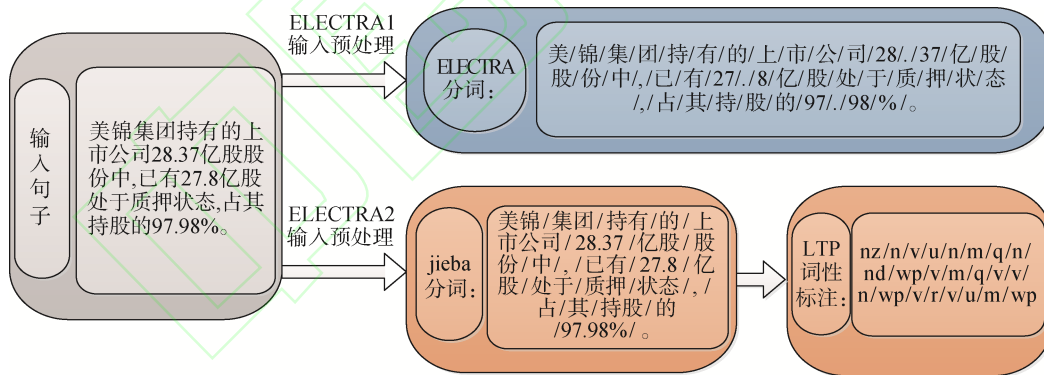


图2 文本预处理过程示例

Fig.2 Example of Text Preprocessing

入句子的第 i 个字, n 为句子所包含字的个数, 输入到 ELECTRA 1 层时, 使用的是预训练模型自带的分词器, 而这个分词器在处理中文时, 是以字进行分词的, 分词后, 在序列后补 0 至统一长度, 得到分词结果 $\omega_i \in \mathbf{R}^m (i = 1, 2, \dots, m)$, ω_i 是句子中第 i 个标记, m 为句子分词后序列的长度。输入到 ELECTRA 2 层时, 需要获取词汇的词性, 因此使用分词工具包 jieba

进行分词, 得到分词结果 $v_j \in \mathbf{R}^t (j = 1, 2, \dots, t)$, v_j 是句子中第 j 个标记, t 为利用 jieba 工具包分词后序列的长度。

(2) 词性标注: 使用哈尔滨工业大学的语言技术平台 (Language Technology Platform, LTP) 进行词性标注。LTP 中采用 863 词性标注集, 对于分词结果 v_j 进行标注得到词性标注结果 $p_k \in \mathbf{R}^t (k =$

$1, 2, \dots, t), p_k$ 是句子中第 k 个标记, t 为根据分词序列进行词性标注后序列的长度。

3.4 预训练模型层

基于预训练模型 ELECTRA^[23] 进行文本表示。ELECTRA 是基于 BERT 模型^[14] 的改进模型。BERT 模型主要架构为 Transformer^[11] 的 Encoder 部分, 使用多层 Encoder 部分结构堆叠而成, 如图 3 所示。

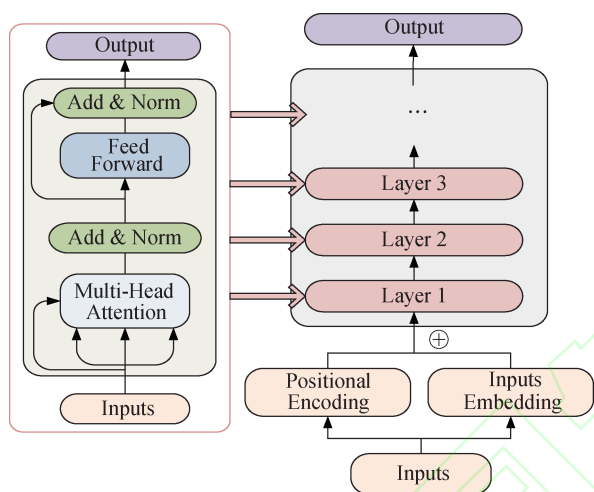


图3 BERT 模型架构

Fig.3 Architecture of BERT

ELECTRA 模型在 BERT 基础主要进行以下两点改进:

(1) ELECTRA 将 BERT 中 MLM 模型中的 [MASK] 标记变为由生成器生成的相对合理的标记。

(2) ELECTRA 训练一个鉴别器 (Discriminator), 区分生成器生成的标记正确与否, 如图 4 所示, 而后直接使用鉴别器作为下游任务的模型, 模型结构与 BERT 一致。

