Computer Science and Application 计算机科学与应用, 2021, 11(5), 1572-1578
Published Online May 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.115162

Hans汉斯

基于BERT-DGCNN的中文事件抽取方法研究

陈安南，叶岩宁，王畅畅，王文举，李博文

合肥工业大学，安徽 宣城
Email: i-can@qq.com

收稿日期：2021年4月26日；录用日期：2021年5月21日；发布日期：2021年5月28日

摘　要

本文构建了一个事件抽取pipeline模型，其旨在对新闻中的信息元进行有效的抽取。在管道抽取模式下，先对文本进行存在事件类型识别，而后再将事件类型与文本一并作为输入传入模型进行事件论元角色抽取，其中事件论元角色采用类似于BERT中SQuAD等阅读理解任务上的双指针输出。两个基本模型都是利用BERT预训练模型产生的词嵌入，使用DGCNN进行编码之后池化，再连接到dense层进行分类。实验结果表明，本模型可对新闻类内容进行高效抽取。

关键词

事件抽取，BERT模型，膨胀门卷积神经网络

Research on Chinese Event Extraction Method Based on BERT-DGCNN

Annan Chen, Yanning Ye, Changchang Wang, Wenju Wang, Bowen Li

Hefei University of Technology, Xuancheng Anhui
Email: i-can@qq.com

Received: Apr. 26th, 2021; accepted: May 21st, 2021; published: May 28th, 2021

Abstract

This paper constructs an event extraction model, which aims at effectively extracting information elements from the news. This model is based on the BERT pre-training model, which first identifies the existing event type of the text, and then inputs the event type and text into the model to extract the event role, in which the event role adopts double-pointer output similar to the reading comprehension task of a SQuAD in BERT. Experimental results show that this model can extract news content efficiently.

文章引用: 陈安南, 叶岩宁, 王畅畅, 王文举, 李博文. 基于 BERT-DGCNN 的中文事件抽取方法研究[J]. 计算机科学与应用, 2021, 11(5): 1572-1578. DOI: 10.12677/csa.2021.115162

Keywords

Event Extraction, BERT Model, Dilated Gated CNN

1. 引言

随着当今社会互联网和自媒体的普及，网络技术的日益革新，新闻文本每天都会大量产生，互联网作为新闻传播的新媒介，存储了海量的信息。但是由于这些文本信息大都是以非结构化的方式存储在互联网中，使得其很难被处理，面对这巨大的信息财富，一个快速从中获取有价值信息的方法显得越来越重要。

新闻一般以事件的形式呈现出来，对新闻中蕴含的事件进行抽取，可达到快速获取主要信息的目的。一般的顺序不敏感模型没有明确的位置标记，可能会导致在处理语法敏感的任务时由于语序或者语法结构的影响不能完全捕获自然语言的语义。因此我们采用上下文敏感的模型进行事件抽取任务。接着很多研究中使用端到端的神经网络进行训练，其性能通常较好，但由于事件抽取任务的复杂性，其通常伴随大量参数从而导致巨大的计算力使用。至此，本文提出一种简洁事件抽取模型架，基于上下文敏感模型产生的预训练词向量且具有更少的参数，输出上采用类似于在 SQuAD 等阅读理解任务上的输出对中文事件进行抽取。

2. 相关研究

一般认为，事件抽取要区分为元事件抽取与主题事件抽取两种任务。元事件表示一个动作的发生或状态的变化[1]，包括参与该事件的主客体，通常由动词或动名词作为触发词。主题事件包括一类核心事件或活动以及所有与之直接相关的事件和活动[2]，可以由多个元事件构成。本文探讨研究的范畴只限定于元事件抽取，下文统称为事件抽取。图 1 描述了一个事件的构成：

