

# Named Entity Recognition Based on Adversarial Generative Network

Jinye Cai<sup>1</sup>, Chao Li<sup>1</sup>, Zhao Qiu<sup>1\*</sup>, Xiangsheng Huang<sup>2</sup>

<sup>1</sup>School of Computer and Cyberspace Security, Hainan University, Haikou Hainan

<sup>2</sup>Institute of Automation, Chinese Academy of Sciences, Beijing

Email: 1062594302@qq.com, 1197648483@qq.com, \*5454734@qq.com, xiangsheng.huang@ia.ac.cn

Received: Jan. 23<sup>rd</sup>, 2020; accepted: Feb. 7<sup>th</sup>, 2020; published: Feb. 14<sup>th</sup>, 2020

---

## Abstract

With the passage of time and the rapid development of Internet technology, how to obtain useful information from a large amount of text data has become a new challenge. Named entity recognition is an important task in information extraction and information retrieval, and its purpose is to identify and classify the components that represent named entities in the text. It is widely used and plays an important role in many fields such as spam filtering, opinion analysis and message classification. Taking into account the advantages of using adversarial generation networks to better learn the feature distribution of sample data and the advantages of using variational autoencoders to better approximate real samples, this article uses the form of adversarial networks to separate the two current stages. The advanced models with characteristics are combined to design a named entity recognition algorithm model based on adversarial generative network to improve the accuracy and effectiveness of feature extraction.

## Keywords

Information Retrieval, Generative Adversarial Network, Named Entity Recognition, Feature Extraction

---

# 基于对抗生成网络的命名实体识别

蔡金晔<sup>1</sup>, 李超<sup>1</sup>, 邱钊<sup>1\*</sup>, 黄向生<sup>2</sup>

<sup>1</sup>海南大学计算机与网络空间安全学院, 海南 海口

<sup>2</sup>中国科学院自动化研究所, 北京

Email: 1062594302@qq.com, 1197648483@qq.com, \*5454734@qq.com, xiangsheng.huang@ia.ac.cn

收稿日期: 2020年1月23日; 录用日期: 2020年2月7日; 发布日期: 2020年2月14日

---

\*通讯作者。

## 摘要

随着时间的流逝, 互联网技术的迅速发展, 如何从大量的文本数据中获取对我们有用的信息成为了一种新的挑战。命名实体识别是信息抽取和信息检索中一项重要的任务, 其目的是识别出文本中表示命名实体的成分, 并对其进行分类。它在垃圾邮件过滤, 舆论分析和邮件分类等许多领域中广泛使用并发挥重要作用。考虑到采用对抗生成网络的形式能够更好地学习到样本数据的特征分布和采用变分自动编码器能够更好地接近真实样本的优点, 本文通过对抗网络的形式, 将两种现阶段具有各自特点的先进模型进行了结合, 综合设计了一种基于对抗生成网络的命名实体识别算法模型用以提高特征提取的准确有效性。

## 关键词

信息检索, 生成对抗网络, 命名实体识别, 特征提取

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

互联网的出现极大地推动了人们之间的交流, 以互联网为代表的信息技术的飞速发展更是极大地推动了整个社会的发展。随着信息产业的不断发展和互联网的飞速进步, 各种移动设备层出不穷, 相应地, 每天都会生成大量不同类型的数据。从海量数据中提取有价值的信息已成为新的挑战。作为 NLP 的经典应用场景之一, 命名实体识别提供了一种方便快捷的方法来为用户获取有用的信息。

作为自然语言处理中非常基本但却十分重要的任务[1]之一。命名实体识别同时也是许多自然语言处理任务中的重要基础工具(例如, 提取信息, 回答问题, 解析和机器翻译)。命名实体的检测是信息提取中的一个子任务, 它的目的是将文本中的命名实体本地化和分类为预定义的类别, 例如时间表达, 数量, 货币价值, 百分比等。检测命名实体的准确性决定了下游任务的有效性, 这是 NLP 的一个非常重要的基本问题。实体边界识别和确定实体类别是命名实体识别过程的两个方面。

现如今命名实体识别仍然存在很多的问题, 影响着信息检索领域的准确性。存在此类问题的原因主要有: 命名实体识别只是在有限的文本类型和实体类别中取得了效果; 与其他信息检索领域相比, 实体命名评测预料较小, 容易产生过拟合; 命名实体识别更侧重高召回率, 但在信息检索领域, 高准确率更重要; 通用的识别多种类型的命名实体的系统性很差。

本文在详细地了解与深入地研究了命名实体识别几个具体过程的基础上, 重点研究并提出了一种基于对抗生成网络的命名实体识别的算法模型, 结合了当前两个先进模型的各自特点, 致力于提高命名实体识别在检索领域的准确性。

## 2. 相关研究

作为自然语言处理中的最重要也是最经典的研究问题之一, 命名实体识别的发展大致经过了三个阶段: (1) 早期的基于词典和规则的方法, (2) 中期的基于传统的机器学习方法, (3) 深度学习的出现, 有关命名实体识别的研究进展的大概趋势大致如图 1 所示[2]。

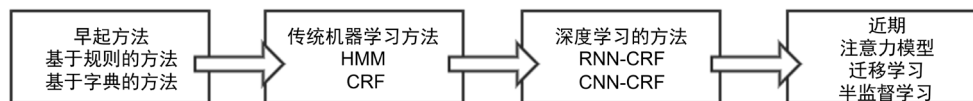


Figure 1. Research trends in named entity recognition  
图 1. 命名实体识别研究发展趋势

## 2.1. 基于规则和字典的方法

早期的基于规则和字典的方法主要是通过人工的方式，由语言学专家使用模式和字符串匹配作为主要方法来手工制定相应规则并进行特征的选取，最后形成对应的规则模板，这类系统大多数都是基于知识库和词典而建立的。基于规则和基于字典的方法是最早的用于命名实体识别的方法，它们基于手动规则系统，都使用命名实体库，并为每个规则分配权重。发生规则冲突时，将选择权重最高的规则以确定命名实体的类型。

## 2.2. 基于统计机器学习的方法

随着命名实体识别的研究不断展开，许许多多更加先进的算法方式和算法模型不断地得以提出，用以改进命名实体识别任务。其中，基于统计机器学习的方法的提出极大的改进了命名实体识别任务的准确率。作为统计机器学习方法的代表，其中最大熵模型、隐马尔可夫模型、条件随机场、支持向量机等著名的算法模型都得到了广泛的应用。

对于基于机器学习的方法来说，NER 一直被视为序列标记的问题，其使用大型语料库来学习出标记模型，用以标注句子的每个位置。在命名实体识别任务中常见的有：生成式模型 HMM、判别式模型 CRF。命名实体识别目前的主流模型是条件随机场[3]。

## 2.3. 基于深度学习的方法

随着大数据的更易获取以及 CPU 处理能力的提升，单词分布式表示形式的出现，神经网络已成为处理许多自然语言任务的有效模型。

这种类型的序列标记任务的处理方法是将单词标记从离散的 one-hot 表示映射到低维空间以表示密集的词嵌入，然后将句子的嵌入序列输入到递归神经网络中。网络强大的特征提取功能可提取文本中的特征，再使用 Softmax 计算每个单词的分数，最后使用 CRF 预测概率最高的标记序列。训练模型不再是常规流程，而是端到端功能，最终模型是一种基于数据驱动的方法，而不是基于特征工程。比较著名的有 Bi-LSTM 模型和 Bi-LSTM + CRF 模型[4]。

## 2.4. GAN 的提出

2014 年，随着生成对抗网络，即 GAN 的提出[5]，在深度学习领域可以说是掀起了又一股热潮，各种算法改进模型也随之层出不穷。同样的，基于生成对抗网络的命名实体识别算法的研究也跨出了一大步。最初，GAN 模型是用于生成图像这样的连续数据的，并不能直接用来生成离散数据。而当离散数据做微小改变时，在映射空间中也许根本就没有对应意义的序列，所以当 GAN 处理 NLP 这种离散数据的任务时，容易出现梯度消失的问题。此外，GAN 无法判断目前生成的某一部分序列的质量，因为它只能给生成的完整序列打分。

但是，通过近几年的研究在解决这些问题的道路上已经有所突破。SeqGAN [6]，WGAN [7]，CGAN [8]模型的提出成功的解决了原始 GAN 在序列标注问题中无法为生成器提供梯度的问题，在 NLP 领域的梯度消失问题，GAN 不能生成特定属性的问题以及训练梯度消失或者梯度爆炸的问题同时也提高了收敛速度。

相关的研究表明, GAN 在 NLP 领域已经取得了杰出的成果, 但在命名实体识别方面并没有太多的研究。因此本文提出了一种基于对抗生成网络的改进算法模型进行命名实体识别, 将变分自动编码器和对抗生成网络的优点相结合增强了命名实体识别中特征提取的准确有效性。

### 3. 基于生成对抗网络的命名实体识别算法模型设计

结合 GAN 可以完美的学习到训练样本的特征分布以及 VAE 更接近真实样本的优点, 本文提出了一种基于对抗性学习的多生成器多判别器命名实体识别算法模型, 旨在提高提取序列文本中特征分布的准确率。

#### 3.1. 设计思路

本小节介绍用于命名实体识别的改进模型的设计思路。如图 2 所示, 本文提出的基于生成对抗网络的命名实体识别算法模型主要分为两个部分: 分别为多生成器模型(G1 和 G2)和多判别器模型(D1 和 D2)。

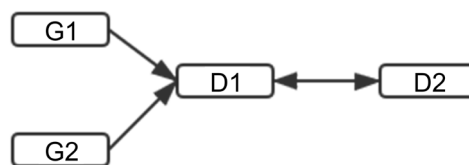


Figure 2. Named entity recognition algorithm model based on generative adversarial network  
图 2. 基于生成对抗网络的命名实体识别算法模型

生成器模型定义了给定句子序列的情况下生成该句子序列对应的命名实体识别标签序列的策略, 判别器模型对生成器模型所提取的命名实体进行判断是否充分和准确。本文采用了双生成器的形式, G1 和 G2 分别为两种不同的生成模型, 对给定的句子序列进行实体提取之后, 同时作为输入, 输入判别器网络 D1, 判别器网络 D1 用于判定分类标签属于哪个生成器网络, 当判别器网络 D1 无法判断分类标签属于哪个生成器网络时, 此时的两个生成网络便同时拥有了两种网络的特性, 再将分类结果输入判别器 D2, 对生成网络所提取出的命名实体的正确性进行判别, 将误差返回判别器 D1, 再由 D1 回传给生成器网络 G1 和 G2, 指导生成器网络进行更准确的实体提取。

#### 3.2. 生成器模型

生成器 G1 采用的是一种基于 GAN 的 Bi-LSTM + Attention + CRF 网络[9], 将句子序列输入到生成器 G1 中, 将未标注的句子作为条件信息, 用于生成标注, 通过双向 LSTM 网络获取每个单词的上下文信息, 然后将前后两个隐藏状态联结起来, 对上下文信息进行标签标记形成全局特征输出。然后将双向 BiLSTM 的输出向量输入到 Attention 层进一步的进行局部特征提取, 该特征表示传入到由专家标注数据训练的 CNN 中, 由 CNN 来判别这些特征分布与专家标注数据分布是否趋于一致, 如果一致则为正样本数据, 可以作为图 1 右侧所示模型的训练语料; 反之则为负样本数据, 重新传递回 BiLSTM 层对该层进行优化。生成器 G1 如图 3 所示。

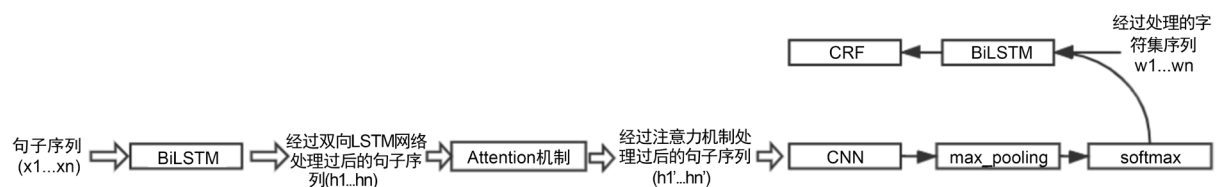


Figure 3. Generator G1  
图 3. 生成器 G1

生成器 G2 采用的是一种基于变分自动编码器的命名实体识别模型[10]，首先将句子序列作为输入，输入到编码器网络将原本句子序列编码成符合正态分布的特征参数，再将这些特征参数输入解码器网络还原生成相应的句子序列，再将生成的句子序列与原始的序列相比较，回传差异值，使编码器网络进行损失函数和参数的调整，直到最后生成序列和原始序列十分相似。

生成器 G2 如图 4 所示。

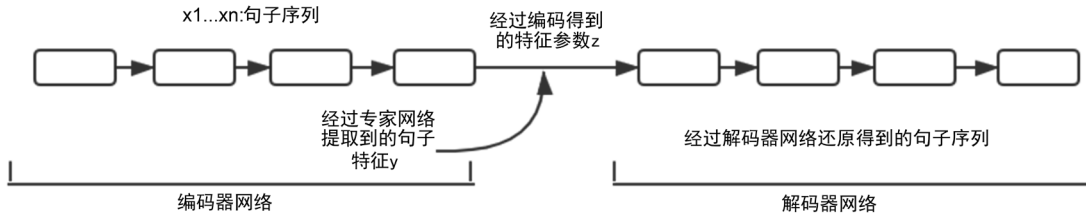


Figure 4. Generator G2  
图 4. 生成器 G2

### 3.3. 判别器模型

本文提出的判别器网络 D1 采用的是一种 CNN 判别模型[9]，该 CNN 是由生成网络 G1 所标注数据所训练的，判别器的输入是一种标记序列，为句子序列经由生成器 G2 处理后所得到的标记序列 S2，由 CNN 来判断这些特征分布与生成器 G1 所生成的标注数据是否一致，如果一致则表明两个生成器 G1 和 G2 已经拥有了相同的提取能力，如果不一致重新传递参数回到生成器 G1 和 G2，对两个生成器进行优化直到经由两个生成器所提取出来的特征标识拥有一致性。判别器 D1 如图 5 所示。

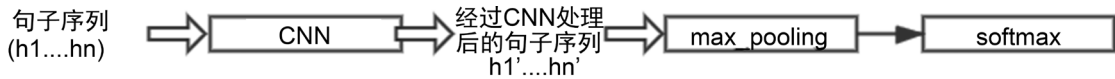


Figure 5. Discriminator network D1  
图 5. 判别器网络 D1

其中池化层我们选择的是 max-over-time pooling，该池化方法将具有最高值的特征视为最重要的特征，能有效的过滤掉大量的单词组合(信息量较少的)，并且可以保证输入的句子长度不会影响到提取的特征。最后使用一个 softmax 层来输出提取到的特征，以此来确定通过 G2 所提取的输入特征是否与 G1 标注数据特征分布一致。

本文提出的判别器网络 D2 是一种 BiLSTM 网络[11]。使用双向长短期网络分别评估每个单词的标签的得分。由于句子的长度是不固定的，所以我们需要将句子转换为具有固定长度的序列，该序列的长度被固定设为双生成器模型中制定的输出设置的最小长度。不同的是判别器的 BiLSTM 的输入有两种，分别为正实例输入(句子序列与对应的标注标签序列的连接)和负实例输入(句子序列与生成器生成的标签序列的连接)。将连接后的向量矩阵作为输入，并将其输入到 BiLSTM 网络，BiLSTM 网络结构为分别计算每个单词的上下文表示向量，并连接每个单词的上下文表示向量，最后通过一个全连接层为每个单词及其标签的连接序列的计算得分。判别器 D2 如图 6 所示。

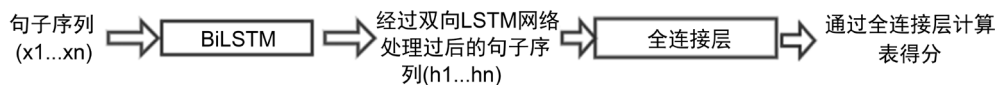


Figure 6. Discriminator network D2  
图 6. 判别器网络 D2

### 3.4. 与其他方法的原理对比分析

从原理上，本文设计的模型结合了两种先进模型分别能完美学习到样本特征分布和生成更加接近真实样本的特点，与现阶段各种模型相比较，解决了现阶段各模型提取特征不尽全面的缺陷，能够提取到更准确的样本特征分布，从而达到提高命名实体识别在信息检索领域的准确性的目的。以下将会用两种现有的命名实体识别模型为例说明：

1) CRF + Attention 模型[12] (如图 7 所示)。

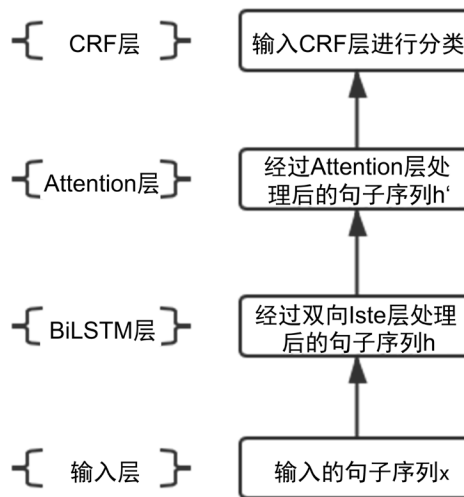


Figure 7. CRF + Attention model  
图 7. CRF + Attention 模型

该模型首先将句子序列输入双向 BiLSTM 网络，对序列中的词与词之间的前后向关系进行提取，再将处理过后的句子序列输入注意力层进行更进一步的特征提取，确保提取到与当前字向量对应的词向量特征，最后输入到 CRF 层通过输出层可以计算得到各时刻概率最高的标签序列，进而得出最优序列。但是该方法提取到的特征分布很大程度上受到了双向 BiLSTM 层提取特征关系是否准确的影响，如果双向 BiLSTM 层所提取到的特征关系并不完备，将会影响到最后的分类结果，使得模型不能完美的学习到样本数据的特征分布。

2) 单一的基于对抗生成网络的 BiLSTM 模型[9] (如图 8 所示)。

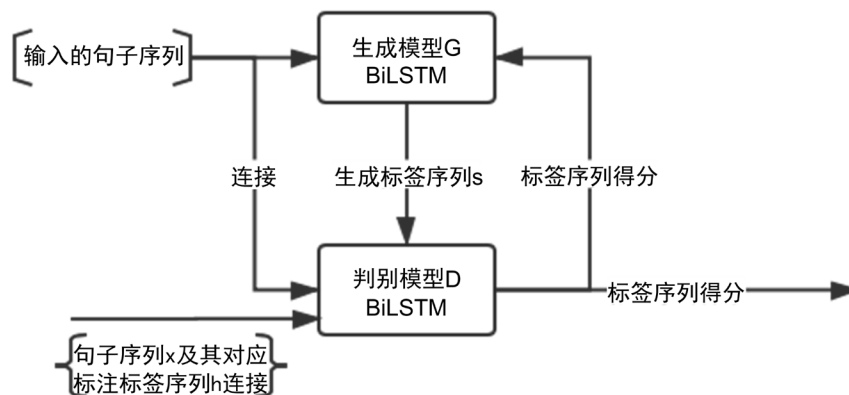


Figure 8. BiLSTM model based on adversarial generation network  
图 8. 基于对抗生成网络的 BiLSTM 模型

该模型生成器采用的是一个 BiLSTM 网络, 将未标注的句子序列输入双向 BiLSTM 层用以提取句子序列之间的上下文关系, 再通过全连接层以及 softmax 层得到每个单词在各个命名实体标签上的概率。判别器模型使用的也是一个 BiLSTM 网络, 将专家标注和生成器生成的标注序列作为输入, 为生成的标注序列的打分, 将标注序列得分返回生成器以达到优化分类的结果。但是对抗生成网络的生成器效果的好坏很大程度上依赖于判别器的返回, 如果判别器的返回不够完善或者甚至是出错, 都会对生成器最后提取到的结果产生很大影响使之不能完美的提取到特征之间的关系。