整体来说, 与 BERT 相比, ELECTRA 在预训练任务上进行修改, 去除了 [MASK] 标记, 不仅与下游任务更加匹配, 而且使用鉴别器来分辨生成的词汇, 训练效率高, 提升了预训练任务的难度, 能让模型学到更好的文本表示。

(1) ELECTRA 1 层

ELECTRA 1 层主要用于获取原始文本的语义特征。分好词的序列 ω_i 通过预训练模型 ELECTRA 1 输入层时, 生成三个向量, 分别是字向量

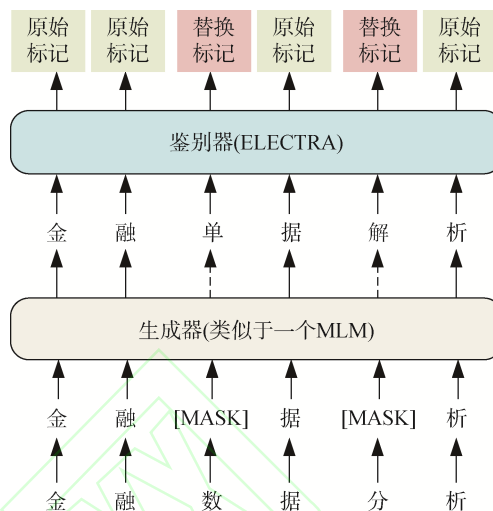


图4 ELECTRA 预训练任务改进

Fig.4 The Improvements of ELECTRA's Pre-training Task

$E_{Token}(\omega_i)$ 、句间分割向量 $E_{Segment}(\omega_i)$ 、位置向量 $E_{Position}(\omega_i)$, 三者相加后的向量, 作为输入向量, 如公式(1)所示。

$$E_{input}(\omega_i) = E_{Token}(\omega_i) + E_{Segment}(\omega_i) + E_{Position}(\omega_i) \quad (1)$$

E_{input} 通过 Transformer^[11] 的 Encoder 部分, 这个部分主要的结构为多头自注意力机制 (Multi-head Self-Attention) 和前馈神经网络。

多头自注意力机制根据输入的向量 E_{input} 生成 Q, K, V 三个向量, 将多个注意力头的信息拼接在一起, 帮助模型提高专注于输入序列不同位置信息的能力, 如公式(2)-公式(4)所示。

$$MultiHead(Q, K, V) =$$

$$Concat(head_1, head_2, \dots, head_k)W^0 \quad (2)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

前馈神经网络将多头注意力层的输出进行映射, 如公式(5)所示。

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (5)$$

序列 ω_i 通过词向量表示层和多层 Transformer 的 Encoder 单元后, 生成的隐藏层向量如公式(6)所示。

$$H_{\omega} = ELECTRA_1(\omega_i) \quad (6)$$

(2) ELECTRA 2 层

ELECTRA 2 层主要根据词性特征获取实体之间的联系。对于给定的词性特征序列 p_k 依据 v_j 分词

后的结果进行 BIO 标注, 然后将其补 0 扩充为与 ω_i 同样的长度, 记为序列 $p_k^w \in \mathbf{R}^m (i = 1, 2, \dots, m)$, 如图 5 所示。