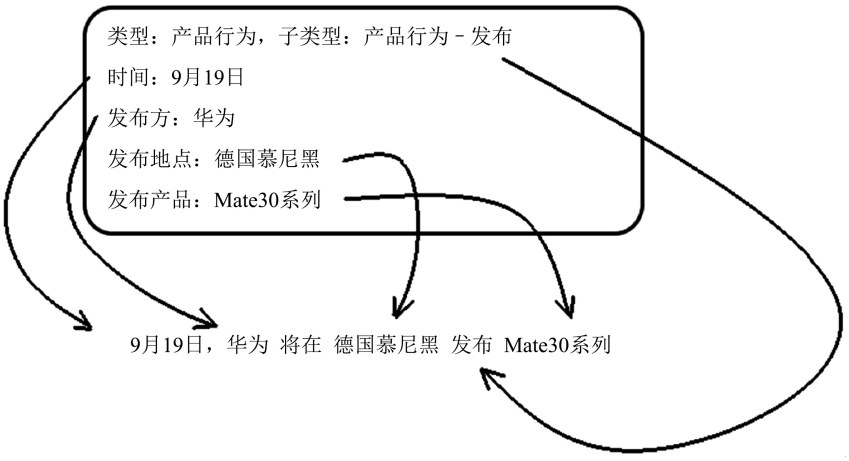

Figure 1. The basic constitutive elements of a “release” event
图 1. “发布”事件的基本构成要素

事件由事件触发词(Trigger)和描述事件结构的元素(Argument)构成。事件抽取作为知识抽取领域的一个子任务，也是其一个非常重要的研究方向，一直受到许多研究者们的注意。对事件抽取的研究主要分为模式匹配和机器学习两大类方法。

2.1. 基于模式匹配的事件抽取

基于模式匹配的方法是在一些人工定义的模式下对事件进行识别和抽取，模式主要用于指明构成目标信息的上下文约束环境[2]，主体思想是融合语言知识与领域知识。ExDisco [3]，GenPAM [4]是两个典型的基于模式匹配的事件抽取系统。基于模式匹配方法适合应用于特定领域，能达到较高的性能需求，但移植性较差，且需要在领域专家的指导下构建。

2.2. 基于机器学习的事件抽取

事件抽取任务在机器学习方法一般中一般被视为分类问题。基于机器学习的方法又可划分为传统机器学习和深度学习。

传统机器学习的研究主要在于对候选触发词和论元的语法、句法和语意等等特征进行研究以进行捕捉，之后基于统计分类模型进行分类。文献[5]用 MegaM 二元分类器和 TiMBL 多元分类器进行事件抽取，在 ACE 英文数据集上取得较好效果。文献[6]结合触发词和二元分类相进行事件抽取，在 ACE 中文数据集上取得较好效果。

随着大数据时代到来，以及计算机算力的快速增长，近年又随之兴起深度学习方法，其能学习到不同于传统机器学习离散型特征的连续型向量特征。文献[7]率先将深度学习方法应用于事件抽取，采用预训练的词向量的同时融入对于单词的语义语法的建模，接着利用位置信息来描述各个词与候选触发词的距离，取得了很好的效果。文献[8]提出一种联合学习事件识别和论元角色分类的基于 RNN 的模型，在论元角色分类任务上达到 SOTA 效果。

2.3. 词向量表示

词向量表示对事件抽取任务有着重要意义，同时其也是 NLP 领域非常热门的一个研究方向。其是将不可计算、非结构化的词转化为可计算、结构化的向量。较早的词向量训练工具 Word2Vec9 [9]将每个词映射到唯一的向量，从而表示词与词之间的关系，但由于其词和向量的一对一关系，所以无法解决一词多义的问题。文献[10]提出的 ELMo 模型，基于双层 Bi-LSTM，双向拼接正向编码器和反向编码器提取特征信息，使上下文无关的静态向量变为上下文相关的动态向量。文献[11]提出了 GPT 模型，与 ELMo 模型的不同点是 GPT 采用 Transformer 替代 LSTM 作为模型进行特征提取，但 GPT 只使用了单向编码。考虑到 ELMo 和 GPT 都不能同时利用两个方向的信息，谷歌提出了使用双向 transformer 的 BERT [12]模型。

