

文章编号: 1003-0077(2018)11-0034-07

基于 Tree-based CNN 的关系抽取

刘 伟,陈鸿昶,黄瑞阳

(国家数字交换系统工程研究中心,河南 郑州 450000)

摘 要: 该文基于语法树和卷积神经网络(Convolutional Neural Network, CNN)构建了一种树形结构的卷积神经网络 Tree-based CNN,以解决在关系抽取的任务中传统 CNN 对句法信息编码不足的问题。该文实验中,Tree-based CNN 相对于 CNN 和 LSTM,在关系抽取方面准确率分别提高了 3% 和 5%。实验证明,Tree-based CNN 在关系抽取任务中,可以同时具有 CNN 和 LSTM 的优势。

关键词: 关系抽取;语法分析;CNN

中图分类号: TP391

文献标识码: A

Neural Relation Extraction Via Tree-based CNN

LIU Wei, CHEN Hongchang, HUANG Ruiyang

(National Digital Switching System Engineering and Technology Research
Center, Zhengzhou, Henan 450000, China)

Abstract: Tree-Based CNN, a kind of tree-structured neural network based on syntax tree, is implemented to extract entities relations in natural language which is designed to enhance tree coding in original CNN model. Experiment in this paper indicates Tree-based has an improvement of 3% and 5% compared to CNN and LSTM, respectively, in the task of relation extraction.

Keywords: relation extraction; natural language processing; CNN

0 引言

随着深度学习在人工智能领域取得突破性的进展,科技工作者们越来越相信以数据为驱动的方法能够在智能化研究的道路上大展鸿图。文本作为信息量最密集的数据形态之一,其优秀的处理方法也越来越受到重视。深度学习的方法自被引入到自然语言处理任务中来,在本领域的多项任务中均取得了相当可观的效果。特别是做实体关系抽取任务时,在引入了 CNN 和 LSTM 后,关系抽取效果有了明显提升。采用深度学习的方法相比传统方法有明显的不同:传统方法更加注重分析语句和句法的结构,从而对关系进行判断分类,而基于深度学习的方法则更加注重词义及其表示。本文尝试通过在神经网络架构中引入句法递归结构,使得句法结构信息更有效地参与到

关系抽取任务中,从而提高抽取效果。

实体(Entity),即真实世界存在的特定的事实信息。关系抽取(Relation Extraction, RE),是通过计算机手段自动寻找文本中所描述的特定的实体关系。例如,“吴恩达成为百度首席科学家”中就包含一个雇佣关系,可以用三元组(吴恩达,雇员,百度)表示,也可以根据需求将关系更加具体化,用三元组(吴恩达,首席科学家,百度)表示。

当前实体关系抽取的主流系统一般采用有监督的学习方法。这类方法往往需要大量的人工标注数据,而人工标注数据的获得往往需要消耗大量的人力物力资源。2007 年 Auer 等建立了 DBpedia^[1], Suchanek 等建立了 YAGO^[2], 2008 年 Bollacker 等建立了 Freebase^[3]等语义数据集,被广泛应用在自然语言处理的各项任务中。尽管这些数据集中包含了大量的关系实例,但对于同一关系实例的数据往

收稿日期: 2017-11-22 定稿日期: 2018-01-16

基金项目: 国家自然科学基金(61521003)

往不足以训练成熟的关系模型。2009 年, Mintz 等提出了远程监督的关系抽取算法^[4], 能够自动关联远程数据库中相应的训练数据。该算法假设如果两个实体 E1、E2 存在特定关系 R, 那么关系 R 将在所有包含实体 E1、E2 的文本材料中体现。这个假设在很多情况下是成立的, 但并不绝对。例如, “吴恩达表示, 现在全世界范围内, 百度在使用超级计算机发展人工智能的公司中是最优秀的”。这句话中虽然也包含“吴恩达”、“百度”两个实体, 但并不能体现“雇员”这层关系, 不过该模型仍然会将其标注为该关系的支持语料。针对上述错误标注问题, 2010 年 Riedel 等^[5]、2011 年 Hoffman 等^[6]、2012 年 Surdeanu 等^[7]基于自然语言处理的相关工具提出了多示例学习的方法及改进方法。