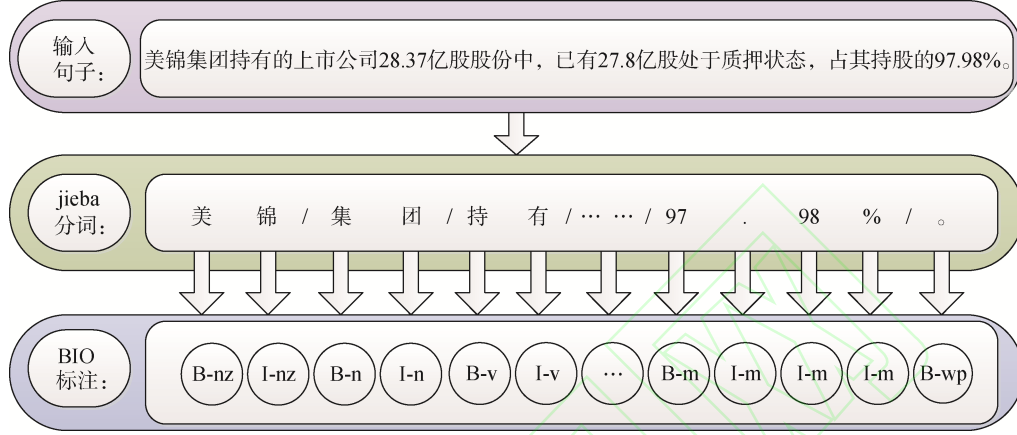


图 5 词性标签预处理示例

Fig.5 Example of Part-of-Speech Tags Preprocessing

综上, p_k 通过预训练模型生成隐藏层向量, 如公式(7)所示。

$$H_p = ELECTRA_2(p_k) \quad (7)$$

(3) ELECTRA 隐藏层融合

通过两个 ELECTRA 预训练模型获得两个隐藏层向量 H_ω 、 H_p , 进行隐藏层融合。将 ELECTRA 1 获得的语义之间的联系与 ELECTRA 2 获得的词性之间的联系进行点乘融合, 增强语义信息, 获得事件要素之间更加清楚的联系, 如公式(8)所示。

$$H_{merge} = H_\omega \otimes H_p \quad (8)$$

其中, \otimes 是点乘操作。

3.5 BiGRU 层

BiGRU 层用于获取输入向量的长距离语义依赖信息。GRU^[24]与循环神经网络相比, 解决了梯度消失、梯度爆炸的问题, 与 LSTM^[25]模型相比, 改进了复杂冗余的模型结构。由于金融事件中出现的主体比较多也比较复杂, 因此通过 GRU 可以进一步提炼通过预训练模型提取之后的特征信息, 获取远距离金融事件要素之间的联系。本文将通过预训练模型的特征向量 H_{merge} 输入 BiGRU 层, 提取金融事件语料关键信息, 如公式(9)和公式(10)所示。

$$\vec{h}_t = \overrightarrow{GRU}(\vec{h}_{t-1}, h_t^{merge}) \quad (9)$$

$$\overleftarrow{h}_t = \overleftarrow{GRU}(\overleftarrow{h}_{t-1}, h_t^{merge}) \quad (10)$$

其中, h_t^{merge} 是当前 GRU 的输入, \vec{h}_t 是前向 GRU 的隐藏层状态向量, \overleftarrow{h}_t 是后向 GRU 的隐藏层状态向量, 得到双向的语义依赖之后进行拼接, 如公式(11)所示。

$$H = [\vec{h}_t, \overleftarrow{h}_t] \quad (11)$$

其中, $[\cdot]$ 是拼接操作。

3.6 CRF 层

CRF 能够对标签之间的依赖关系进行建模, 克服标签偏差问题^[9]。将获取了上下文语义依赖的特征向量 H 传入线性层之后, 得到 $m \times n$ 的矩阵 P , 其中 $P_{i,j}$ 是第 i 个标记中属于第 j 个标签的得分, m 是模型设定的句子最大长度, n 是标签个数。对于一个输入序列 $x = \{x_1, x_2, \dots, x_m\}$ 、输入标签序列 $y = \{y_1, y_2, \dots, y_m\}$ 到输出的标签序列得分如公式(12)所示^[26]。

$$S(x, y) = \sum_{i=0}^m A_{y_i, y_{i+1}} + \sum_{i=1}^m p_{i, y_i} \quad (12)$$

其中, A 是转移概率矩阵, $A_{i,j}$ 是标签 i 到标签 j 的转换分数。训练过程中, 优化 $\{x_i, y_i\}$ 的最大似然函数, 如公式(13)^[27]所示。

$$L = \sum_{i=1}^m \log(P(y_i | x_i)) + \frac{\lambda}{2} \|\theta\|^2 \quad (13)$$

其中, λ 和 θ 为正则参数, $P(y_i|x_i)$ 为原序列到预测序列的概率。依据最大似然函数, 得到最后的序列标签。

3.7 输出层

根据 CRF 层输出的标签 $y = \{y_1, y_2, \dots, y_m\}$, 进行 BIO 转换, 输出事件元素和元素角色, 以上述例句为例进行抽取, 抽取结果如表 1 所示。

表 1 事件抽取输出结果

Table 1 Results of Event Extraction

| 抽取框架 | 抽取结果 |
|------|--|
| 原始内容 | 美锦集团持有的上市公司 28.37 亿股股份中, 已有 27.8 亿股处于质押状态, 占其持股的 97.98%。 |
| 事件类型 | 质押 |
| 事件元素 | “美锦集团”: 质押公司 |
| | “股份”: 质押物 |
| | “27.8 亿”: 质押数量 |
| | “97.98%”: 质押比例 |