3. 基于 BERT-DGCNN 的新闻类内容事件抽取模型

本模型分为两部分：第一部分为事件类型预测模型，第二部分为对事件角色抽取模型，其中第一部分的抽取结果同时与文本作为第二部分的输入。

3.1. 事件类型抽取模型

模型共分为三部分：句子级编码层、膨胀门卷积网络层与平均池化层、分类层。图 2 描述了事件类型抽取模型的整体框架：

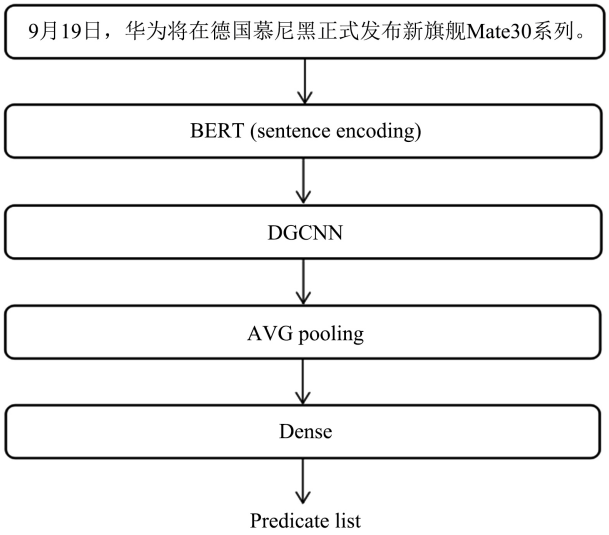

Figure 2. The framework of the event type predication model
图 2. 事件类型预测模型框架

3.1.1. 句子级编码层

本层使用 BERT (Bidirectional Encoder Representation from Transformers) [12]得到句子级编码特性。它是由谷歌在 2018 年提出的一种预训练模型，基于双向 transformer，创新地引入了 MASK 语言模型与 Next Sentence Prediction 机制，前者用来解决 transformer 只能利用单向信息的问题，后者能够让模型学习到句子间的关系。

BERT 的输入由三部分组成：Token Embeddings、Segment Embeddings、Position Embeddings。其中 Token Embeddings 是词向量，在中文 NLP 任务中，只需要在句子本文开头添加一个特殊的 token：[CLS]，文本末尾添加 token [SEP]，以及未登录词典字换为[UNK]即可；Segment Embeddings 用于区分输入中的第一个句子和第二个句子，本任务输入只有一个句子，全部设置为 0 即可；Position Embeddings 用于对文本的字符进行位置编码，因为一般认为 transformer 没有对输入进行位置编码。

本文实验中使用的是 Google 发布的 BERT-Base 中文版，该模型使用 12 层的 Transformer，隐藏层维度大小为 768，自注意力 multi-head = 12，模型所有参数为 12 × 768 × 12 = 110 M，使用 GPU 内存 7G 多。

3.1.2. 膨胀门卷积神经网络(Dilated Grated CNN)与池化

膨胀门卷积神经网络，将门卷积与膨胀卷积融合使用。令门卷积假设处理序列为：$X=[x_1,x_2,\cdots,x_n]$，我们对普通的一维卷积添加一个门：

$$Y=\text{Conv1D}_1(X)\otimes\sigma(\text{Conv1D}_2(X))$$

使用两个形式相同但权值互不共享的 Conv1D 网络，其中后者使用 sigmoid 函数激活，而前者没有使用，之后再逐位相乘。sigmoid 激活后的 Conv1D 函数值域为(0, 1)，类似于将其每个输出施加一个阀门。

接着使用残差结构：输出以$(1-\xi)$的概率直接通过，ξ的概率经过变换后通过，如公式表达：

$$Y=X\otimes(1-\xi)+\text{Conv1D}_1(X)\otimes\xi$$

$$\xi=\sigma(\text{Conv1D}_2(X))$$

其中 σ 即为 sigmoid 操作，使用残差结构的意义一方面在于使得梯度消失的可能性更低，另一方面可以在多通道传输信息。

之后我们使用膨胀卷积代替普通卷积，普通卷积捕捉与中心相邻的输入，而膨胀卷积捕捉到的是与中心有间隔的输入，像是一个窗口内被挖空若干单元的卷积，所以也叫空洞卷积，这样可以在不给模型添加参数的情况下，使得 CNN 能够捕捉更远距离的特征。