近年来, 深度学习方法被应用到关系抽取的工作中。2012 年 Socher 等在句子级的人工标注数据上搭建了深度神经网络分类器^[8], 但由于数据不易获得因而难以应用在大规模数据集上。因此, 2015 年 Zeng 等^[9]将远程监督数据的思想应用到神经网络的训练中, 取得了突破性进展。这种方法假设在包含两个实体的所有语句中, 至少有一个语句可以表达它们之间的关系。然后, 选择最能表达该关系的语句进行训练和预测。这个过程可以有效地排除那些对关系表达毫无作用的语句, 但同时也放弃了可能包含重要信息的语句。为了能够抗干扰, 同时更充分地利用有效语句, 2016 年 Lin 等^[10]将注意力模型与 CNN 结合来处理关系抽取任务。这种方法可以将选出的几个有用关系的语句信息进行加权组合, 提炼出对于关系抽取更准确的信息, 有效地提高了关系抽取的准确率。由于该方法在语句编码阶段仍然对句法信息表达不足, 故本文设计了一种基于句法树的 CNN 网络, 来提高对句法信息的利用。

本文主要工作是在句子级的水平上基于注意力机制的远程关系抽取中的注意力模型进行了优化。如图 1 所示, 本文利用 CNN 方法构建句子的语义信息。为了更准确地构建注意力机制模型, 通过两层相关性分析将语义信息投影在抽取关系上, 又通过利用来自《纽约时报》的实际数据对模型进行训练和测试。实验表明, 本文所提出的方法可以有效提升关系抽取的效果及运算效率。

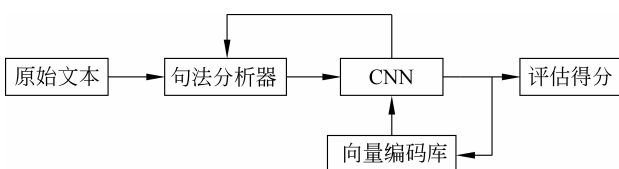


图 1 算法结构框图

1 相关研究

关系抽取是自然语言处理领域最重要的任务之一。远程监督关系抽取由于其基于远程数据库关联的特点, 是近年来热门的研究方法。但是远程监督关系抽取常常伴随着错误标注的问题。为了解决这个问题, 科研人员做出了不懈地努力。2010 年 Riedel 等提出了解决多示例单标签问题的远程监督模型^[5]; 随后, Hoffmann^[6]和 Surdeanu^[7]等在多示例多标签远程监督模型上提出了各自的方法。

自 2009 年以来, 深度学习已经在包括计算机视觉、语音识别等领域取得了突破性的进展, 在自然语言处理领域也是如此。2013 年 Socher 等的句法分析是深度学习在自然语言处理领域的成功应用。深度学习应用在关系抽取工作中是从 Socher 等将句法分析树结点表示为向量开始的。Zeng 等在 2014 年将端到端卷积神经网络应用到关系抽取中^[9], Lin 等在 2016 年将注意力模型应用在卷积神经网络之上, 提高了关系抽取的准确率^[10]。

但是, 难以获得充足的人工标注训练数据, 一直是关系抽取任务的困难之一。因而 2015 年 Zeng 等尝试用远程监督数据训练了多示例关系抽取的神经网络模型^[9]。但是该模型假定只有一条语句可以表达实体对所对应的关系, 大量有用信息被应用丢失。2016 年 Lin 等利用注意力模型对其进行了改进, 以充分利用所有语料信息。该模型认为通过这种方法可以充分利用所有的语料信息, 从而有效地进行关系抽取^[10]。在远程监督关系抽取时无法避免地会引入各种各样的噪声, 注意力模型具有一定的抗噪声能力, 我们希望能够提高这种能力。并且, 我们希望将更多的信息应用到关系抽取中, 除了语料本身的信息之外, 其语言环境、社会环境信息是不可以忽略的, 我们尝试将这部分信息应用到关系抽取中。

2014 年, Mou 等^[11]使用树形结构的卷积神经网络(Convolutional Neural Network, CNN)对程序语言进行了分析处理, 取得了很好的效果。但程序语言有严格固定且相对简单的语法结构, 因而并不能直接应用在自然语言上。2016 年, Dligach 等^[12]使用 CNN 和 LSTM 在病例语料上尝试了短时关系抽取, 发现 CNN 可以取得更好的效果。我们在此基础上更进一步, 通过对自然语言的时态等进行预处理后, 在自然语言的句法树上建立可控深度的 CNN 迭代模型, 并将其和一般 CNN 和 LSTM

进行比较,取得了良好效果。

2 算法模型

给定包含指定实体对 $(e1, e2)$ 的语句 s 及关系集 $Re = \{r_1, r_2, \dots, r_{m-1}, NA\}$,其中 NA 表示没有关系或其它关系。本文所提出模型最终将给出 $(e1, e2)$ 在语句 s 中存在关系 r 的概率 $P(r_i) = P_{e1, e2}(r_i | s)$,其中 $1 \leq i \leq m, r_m = NA, r_i \in Re$ 。模型主要由以下两个部分组成。

(1) 语句编码: 给定标出目标实体对 $(e1, e2)$ 的原始文本语句 s ,模型利用基于树形结构的卷积神经网络(Tree-Based CNN)将文本语句 s 表达为与 $(e1, e2)$ 相关的分布式向量 s ,该过程记为 $s = \text{Encoder}(s)$ 。

(2) 注意力机制: 利用已经表示好的句向量进行自适应加权,与关系向量进行匹配。

2.1 Tree-based CNN 语句编码

如图2所示,在对文本语句进行向量编码时,我们将CNN在句法树上进行逐级应用。即每级CNN只对语句中的一部分短语进行编码,然后再将短语的编码代入到CNN中,按照句法树对语句进行逐级编码,直到整个语句被编码为一个单一的向量为止。

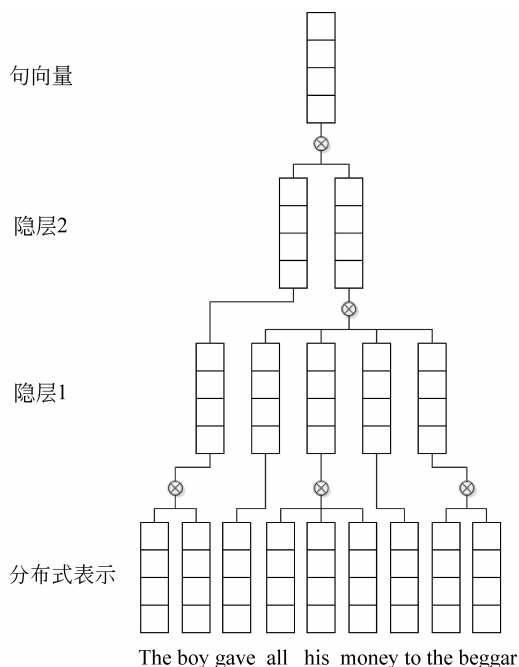


图2 语句编码示意图

对于每一层短语的编码方法则如图3所示,我

们利用前级输出的向量(或词向量)及该向量在本短语中的相对位置表示向量作为输入,然后采用一层卷积神经网络对其进行编码,经过卷积层、最大池层及非线性层,最终得到该短语的向量表示。

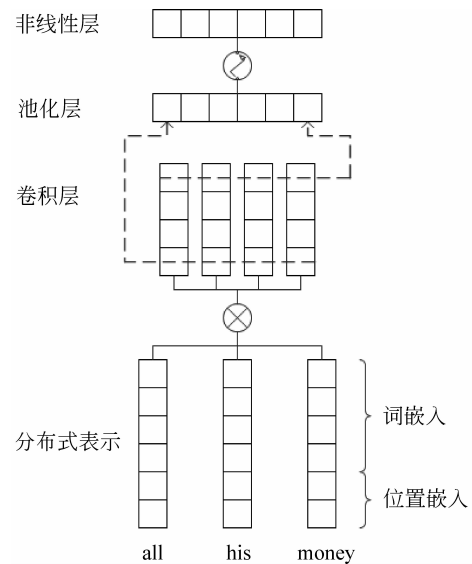


图3 短语编码示意图

2.1.1 输入向量表示

对于每个词或短语,我们通过向量库查询的方法找到其对应的向量表示,同时在其基础上嵌入位置表示。每个词的表示向量由两部分嵌入组成: 词义嵌入和位置嵌入。词义嵌入是为了捕捉词的词义信息,位置嵌入的作用是捕捉词相对实体的位置信息。

(1) 词义嵌入: 若查询矩阵为 $Q, Q \in R^{l^w \times |V|}$,其中 l^w 表示词义嵌入的长度, V 表示所有词的集合,则每个词的词义嵌入可以通过该矩阵进行编码。在图3中,词义嵌入对应词表示的 word embedding 部分。

(2) 位置嵌入: 为了跟踪词与目标实体的位置关系,词的位置信息需要嵌入到词的表示中去。若位置嵌入函数为 $pe(\cdot)$,则对于相对两实体位置分别为 $p1, p2$,位置嵌入表示为 $pe(p1, p2)$ 。其嵌入编码只与词与目标实体的相对位置有关。在图3中位置嵌入对应词表示的 position embedding 部分。

这种词表示的方法在实体关系抽取任务中可以有效地同时捕捉词的词义信息和位置信息。该方法2014年曾被Zeng使用^[9],2016年被YanKai Lin延伸^[10]。

2.1.2 短语向量表示

我们认为短语的向量表示和词的向量表示在向

量结构上并没有差异,即词向量与短语向量具有相同的维度。这样,当我们通过词向量表示整个句子的时候,可以按照句子结构通过对短语向量进行迭代分析,从而有效关注到句子的结构信息和词义信息。

本文对于短语的编码方法,采用基于门卷积的 CNN 方法。对于给定的一个短语,若其每个词的向量表示分别为 w_1, w_2, \dots, w_n , 其中 n 表示短语的长度,每个词向量表示的长度为 $l = l^w + l^p$, l^w 和 l^p 分别为词义嵌入和位置嵌入的长度。卷积核为 W , $W \in R^{l \times c}$, 其中 c 表示卷积核的宽度。那么在图 3 中卷积层 (convolutional layer) 的计算过程如式 (1) 所示。

$$p = W * [0 \cdots 0 w_1 w_2 \cdots w_n 0 \cdots 0] \quad (1)$$

其中 $[0 \cdots 0 w_1 w_2 \cdots w_n 0 \cdots 0]$ 为词表示向量按序排列并在前后分别补 $(c-1)$ 个零向量后组成的矩阵,其大小为 $l \times (2c+n-2)$, 表示矩阵间的卷积运算, p 为行向量。

为了保证短语向量与词向量在维度上兼容,对每条语句,我们取 l 个卷积核进行卷积运算。于是我们得到 l 个行向量,如式 (2) 所示。

$$p_i = W_i * [0 \cdots 0 w_1 w_2 \cdots w_n 0 \cdots 0] + b_i \quad (2)$$

若经过最大池层的输出的向量为 $v, v \in R^l$, 则 v 的第 i 个元素可以表示为式 (3)。

$$v_i = \max(p_i) \quad (3)$$

其中 $\max(p_i)$ 表示求 p_i 中值最大的元素。

最后经过非线性层,若非线性函数取双曲正切函数,最后输出的短语表示向量为 $x, x \in R^l$, 则 x 的第 i 个元素可以表示为式 (4)。

$$x_i = \tanh(v_i) \quad (4)$$

这个表示过程可以用图 3 表示,最后我们得到一个与词向量同维度的短语表示向量,进而可以将每条短语和词代入同一个语句表示过程进行一并计算。

2.1.3 语句向量表示

语句向量表示是按照语句结构对短语向量表示过程的一个迭代运算。给定任意一条语句,可以分为 m 条短语,若对于每条短语都已经得到一个向量表示则整个语句可以看作一条短语,进而用短语的向量表示方法将其表示为一个向量。若上节的短语向量表示方法用 $\text{encode_phrase}(\cdot)$ 表示,语句为 x_1, x_2, \dots, x_m , 则整个语句的表示向量 s 可以表示为式 (5)。

$$s = \text{encode_phrase}(x_1, x_2, \dots, x_m) \quad (5)$$

这个表示过程的迭代过程可以用图 2 表示,通过该方法,每条语句将被表示为一个长度固定为 l 的向量。

此外为了防止其过度迭代而破坏语句的基本结构,因而设置迭代深度控制参数 γ 。当迭代器迭代到指定深度时,直接将现有短语表示序列按照 CNN 的方法表示为一个向量。当 $\gamma=1$ 时,本实验中所提的方法将退化到基于普通 CNN 的关系抽取方法。

2.2 关注点选择方法

针对同一对实体 (head, tail), 若被 n 条语句 s_1, s_2, \dots, s_n 提及,且假定在这个 n 条语句中的实体对有相同的关系 $r, r \in Re, Re$ 为已知的关系集合。

首先,我们对这 n 条语句进行加权融合,如式 (6) 所示。

$$s = \sum_{i=1}^n \alpha_i \cdot s_i \quad (6)$$

其中 α_i 为关注系数。

由于并不是所有的语句 s 都能准确地对目标关系 r 进行描述,因此对于不同的语句应有不同的权重。关注系数 α_i 可由式 (7) 获得

$$\alpha_i = \frac{\exp(s_i A r)}{\sum_j \exp(s_j A r)} \quad (7)$$

其中 r 为关系向量, A 为参数矩阵。

最后我们将得到的融合语句向量 s 送入全连接层,输出向量 y 可表示为式 (8)。

$$y = M s + b \quad (8)$$

其中 b 为偏置向量。于是我们可以对神经网络的输出得用 softmax 分类器进行分类,从而找出该实体对最可能存在的关系 r 的概率为式 (9)。

$$P(r | s) = \frac{\exp(y_i)}{\sum_j \exp(y_j)} \quad (9)$$

其中 y_i, y_j 分别为 y 的第 i, j 维的值。

此处我们延用 2015 年 Zeng 等提出的假设,即在给定的语句中至少有一条语句可以反应出给定实体对存在的关系。经过该式的计算,对于融合向量 s 中蕴含的关系,该式会趋向于 1。而对于其他不能描述该关系的语句,该式将趋向于 0。

2.3 优化方法

我们采用交叉熵作为该算法的损失函数,其定义如式 (10) 所示。

$$J(A, M, b) = - \sum_i \log[P(r_i | S_i, A, M, b)] \quad (10)$$

其中 A, M, b 分别对应式(7)和式(8)中的同字母参数,是需要训练的神经网络参数。

为确定模型中的参数,我们需要用标注数据对其进行训练。为了加速训练,训练方法采用随机梯度下降法(SGD)。每次随机选择数据对其进行训练,直到收敛。

另外,每次在训练数据时需要对数据进行小的随机变动,以防止过拟合。

3 实验结果及分析

本文实验中将在广泛采用的来自《纽约时报》(NYT)的数据集进行测试,并和基于 CNN 和 LSTM 的典型方法以准确率和召回率作为评价指标进行对比。

3.1 数据来源及评测指标

本文采用一个在实体关系抽取领域广泛采用《纽约时报》(NYT)的公开数据集。该数据集曾在 2010 年被 Ridel 等^[5]、在 2011 年被 Hoffmann 等^[6]、在 2012 年被 Surdeanu 等^[7]、在 2016 年被 Yankai Lin 等引用^[10]。该数据集是结合《纽约时报》中的语料和 Freebase 中的实体集和关系集,利用斯坦福的命名实体识别工具,匹配结合而成。这个数据集被分为两部分,一部分是 2005~2006 年的语料,包含 522 611 条语句和 281 270 对实体,存在 18 252 种关系实例。另一部分是 2007 年的语料,包含 172 252 条语句,96 678 对实体,存在 1 950 种关系实例。本文实验中,将用前一部分语料进行训练,并用后一部分语料进行测试。与过去的评测关系抽取结果的方法相同,本实验中我们采用准确率和召回率进行评价。

3.2 实验参数设置及数据预处理

为了更好地验证提升效果,我们将分别对基于 Tree-based CNN 方法和基于普通 CNN 的方法和基于 LSTM 的方法进行实验,并在准确率和召回率两个指标下进行比较。

算法中存在一些经验参数设置,本实验中对于所有方法(Tree-based CNN /CNN/LSTM)的优化参数设置相同。根据以往的实验经验,同时借鉴 Yankai Lin 等^[10]的参数值设置,我们手动确定了一些可调参数。它们分别为 SGD 学习率 $\lambda=0.01$,词句表示维度 $l^w=50$,位置嵌入维度 $l^p=5$ 。此外,基

于 CNN 的和 Tree-based CNN 方法中卷积核的宽度都为 $c=3$,且 Tree-based CNN 的迭代深度控制参数 $\gamma=3$ 。事实上,与其他模型相比,这样的参数设置在实验中也的确使模型表现出了不错的效果。

另外,为了剔除一些不必要的词语信息干扰,我们除去了语料中包含出现次数少于 100 次的单词的语句。同时我们按照所提的标准,对语句进行了统一规范化的处理。

3.3 语句语法解析树及词向量

本文提出的算法,首先对语句进行了语法解析,其解析工具来源于斯坦福的自然语言处理工具^①。而对于词向量,我们则应用谷歌的 Word2Vec^② 工具对经过预处理后的 NYT 语料进行了训练。

3.4 实验效果

为了方便比较不同的方法对关系抽取的效果,我们分别利用 CNN、LSTM 和本文应用的 Tree-based CNN 方法在同一数据集上进行训练和测试。

表 1 记录了分别利用 CNN、LSTM 和本文应用的 Tree-based CNN 方法时,召回率分别为 0.2、0.3 和 0.4 时的准确率。必然的,当提高准确率时,三种方法的准确率都会有所下降,但是可以看出在这三个召回率下,本文所应用的方法都明显优于传统的 CNN 和 LSTM 算法。TBCNN 相对 LSTM,当召回率设置为 0.2 时,准确率提高最大,为 0.054;TBCNN 相对于 CNN,当召回率设置为 0.4 时,准确率提高最大,为 0.036。

表 1 不同算法准确率(Precise)对照表

	Recall=0.2	Recall=0.3	Recall=0.4
TBCNN	0.611	0.522	0.464
CNN	0.586	0.497	0.428
LSTM	0.557	0.474	0.417

图 4 中,我们以准确率和召回率为横纵坐标画出了 precise-recall 曲线。从该图我们可以看出,Tree-based CNN 整体优于一般的 CNN 和 LSTM 方法:一般的 CNN 方法,在提高召回率的同时,会

① <https://github.com/stanfordnlp/CoreNLP>

② <https://code.google.com/p/word2vec>

伴随准确率的急剧下降,而 LSTM 在要求高准确率的情况下,召回率则明显不足;而应用 Tree-based CNN 处理时,则兼顾了准确率和召回率。无论是在准确率还是在召回率水平上,Tree-based CNN 方法都能够发挥其优势。并在召回率为 0.3 时,Tree-based CNN 相对于 CNN 和 LSTM 明显占优,其准确率相对于 CNN 和 LSTM 分别提高了约 3% 和 5%。

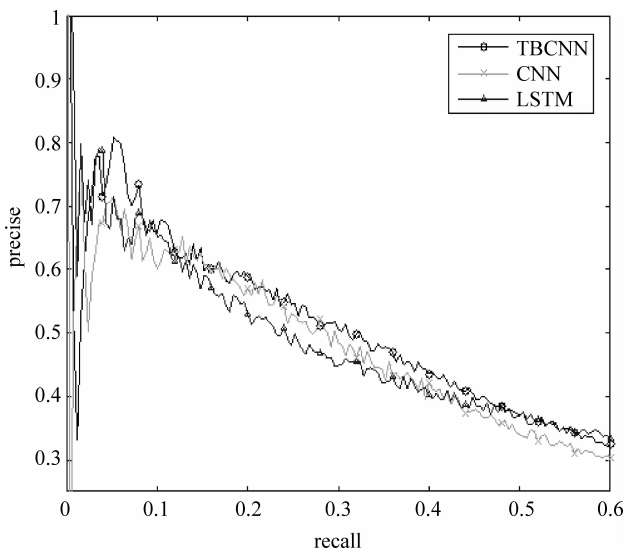


图 4 precise-recall 曲线

图 5 和图 6 分别展示了三种方法的 F_1 值和 $F_{0.5}$ 值曲线。可以看出,无论是 F_1 值还是 $F_{0.5}$ 值在所有相同召回率控制条件下,Tree-based CNN 处理结果都不低于 CNN 和 LSTM 的处理结果。且当召回率控制在 0.3~0.45 之间时,Tree-based CNN 的处理结果有明显优势。

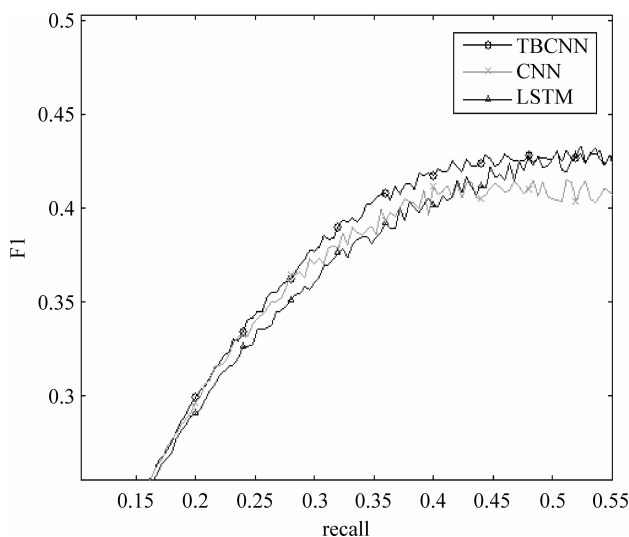


图 5 F_1 -recall 曲线

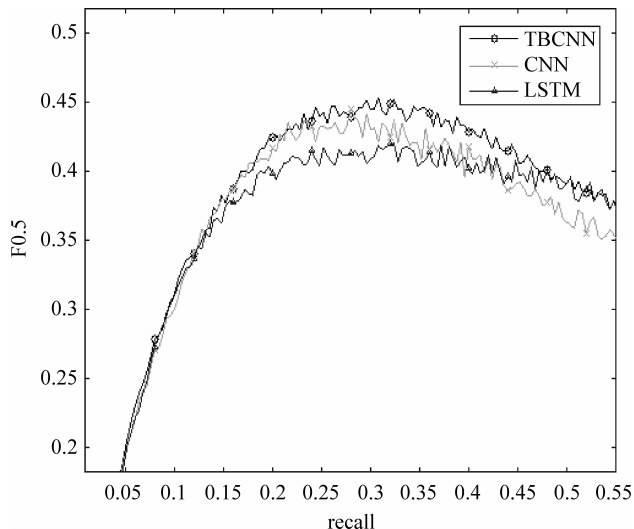


图 6 $F_{0.5}$ -recall 曲线

3.5 深入分析

基于 Tree-based CNN 的关系抽取方法是从基于 CNN 的关系抽取方法发展而来,其本质上是对 CNN 方法按照句法树的递归调用。当迭代深度控制参数 $\gamma = 1$ 时,Tree-based CNN 方法将退化到 CNN 方法。因而当召回率控制较低时,Tree-based CNN 与 CNN 在性能上表现出了相似性。但随着召回率的增加,其能够融合更多句法信息的优势则体现了出来,因而可以达到更高的准确率。

并且其递归调用的方法,将原来层状结构的神经网络变为了树状结构。由于其递归调用的方法,可以使得其对更长距离的文本信息进行融合提炼,在一定程度上讲,其与 LSTM 有一定的相似性。因而在控制高召回率时,Tree-based CNN 和 LSTM 在性能上表现出了相似性。当召回率控制较低时,Tree-based CNN 在关系抽取方面继承了 CNN 的优势,因而可以在准确率上相对 LSTM 达到更高。

4 结束语

本文在句子级水平上针对句子的句法树结构提出了利用 Tree-based CNN 对语句进行编码,并在关系抽取任务中,提高了 3%~5% 的精度。Tree-based CNN 是一种兼容 CNN 的递归神经网络,通过控制迭代深度可以退化到一般 CNN 模型。通过实验结果分析可知,基于 Tree-based CNN 的编码方式,可以在实体关系抽取任务中兼具 CNN 和

LSTM 的优势。

下一步,我们将继续发展 Tree-based CNN 的动态结构优势,并希望其在自然语言的其他任务中发挥优势。

参考文献

- [1] Auer S, et al. DBpedia: A nucleus for a Web of open data[C]//Proceedings of ISWC 2007, ASWC 2007, 2007: 722-735.
- [2] Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge[C]//Proceedings of the 16th International Conference on World Wide Web. ACM, 2007: 697-706.
- [3] Bollackerk, et al. Freebase: a collaboratively created graph database for structing human knowledge[C]//Proceedings of SIGMOG Conference. 2008: 1247-1250.
- [4] Mintz M, et al. Distant supervision for relation extraction without labeled data[C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. Association for Computational Linguistics, 2009: 1003-1011.
- [5] Riedel S, Yao L, McCallum A. Modeling relations and their mentions without labeled text[C]//Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases, 2010: 148-163.
- [6] Hoffmann R, et al. Knowledge-based weak supervision for information extraction of overlapping relations [C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011: 541-550.
- [7] Surdeanu M, et al. Multi-instance multi-label learning for relation extraction[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012: 455-465.
- [8] Socher R, Bauer J, Manning C D, et al. Parsing with Compositional Vector Grammars[C]//Proceedings of ACL 2013. 2013: 455-465.
- [9] Zeng, Dao J, et al. Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. [C]//Proceedings of the EMNLP 2015. 2015: 1753-1762.
- [10] Lin Y, et al. Neural relation extraction with selective attention over instances [C]//Proceedings of ACL 2016. 2016, 1: 2124-2133.
- [11] Mou L, et al. Convolutional neural networks over tree structures for programming language processing [C]//Proceedings of the 30th AAAI Conference. 2014: 1278-1293.
- [12] Dligach D, et al. Neural Temporal Relation Extraction [C]//Proceedings of the EACL 2017, 2017: 746.
- [13] Mikolov T, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv: 1301.3781, 2013.
- [14] Le Q, Mikolov T. Distributed representations of sentences and documents[C]//Proceedings of the 31st International Conference on Machine Learning (ICML-14). 2014: 1188-1196.



刘伟(1992—),硕士研究生,主要研究领域为自然语言处理。

E-mail: wliu_lcbx@126.com



黄瑞阳(1986—),助理研究员,主要研究领域为文本挖掘、图挖掘。

E-mail: 277433109@qu.com



陈鸿昶(1964—),教授,博士生导师,主要研究领域为网络分析。

E-mail: chc@ndsc.com.cn