4 实验与分析

4.1 实验数据

本文使用“CKKS 2020: 面向金融领域的小样本跨类迁移事件抽取”的训练集数据^①, 该数据集包含 5 个金融事件, 分别是质押、股份股权转让、投资、起诉以及减持, 每个事件的事件框架如表 2 所示。

数据集中共有 2 729 条样本, 由于每条样本中可

表 2 事件框架

Table 2 Event Framework

| 事件类型 | 事件框架 |
|--------|--|
| 质押 | 触发词、质押公司、质押人、质权公司、质权人、质押物、质押日期、质押金额、质押数量、质押比例 |
| 股份股权转让 | 触发词、股份股权转让公司、股份股权转让人、受转让公司、受转让人、股份股权转让物、转让日期、转让交易金额、转让数量、转让比例、标的公司 |
| 投资 | 触发词、原告(个人)、原告(公司)、被告(个人)、被告(公司)、起诉日期 |
| 起诉 | 触发词、发起投资的组织或单位、被投资的组织或单位、投资金额、日期 |
| 减持 | 触发词、减持方、减持方的职务、日期、减持的股份占个人股份百分比、减持的股份占公司股份的百分比、减持方所在组织或单位 |

能包含多个事件, 具体的事件类型分布如表 3 所示, 训练集、验证集和测试集按照 7: 1: 2 的比例进行划分。

表 3 事件个数统计

Table 3 Number of Events Statistics

| 事件类型 | 质押 | 股份股权转让 | 投资 | 起诉 | 减持 | 共计 |
|------|-----|--------|-------|-----|-----|-------|
| 事件数 | 851 | 1 572 | 1 081 | 533 | 739 | 4 776 |

4.2 评价标准与环境配置

采用准确率(Precision, P), 召回率(Recall, R), F 值(F_1)作为衡量该模型事件元素抽取好坏的指标, 三者的计算方法如公式(14)~公式(16)所示。

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F_1 = \frac{2PR}{P + R} \quad (16)$$

其中, TP 为模型正确抽取的事件元素个数, FP 为模型错误抽取的事件元素个数, FN 为模型没有抽取到的正确事件元素个数。实验环境如表 4 所示。

表 4 实验环境

Table 4 Experimental Environment

| 实验环境 | 环境配置 |
|--------|---------------------------------|
| 操作系统 | Ubuntu |
| CPU | Intel Xeon E5-2620 2.10 GHz |
| GPU | TITAN X(12GB) |
| Python | 3.6 |
| 内存 | 128GB |
| 深度学习框架 | TensorFlow(1.14.0)+Keras(2.2.4) |

4.3 实验参数设置

采用 jieba 工具包对语料进行分词, LTP 工具包对语料进行词性标注。深度学习模型参数如表 5 所示。

4.4 对比实验设置

(1) 基准算法对比

为验证 ELECTRA-POS-BiGRU-CRF 模型的有效性, 分别将其与以下模型进行对比:

①IDCNN-CRF 模型: 由膨胀卷积神经网络和条

^①https://www.biendata.xyz/competition/ckks_2020_3/.

表 5 模型参数

Table 5 Parameter Settings

| 参数 | 参数值 |
|------------|----------|
| 学习率 | 0.000 02 |
| 最大序列长度 | 256 |
| Batch Size | 8 |
| GRU 单元数 | 128 |
| Epoch | 30 |
| Dropout | 0.5 |
| Optimizer | Adam |

件随机场组成。膨胀卷积网络增大了传统卷积神经网络的感知视野,与循环神经网络相比,收敛速度更快。

②BiLSTM-CRF 模型^[28]:由双向长短时记忆网络和条件随机场组成。BiLSTM 解决了循环神经网络梯度消失、梯度爆炸问题,可以更好地提取上下文的语义依赖。

③BERT-IDCNN-CRF 模型^[27]:由预训练模型 BERT、膨胀卷积神经网络和条件随机场组成。使用 BERT 获取语料表示,IDCNN 抽取特征,CRF 克服标签偏差问题。

④BERT-BiLSTM-CRF 模型:由预训练模型 BERT、双向长短时记忆网络和条件随机场组成。使用 BERT 获取语料表示,BiLSTM 获取上下文语义依赖,CRF 克服标签偏差。

本文算法与各个基准算法的对比结果如表 6 所示。

表 6 基准算法对比

Table 6 Comparison with Baselines

| 模型 | $P(\%)$ | $R(\%)$ | $F_1(\%)$ |
|-----------------------|--------------|--------------|--------------|
| IDCNN-CRF | 72.10 | 33.14 | 45.41 |
| BiLSTM-CRF | 79.19 | 36.77 | 50.22 |
| BERT-IDCNN-CRF | 86.10 | 56.13 | 67.96 |
| BERT-BiLSTM-CRF | 87.03 | 55.81 | 68.01 |
| ELECTRA-POS-BiGRU-CRF | 87.56 | 59.64 | 70.96 |

对比模型①与模型②、模型③与模型④可以发现,BiLSTM 在金融事件抽取任务上的表现要优于 IDCNN,这说明,尽管 IDCNN 可以增大模型对语料局部的感受能力,对实体边界识别起到重要作用,但

是面对金融事件抽取这类任务,局部感知能力无法捕捉上下文金融实体的联系,无法解决长距离语义信息丢失的问题。

对比模型①与模型③、模型②与模型④可以发现,加入预训练模型 BERT 后, F_1 值有了大幅提升,说明预训练模型在大规模数据上进行预训练,将学习到的知识进行迁移,具有更好的提取语料特征的能力,能够大幅提升抽取金融事件元素的效果。

与这些经典的模型相比,本文模型的准确率、召回率以及 F_1 值都达到了最高,这说明加入词性特征可以有效提取实体之间的联系,避免实体间依赖信息的丢失,增强模型关注重要实体的能力,从而更有效地从语料中抽取金融事件的关键信息,实现更准确的事件抽取。

(2) 消融实验

为确定模型中每个部分提升金融事件要素抽取效果的作用,本文进行了消融实验,结果如表 7 所示。

表 7 消融实验

Table 7 Ablation Experiment

| 模型 | $P(\%)$ | $R(\%)$ | $F_1(\%)$ |
|-----------------------|--------------|--------------|--------------|
| BiGRU-CRF | 75.33 | 35.34 | 48.11 |
| ELECTRA-CRF | 84.04 | 58.65 | 69.09 |
| ELECTRA-BiGRU-CRF | 85.59 | 58.93 | 69.80 |
| ELECTRA-POS-BiGRU-CRF | 87.56 | 59.64 | 70.96 |

BiGRU-CRF 加入预训练模型 ELECTRA 后, F_1 值提升 21.69 个百分点,这说明加入 ELECTRA 后能够更明确地捕捉上下文语义信息,对于金融事件抽取任务更为有利。ELECTRA-CRF 加入 BiGRU 网络后,进一步获取远距离的上下文语义依赖关系,抽取金融事件要素的能力增强, F_1 值提升 0.71 个百分点。ELECTRA-BiGRU-CRF 加入词性特征后,模型能够同时利用语料中的原始信息以及词性特征信息,识别金融事件要素边界的能力提升,捕捉语料中金融事件要素联系的敏感度上升,模型能够更加准确地理解语料中描述的金融事件信息, F_1 值提升 1.16 个百分点。

(3) 预训练模型效果对比

为验证预训练模型 ELECTRA 的有效性,分别在基于 BERT 的改进预训练模型上进行对比实验。

①RoBERTa

RoBERTa (A Robustly Optimized BERT Pretraining Approach) 是 Facebook 在 2019 年提出的一个模型^[15],它在 BERT 的基础上主要进行了以下改进:

1)使用动态遮蔽语言模型代替静态遮蔽语言模型。

2)采用 FULL-SENTENCES 输入。

3)使用更大的批处理大小进行训练。

4)采用更大的字节级 BPE 词汇表进行训练。

5)使用更多的数据进行训练。

②ALBERT

ALBERT (A Lite BERT) 是谷歌在 2019 年提出的预训练模型^[29]。由于预训练模型存在参数规模较大、训练时间较长的问题,因此,ALBERT 针对上述问题,在 BERT 的基础上进行以下改进:

1)进行嵌入层参数分解化 (Factorized Embedding Parameterization)。

2)使用跨层参数共享 (Cross-layer Parameter Sharing)。

3)使用句间连贯性损失 (Inter-sentence Coherence Loss)。

③NEZHA

NEZHA (Neural Contextualized Representation for Chinese Language Understanding) 是华为诺亚方舟实验室 2019 年提出的预训练模型^[30]。NEZHA 在 BERT 上进行了以下改进:

1)采用函数相对位置编码 (Functional Relative Positional Encoding)。

2)采用全词覆盖 (Whole Word Masking, WWM) 策略。

3)采用混合精度训练 (Mixed Precision Training)。

4)采用 LAMB 优化器^[31]。

本文使用 5 种预训练模型,分别是 BERT-base^[32]、NEZHA-base、ELECTRA-base、ALBERT-large^①和 RoBERTa-base,各个预训练模型的参数如表 8 所示。