接着对 DGCNN 的输出进行池化，池化的作用是使得要处理的特征个数和参数减少。常用的池化有最大池化、平均池化和随机池化。最大池化一般能获取最显著的特征，平均池化可以保留数据整体的特征，随机池化中元素值大小和元素被选中的概率成正相关。本模型使用的是平均池化。

3.1.3. 分类层

一段文本可能含有多个事件类型，所以本文将其视作一个多标签任务，分类层使用一个全连接层后，以池化结果作为输入，使用 sigmoid 函数进行输出。

3.1.4. 损失函数的设计

下面给出损失函数的设计：

$$L_R^* = -\frac{\sum_{i=1}^{N} \delta_i y_{ti} \log y_i}{\sum_{i=1}^{N} y_{ti}}$$

其中 y_{ti} 为标签值，y_i 为预测值，$y_{ti} \log y_i$ 为原始交叉熵损失函数。δ_i 作为边界指示函数，只惩罚模型还未充分学到的样本，忽略已可分样本用于解决样本噪声与类别不平衡问题，δ_i 其定义如下：

$$\delta_i = \begin{cases} 1, & (0.5 - m < y_i < 0.5 + m) \text{ or } \theta_i < 0 \\ 0, & \text{otherwise} \end{cases}$$

其中又有 $\theta_i = \text{sgn}\left((y_i - 0.5)(y_{ti} - 0.5)\right)$，用来表示已可分样本，$m$ 用来设定样本被正确识别的阈值。

3.2. 事件角色抽取模型

这部分的模型结构与事件类型预测模型除输出部分外基本一致。

输入部分：将前一模型预测结果与事件角色拼接作为 BERT 输入 sequence 1，文本作为输入 sequence 2，如图 3 所示：

[CLS]产品行为 - 发布，时间[SEP] 9月19日，华为将在德国慕尼黑正式发布新旗舰Mate30系列。[SEP]

Figure 3. Event extraction model input sample
图 3. 事件抽取模型输入示例

输出部分：使用类似于在 SQuAD 等阅读理解任务上的输出，每个 token 有两个作为事件论元角色首末地址的概率输出，这些概率不互斥所以使用 Sigmoid 输出，并且便于加入先验分数、选取多个候选项和进行阈值筛选等。

4. 实验设计和实验结果分析

4.1. 数据集

本文采用的是目前中文事件抽取 DuEE 数据集，其定义了 8 个主事件类别，每个类别包含若干个子事件类型，共 65 个事件类型和 121 个事件角色类别。数据集总共包括 1.7 万个中文句子，其中 1.3 万作为训练集，0.15 万作为验证集，0.35 万作为测试集。

4.2. 实验环境与超参数设置

本文实验采用谷歌 Cloab 平台进行，它是一种托管式 Jupyter 笔记本服务，免费提供 GPU 甚至 TPU 算力，十分适合机器学习任务。本任务使用 Linux 18.04.5 LTS 操作系统，GPU 算力是 Nvidia K80，python 版本为 3.6，BERT 版本为中文版 BERT-BASE。

BERT 的训练超参数如表 1 所示：

Table 1. The hyperparameters of BERT model
表 1. BERT 模型超参数设置

超参数名	超参数值
序列长度(max_length)	256
每批大小(batch_size)	32
学习率(learning_rate)	2e−5
dropout	0.1
迭代轮次	9