由此我们看出出现阶段的很多模型还存在着不能完美的学习到样本数据的特征分布以及不能完美的提取到特征之间的关系的的问题.因此文本提出了一种变分自动编码器和对抗生成网络的混合模型用于命名实体识别, 正好能够解决现阶段很多模型所存在的不足, 提高命名实体识别的准确性。

#### 4. 结语

随着科技的不断进步, 用于命名实体识别任务的算法和模型也不断地得以研究和发展, 近年来最具有代表性的算法模型便是基于生成对抗网络的命名实体识别模型和基于变分自动编码器的半监督命名实体识别。两种模型都有其明显的优势, 生成对抗网络能够完美地学习到样本特征的分布, 以及变分自动编码器网络能够更接近真实样本。同时也发现因为模型本身的限制, 两者都具有一定的限制性, 生成式对抗模型只能区分出生成样本的“真”、“假”, 没有办法强制它所生成样本必须看起来像原始样本, 变分自动编码器直接采用均方误差而非对抗网络, 其神经网络倾向于生成更为模糊的样本。

笔者综合考虑到了两种模型的优势和缺点, 于文中主要设计了一种基于对抗性生成网络的改进的命名实体识别算法模型, 将基于变分自动编码器的命名实体识别和基于 GAN 的命名实体识别模式通过双生成网络的形式结合在一起, 最终的模型将同时具有变分自动编码器更接近真实样本的特性和对抗性生成网络能完美学习到训练样本的分布特性, 该模型旨在提高命名实体识别问题中特征提取的高有效性和高准确性。

#### 基金项目

海南省重点研发计划项目“基于深度学习的智能客服系统研发(No. ZDYF2018017)”; 国家自然科学基金项目“复杂物体表面纹理获取和三维重建的关键技术研究(No. 61573356)”; 海南省自然科学基金项目“基于 YOLOv2 的监控视频人数统计研究(No. 618MS028)”。

#### 参考文献

- [1] 刘浏, 王东波. 命名实体识别研究综述[J]. 情报学报, 2018, 37(3): 329-340.
- [2] 王子牛, 姜猛, 高建瓴, 陈娅先. 基于 BERT 的中文命名实体识别方法[J]. 计算机科学, 2019, 46(S2): 138-142.
- [3] Zhao, H. and Kit, C. (2007) Unsupervised Segmentation Helps Supervised Learning of Character Tagging for Word Segmentation and Named Entity Recognition. *Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing*.
- [4] Huang, Z., Xu, W. and Yu, K. (2015) Bidirectional LSTM-CRF Models for Sequence Tagging. *Computer Science*. arXivpreprint arXiv:1508.0199
- [5] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., et al. (2014) Generative Adversarial Nets. *Proceedings of Neural Information Processing System*, Curran Associates, New York, 2672-2680.
- [6] Yu, L., Zhang, W., Wang, J., et al. (2016) SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient.
- [7] Arjovsky, M., Chintala, S. and Bottou, L. (2017) Wasserstein GAN.
- [8] Mirza, M. and Osindero, S. (2014) Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784.
- [9] 张晗, 郭渊博, 李涛. 结合 GAN 与 BiLSTM-Attention-CRF 的领域命名实体识别[J]. 计算机研究与发展, 2019, 56(9): 1851-1858.

- 
- [10] Xu, W., Sun, H., Deng, C., *et al.* (2016) Variational Autoencoders for Semi-Supervised Text Classification.
- [11] 冯建周, 马祥聪, 刘亚坤, 等. 关于命名实体识别的生成式对抗网络的研究[J]. 小型微型计算机系统, 2019, 40(6): 1191-1196.
- [12] 单义栋, 王衡军, 黄河, 闫倩. 基于注意力机制的命名实体识别模型研究——以军事文本为例[J]. 计算机科学, 2019, 46(S1): 111-114+119.