实验结果如表 9 所示。可以看出,ALBERT 在模型大小只有 BERT 的 1/5 的情况下,准确率达到最

表 8 预训练模型参数

Table 8 Parameters of Pre-trained Models

| 预训练模型 | 层数 | 隐藏层单元 | 注意力头 | 模型大小 |
|--------------|----|-------|------|----------|
| BERT-base | 12 | 768 | 12 | 392 MB |
| NEZHA-base | 12 | 768 | 12 | 1 173 MB |
| ELECTRA-base | 12 | 768 | 12 | 102 MB |
| ALBERT-large | 24 | 1 024 | 16 | 64 MB |
| RoBERTa-base | 12 | 768 | 12 | 392 MB |

高,不过这也牺牲了召回率和 F_1 值,导致召回率和 F_1 值是 5 个预训练模型中最低的,这说明 ALBERT 在 BERT 上面进行低秩分解、共享参数以及 NSP 改进,从某种程度上是有效的,但是在金融事件抽取的整体效果上要逊色一些。RoBERTa 的召回率是 5 个模型中最高的,不过准确率是 5 个模型中最低的。与 BERT 相比, F_1 值得到提升,这说明通过大规模数据、大批量训练等改进措施,下游任务的效果会有一定的提升。与 ELECTRA 相比,模型大小是 ELECTRA 的近 4 倍, F_1 值也低于 ELECTRA,这说明 ELECTRA 的预训练任务改进方式在金融事件抽取上更为有效。ELECTRA 的 F_1 值最高,并且准确率、召回率与其他模型相比具有可竞争性。与 BERT 相比, F_1 值提升 1.31 个百分点,说明通过改进 MLM 任务,使其与下游任务更加匹配,下游任务的效果会大幅度提高。

表 9 与其他预训练模型对比情况

Table 9 Comparison with Other Pre-trained Models

| 预训练模型 | $P(\%)$ | $R(\%)$ | $F_1(\%)$ |
|-----------------------|--------------|--------------|--------------|
| BERT-POS-BiGRU-CRF | 85.19 | 58.91 | 69.65 |
| NEZHA-POS-BiGRU-CRF | 86.44 | 58.99 | 70.13 |
| RoBERTa-POS-BiGRU-CRF | 84.62 | 60.11 | 70.29 |
| ALBERT-POS-BiGRU-CRF | 87.63 | 55.67 | 68.08 |
| ELECTRA-POS-BiGRU-CRF | 87.56 | 59.64 | 70.96 |

(4) 金融事件元素抽取结果对比

为更加详细地说明 ELECTRA-POS-BiGRU-CRF 神经网络模型的效果,展示不同算法的金融事件元素抽取结果示例如表 10 所示。相较于其他算法,ELECTRA-POS-BiGRU-CRF 模型更准确地抽取该金融事件的相关事件元素。

① https://github.com/brightmart/albert_zh.

表 10 金融事件元素抽取结果对比示例

Table 10 Examples of Financial Event Arguments Extraction Results

| 输入句子 | 事件元素 | 抽取方法 | 元素抽取结果 |
|---|------|-----------------------|--|
| 公告显示,中南建设本次质押股数 2 400 万股,占其所持股份比例为 1.19%,占公司总股本比例 0.64%,质押日期自 2019 年 11 月 15 日至 2021 年 4 月 18 日,质权人为华夏银行股份有限公司南通分行。 | | IDCNN-CRF | '股份比例为 1.19%,占公司': ('股份股权转让', 'obj-org'), '64%': ('投资', 'trigger') |
| | | BiLSTM-CRF | '股份比例为 1.19%,占公司': ('投资', 'sub'), '64%': ('股份股权转让', 'trigger'), ',质押日': ('股份股权转让', 'target-company'), '期自': ('股份股权转让', 'collateral')} |
| | | BERT-IDCNN-CRF | '中南建设': ('质押', 'sub-org'), '质押': ('质押', 'trigger'), '2400 万': ('质押', 'number'), '1.19%': ('质押', 'proportion'), '0.64%': ('质押', 'proportion'), '华夏银行股份有限公司南通分行': ('质押', 'obj-org') |
| | | BERT-BiLSTM-CRF | '中南建设': ('质押', 'sub-org'), '质押': ('质押', 'trigger'), '股': ('质押', 'collateral'), '2400 万': ('质押', 'number'), '股份': ('质押', 'collateral'), '1.19%': ('质押', 'proportion'), '2019 年 11 月 15 日至 2021 年 4 月 18 日': ('质押', 'date'), '华夏银行股份有限公司南通分行': ('质押', 'obj-org') |
| | | ELECTRA-POS-BiGRU-CRF | '中南建设': ('质押', 'sub-org'), '质押': ('质押', 'trigger'), '2400 万': ('质押', 'number'), '股份': ('质押', 'collateral'), '1.19%': ('质押', 'proportion'), '华夏银行股份有限公司南通分行': ('质押', 'obj-org') |

5 结 语

本文提出一种基于 ELECTRA 模型和词性向量的 ELECTRA-POS-BiGRU-CRF 神经网络模型。该模型使用两个 ELECTRA 预训练模型,分别获取文本的语义信息以及词性特征之间的关系,通过点乘融合进行语义增强;通过 BiGRU 获取长距离的语义依赖;利用 CRF 进行克服标签偏差问题,充分发挥词性特征在金融事件抽取中的关键作用,从而获取更加准确的金融事件抽取结果。三组对比实验表明,本文模型具有较好的金融事件抽取性能,验证了将词性特征加入模型的有效性。但是,本文只将词性特征加入模型中,未来可以利用句法依存树,进一步研究如何提升模型抽取金融事件关键实体的性能。此外,由于预训练模型较大,对设备要求较高,如何使用模型压缩技术减小模型的大小并且不损害

模型的性能也是未来的研究方向。

参考文献:

- [1] Ding X, Liao K, Liu T, et al. Event Representation Learning Enhanced with External Commonsense Knowledge[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 4894-4903.
- [2] Liang X, Cheng D, Yang F, et al. F-HMTC: Detecting Financial Events for Investment Decisions Based on Neural Hierarchical Multi-Label Text Classification[C]// Proceedings of the 29th International Joint Conference on Artificial Intelligence. 2020: 4490-4496.
- [3] Wollmer M, Weninger F, Knaup T, et al. YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context[J]. IEEE Intelligent Systems, 2013, 28(3): 46-53.
- [4] Feldman R, Rosenfeld B, Bar-Haim R, et al. The Stock Sonar -

- Sentiment Analysis of Stocks Based on a Hybrid Approach[C]//Proceedings of the 23rd Conference on Innovative Applications of Artificial Intelligence. 2011: 1642-1647.
- [5] Arendarenko E, Kakkonen T. Ontology-Based Information and Event Extraction for Business Intelligence[C]//Proceedings of the 15th International Conference on Artificial Intelligence: Methodology, Systems, and Applications. Springer Berlin Heidelberg, 2012:89-102.
- [6] Hogenboom A, Hogenboom F, Frasinca F, et al. Semantics-based Information Extraction for Detecting Economic Events[J]. Multimedia Tools and Applications, 2012, 64(1): 27-52.
- [7] Ronnqvist S, Sarlin P. Bank Distress in the News: Describing Events Through Deep Learning[J]. Neurocomputing, 2016, 264(11): 57-70.
- [8] Jacobs G, Lefever E, Hoste V. Economic Event Detection in Company-Specific News Text[C]//Proceedings of the 1st Workshop on Economics and Natural Language Processing. 2018: 1-10.
- [9] Yang H, Chen Y, Liu K, et al. DCFEE: A Document-Level Chinese Financial Event Extraction System Based on Automatically Labeled Training Data[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 50-55.
- [10] Zheng S, Cao W, Xu W, et al. Doc2EDAG: An End-to-End Document-level Framework for Chinese Financial Event Extraction[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 337-346.
- [11] Vaswani A, Shazeer N, Parmar N, et al. Attention is All You Need [C]//Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017: 5998-6008.
- [12] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[OL]. arXiv Preprint, arXiv:1301.3781.
- [13] Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014:1532-1543.
- [14] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[OL]. arXiv Preprint, arXiv:1810.04805.
- [15] Liu Y, Ott M, Goyal N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach[OL]. arXiv Preprint, arXiv: 1907.11692.
- [16] Yang S, Feng D, Qiao L, et al. Exploring Pre-trained Language Models for Event Extraction and Generation[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 5284-5294.
- [17] Du X, Cardie C. Event Extraction by Answering (Almost) Natural Questions[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP). 2020: 671-683.
- [18] Liu J, Chen Y, Liu K, et al. Event Extraction as Machine Reading Comprehension[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 1641-1651.
- [19] Ein-Dor L, Gera A, Toledo-Ronen O, et al. Financial Event Extraction Using Wikipedia-Based Weak Supervision[C]//Proceedings of the 2nd Workshop on Economics and Natural Language Processing. 2019: 10-15.
- [20] Li Q, Zhang Q. A Unified Model for Financial Event Classification, Detection and Summarization[C]//Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI-20). 2020:4668-4678.)
- [21] Zhao L, Li L, Zheng X. A BERT Based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts[OL]. arXiv Preprint, arXiv:2001.05326.
- [22] 赵军. 知识图谱[M]. 北京: 高等教育出版社, 2018. (Zhao Jun. Knowledge Graph[M]. Beijing: Higher Education Press, 2018.)
- [23] Clark K, Luong M T, Le Q V, et al. Electra: Pre-training Text Encoders as Discriminators Rather Than Generators[OL]. arXiv Preprint, arXiv: 2003. 10555.
- [24] Cho K, van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation[C]//Proceedings of the 19th Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1724-1734.
- [25] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [26] Du X, Cardie C. Document-Level Event Role Filler Extraction Using Multi-Granularity Contextualized Encoding[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics(ACL). 2020: 8010-8020.
- [27] 李妮, 关焕梅, 杨飘, 等. 基于BERT-IDCNN-CRF的中文命名实体识别方法[J]. 山东大学学报(理学版), 2020, 55(1): 102-109. (Li Ni, Guan Huanmei, Yang Piao, et al. BERT-IDCNN-CRF for Named Entity Recognition in Chinese[J]. Journal of Shandong University (Natural Science), 2020, 55(1): 102-109.)
- [28] Lample G, Ballesteros M, Subramanian S, et al. Neural Architectures for Named Entity Recognition[C]//Proceedings of the 2016 Conference on North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 260-270.)
- [29] Lan Z, Chen M, Goodman S, et al. ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations[OL].

arXiv Preprint, arXiv: 1909.11942.

- [30] Wei J, Ren X, Li X, et al. NEZHA: Neural Contextualized Representation for Chinese Language Understanding[OL]. arXiv Preprint, arXiv: 1909.00204.
- [31] You Y, Li J, Hseu J, et al. Reducing BERT Pre-Training Time from 3 Days to 76 Minutes[OL]. arXiv Preprint, arXiv: 1904.00962.
- [32] Cui Y, Che W, Liu T, et al. Revisiting Pre-Trained Models for Chinese Natural Language Processing[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings (EMNLP). 2020: 657-668.

作者贡献声明:

陈星月: 实现研究方案、数据分析、编写代码, 论文撰写、修改和修订;
倪丽萍: 提出研究思路, 论文最终版本修订;

倪志伟: 对研究方案进行修订, 修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: asterexy@163.com。

- [1] 陈星月. train_base.json. “CCKS 2020: 面向金融领域的小样本跨类迁移事件抽取”的训练数据集。
- [2] 陈星月. ccks_codes.zip. 各个模型的源码。
- [3] 陈星月. Model.zip. 实验保存的模型。
- [4] 陈星月. Results.xlsx. 模型实验记录结果。

收稿日期: 2020-12-26

收修改稿日期: 2021-04-09

Extracting Financial Events with ELECTRA and Part-of-Speech

Chen Xingyue Ni Liping Ni Zhiwei

(School of Management, Hefei University of Technology, Hefei 230009, China)

(Key Laboratory of Process Optimization & Intelligent Decision-making, Ministry of Education,
Hefei University of Technology, Hefei 230009, China)

Abstract: [Objective] This paper proposes a method to extract financial events based on the ELECTRA model and part-of-speech, aiming to address the issues of blurred entity boundaries and inaccurate extractions. [Methods] First, we input corpus to two models pre-trained by ELECTRA, which identified key entities, the original semantic information, and part-of-speech. Then, we used the BiGRU model to extract contextual semantic dependency and generated the original sequence tags. Finally, we addressed the issues of label deviation with the CRF model and extracted the financial events. [Results] We examined the new model with financial event dataset and found its F-value reached 70.96%, which was 20.74 percentage point higher than the BiLSTM-CRF model. [Limitations] The number of events in the dataset needs to be increased. The size of pre-trained model is large, which might be limited by the memory of GPU/TPU. [Conclusions] The model based on ELECTRA and part-of-speech could effectively identify the relationships among financial events to extract them.

Keywords: ELECTRA Part-of-Speech Financial Event Extraction Pre-trained Model