4.3. 评价指标及评价方法

此次实验按字级别匹配进行打分，选用精确率(P)、召回率(R)以及 F1 值作为评价指标，计算公式为：

$$P = \frac{\text{预测论元得分总和}}{\text{所有预有预测论元的数量}}$$

$$R = \frac{\text{预测论元得分总和}}{\text{所有人工标有人工标注量}}$$

$$F1 = \frac{2 * P * R}{(P + R)}$$

其中，预测论元得分 = 论元角色是否准确 * 事件类型是否准确 * 字级别匹配F1值；

字级别匹配F1值 = 2 * 字级别匹配P值 * 字级别匹配R值/(字级别匹配P值 + 字级别匹配R值)；

字级别匹配P值 = 预测论元和人工标注论元共有字的数量/预测论元字数；

字级别匹配R值 = 预测论元和人工标注论元共有字的数量/人工标注论元字数 。

4.4. 实验结果

将本文提出的 BERT-DGCNN 模型与 ERNIE-CRF、BERT-base、BERT-CRF 三种模型进行比较，实验结果如下表所示：

Table 2. The experiment results of each event extraction model
表 2. 各事件抽取模型实验结果

模型	精确率	召回率	F1 值
ERNIE-CRF	80.4%	68.6%	74.0%
BERT-Base	74.5%	79.7%	77.0%
BERT-CRF	80.1%	75.8%	77.9%
BERT-DGCNN	81.3%	77.4%	80.0%

从表 2 可以看到，本文的模型在百度 DuEE 数据集上取得了相对不错的结果，ERNIE-CRF 和 BERT-CRF 差距比较大的原因可能是初始预训练语料存在一定差异同时对模型的迭代方式存在一定差别。BERT-CRF 相对于 BERT-Base 在 F1 值上提升的效果并不明显可能是因为加上 CRF 使得模型对训练集的拟合效果较好，但相对的泛化能力有所减弱。BERT-DGCNN 模型相较于 BERT-CRF 在准确率、回率和 F1 值三个指标上都有所提升，说明了本模型在中文事件抽取任务上效果有所提升。

5. 结束语

本文提出一种基于 BERT 预训练模型的中文事件抽取 pipeline 模型，利用 BERT 模型产生词嵌入，DGCNN 对上层产生的向量化序列进行编码，之后将 DGCNN 池化结果送入一个全连接层进行分类。实验结果显示，本文提出的模型在 DuEE 数据集上实验效果较好。在接下来工作中，计划针对特定领域，对模型进行优化，以技术支持特定领域知识图谱的构建等工作。

基金项目

合肥工业大学 2019 年大学生创新创业训练计划项目国家级项目：基于知识图谱的新闻类内容校验系统(项目编号：201910359097)。

参考文献

[1] 吴刚. 基于主题的中文事件抽取技术研究及应用[D[: [硕士学位论文]. 苏州: 苏州大学, 2009.

[2] 郑家恒, 王兴义, 李飞. 信息抽取模式自动生成方法的研究[J]. 中文信息学报, 2004, 18(1): 48-54.

[3] Yangarber, R. (2001) Scenario Customization for Information Extraction. Defense Advanced Research Projects Agency, Arlington, VA.

[4] 姜吉发. 自由文本的信息抽取模式获取的研究[D]: [博士学位论文]. 北京: 中国科学院计算技术研究所, 2004.

[5] Ahn, D. (2006) The Stages of Event Extraction. *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, Association for Computational Linguistics, 1-8. https://doi.org/10.3115/1629235.1629236

[6] 赵妍妍. 中文事件抽取的相关技术研究[D[: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2007.

[7] Chen, Y., Xu, L., Liu, K., *et al.* (2015) Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks. *Proceedings of the* 53*rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (*Volume* 1: *Long Papers*), Association for Computational Linguistics, 167-176. https://doi.org/10.3115/v1/P15-1017

[8] Nguyen, T.H., Cho, K. and Grishman, R. (2016) Joint Event Extraction via Recurrent Neural Networks. *Proceedings of the* 2016 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, Association for Computational Linguistic, 300-309. https://doi.org/10.18653/v1/N16-1034

[9] Mikolov, T., Chen, K., Corrado, G., *et al.* (2013) Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781 [cs.CL]

[10] Peters, M., Neumann, M., Iyyer, M., *et al.* (2018) Deep Contextualized Word Representations. *Proceedings of the* 2018 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, *Volume* 1 (*Long Papers*), Association for Computational Linguistics, 2227-2237. https://doi.org/10.18653/v1/N18-1202

[11] Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I. (n.d.) Improving Language Understanding by Generative Pre-Training. https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

[12] Devlin, J., Chang, M.W., Lee, K., *et al.* (2018) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL]