

文章编号: 2095-2163(2021)10-0155-05

中图分类号: TP399

文献标志码: A

# 基于生成式对抗网络的开放式信息抽取

韩家宝, 王宏志

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

**摘要:** 开放信息抽取(Open IE)是自然语言处理(NLP)的核心任务。尽管在这方面工作投入很多,但仍有许多问题需要解决。传统的开放式信息抽取方法使用一组手工定制的抽取模式从语料库中提取关系元组。同时,在程序中使用了许多自然语言处理工具;因此,将面临误差传播问题。为了解决这些问题,并受到最近成功的生成式对抗网络(GANs)的启发,文中采用了一种对抗训练架构,将其命名为 Adversarial-OIE。在 Adversarial-OIE 中,开放式信息抽取模型的训练由一个鉴别器辅助,这是一个卷积神经网络(CNN)模型。该鉴别器的目标是将 Open IE 模型生成的提取结果与训练数据进行区分。开放 IE 模型的目标是产生高质量的三元组来欺骗鉴别器。利用策略梯度方法对 Open IE 模型和鉴别器进行联合训练。通过实验证实,本文的方法明显优于许多现有的基准测试。

**关键词:** 生成式对抗网络;神经网络;开放式信息抽取

## Generative Adversarial Network based open information extraction

HAN Jiabao, WANG Hongzhi

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

**[Abstract]** Open Information Extraction (Open IE) is a core task of Natural Language Processing (NLP). Even many efforts have been made in this area, and there are still many problems that need to be tackled. Conventional Open IE approaches use a set of handcrafted patterns to extract relational tuples from the corpus. Meanwhile, many NLP tools are employed in their procedure; therefore, error propagation is highlighted. To address these problems and inspired by the recent success of Generative Adversarial Networks (GANs), the paper employs an adversarial training architecture, named as Adversarial-OIE. In Adversarial-OIE, the training of the Open IE model is assisted by a discriminator, which is a Convolutional Neural Network(CNN) model. The goal of the discriminator is to differentiate the extraction result generated by the Open IE model from the training data. The goal of the Open IE model is to produce high-quality triples to cheat the discriminator. A policy gradient method is leveraged to co-train the Open IE model and the discriminator. An empirical study shows that the proposed approach significantly outperforms many existing baselines.

**[Key words]** Generative Adversarial Networks; Neural Networks; open information extraction

## 0 引言

信息抽取是自然语言处理领域内的重要研究内容之一。研究目的是将文本结构化,结构化后的内容可以是三元组的表示形式,也可以是多元关系的表示形式<sup>[1]</sup>。传统的信息抽取方法侧重于在语料库上建立一套预定义的抽取范式<sup>[2-4]</sup>。因此,这些方法离不开大量的人工参与,为了减少信息抽取中的人工参与,研究人员提出了开放式信息抽取<sup>[5-6]</sup>,开放式信息抽取不局限于一组与定义的目标关系,而是提取文本中发现的所有类型关系。近年来,开放式信息抽取取得了可观的研究成果<sup>[4,7-20]</sup>,一系列的开放式信息抽取系统陆续被应用到各种工具当

中。这些系统普遍采用多种自然语言处理工具,因此也无一例外地面临着错误积累和传播的问题<sup>[21]</sup>。近期的工作多采用端到端的神经网络方法来进行开放式信息抽取的研究,这些方法虽然有效避免了错误积累和传播问题,但却没有考虑到曝光偏差问题,即:使用编码器-解码器架构处理序列到序列问题时,训练阶段与测试阶段所使用的数据不一致。

本文中,提出了基于生成式对抗网络(GAN)<sup>[6,22]</sup>的模型来应对开放式信息抽取任务,GAN模型不仅避免了传统方法带来的错误累积和传播问题,还能很好地解决曝光偏差问题<sup>[5,23-25]</sup>。此模型包含一个序列生成器,一个鉴别器。其中,序列生成器负责生成开放式信息抽取的结果,鉴别器用于鉴定生成器

**基金项目:** 国家自然科学基金(U1866602, 71773025);国家重点研发计划(2020YFB1006104)。

**作者简介:** 韩家宝(1989-),男,博士研究生,主要研究方向:大数据、自然语言处理、人工智能;王宏志(1978-),男,博士,教授,主要研究方向:数据质量、大数据、Web数据库。

**通讯作者:** 王宏志 Email: wangzh@hit.edu.cn

**收稿日期:** 2021-07-01

的结果是否来自训练数据。文中也对一个大型的 Open IE 基准数据集进行了研究,实验结果表明,该算法的性能优于几种常用的基准。同时,也证实了本文提出的模型要比单一的编码器-解码器模型好得多。此外,性能上的明显改进也证实了 GAN 模型在解决开放信息抽取任务中的曝光偏差问题上的有效性。

## 1 模型架构

生成式对抗网络已经成功应用于计算机视觉等领域以及一些其他重要的领域<sup>[26-30]</sup>。GAN 由 2 个模块组成,分别是:生成器  $G$  和鉴别器  $D$ ,其目的旨在估计当前输入施工训练数据而不是从生成器中产生的概率。在开放式信息抽取任务中使用 GAN 模型的价值函数为:

$$\min(G) \max(D) V(D, G) = E_{(x, y) \sim P_d(x, y)} [\log D(x, y)] + E_{(x, \hat{y}) \sim P_G(x, y)} [\log(1 - D(x, \hat{y}))] \quad (1)$$

其中,  $(x, y)$  表示训练实例;  $x$  表示输入;  $y$  表示输出;  $P_d$  表示数据分布;  $P_G$  表示生成器的参数分布。

如式(1)目标函数所述,鉴别器的学习目标是确定当前数据是来自训练数据、还是来自生成器的结果。生成器的训练目标是产生与训练数据相当的结果来混淆鉴别器。在实际的模型训练过程中,常用策略梯度<sup>[31]</sup>法来计算生成器的梯度。更新生成器参数模型前,先采样候选答案并使用鉴别器来计算奖励分数,继而利用所得奖励分数来计算生成器的梯度,再使用反向传播算法来更新一代参数。在训练过程中,将训练数据中的序列和采样后的候选序列分别作为鉴别器训练的正例和反例。为了解决鉴别器的过拟合问题,使用 MLE<sup>[32-34]</sup>提高生成器训练的稳定性。

图 1 中展示了 GAN 模型的整体架构,该模型是由 Transformer 模型构成的序列生成器和 CNN 模型构成的鉴别器组成。

### 1.1 序列生成器

对于序列生成器,编码器结构将输入序列映射到高维空间中的向量表示,解码器根据中间向量表示进行解码,与传统的网络结构不同,Transformer 模型不包含任何循环单元和卷积单元<sup>[35]</sup>。研究中,是通过正弦位置编码来记录单词的相对位置,具体公式可写为:

$$P(pos, 2i) = \sin(pos/10\ 000^{2i/d}) \quad (2)$$

$$P(pos, 2i + 1) = \cos(pos/10\ 000^{2i/d}) \quad (3)$$

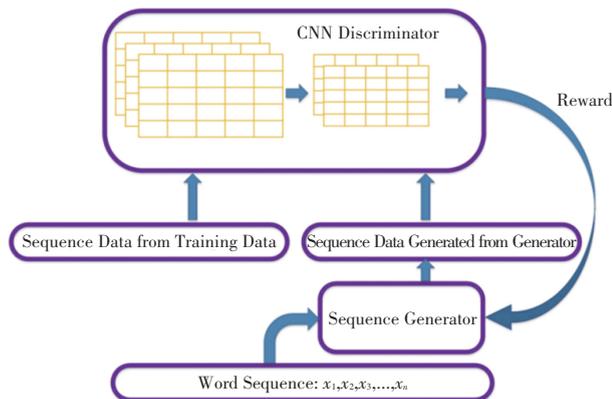


图 1 模型架构

Fig. 1 Model architecture

其中,  $pos$  表示单词的位置;  $i$  表示输入句子的维度;  $d$  表示编码器输出的维度。在第一个编码器和解码器部分,位置编码通过式(2)、式(3)进行计算。第  $j$  个编码器的输出  $S_e^j$  是通过一个自注意力层和一个前馈神经网络层进行计算得到:

$$O_e^{(j)} = LN(S_e^{(j-1)} + SelfAtt(S_e^{(j-1)})) \quad (4)$$

$$S_e^{(j)} = LN(O_e^{(j)} + FFN(O_e^{(j)})) \quad (5)$$

其中,  $O_e^{(j)}$  表示第  $j$  个编码器的注意力层的输出;  $LN(\cdot)$  表示的是归一化层;  $S_e^{(0)}$  表示的是编码器的输入;第  $j$  个解码器的输出  $S_d^{(j)}$  是由一个编码器-解码器的注意力层  $EDATT()$  产生的。此处需涉及的数学公式可写为:

$$O_d^{(j)} = LN(S_d^{(j-1)} + SelfAtt(S_d^{(j-1)})) \quad (6)$$

$$I_d^{(j)} = LN(O_d^{(j)} + EDATT(O_d^{(j)})) \quad (7)$$

$$S_d^{(j)} = LN(I_d^{(j)} + FFN(O_d^{(j)})) \quad (8)$$

其中,  $S_d^{(0)}$  表示解码器的输入;  $O_d^{(0)}$  表示第  $j$  个解码器自注意力层的输出;  $I_d^{(j)}$  表示第  $j$  个编码器的输入;最后一个解码器层的输出  $S_d^{(n)}$  被线性映射到  $V$  维的矩阵中,这里的  $V$  是输出词汇表的长度。

### 1.2 对抗模型

对于给定的输入序列  $x$ ,鉴别器需要区分当前结果是来自生成器的输出  $\hat{y}$ ,还是来自训练数据中的实序列  $y$ 。因此,研究使用卷积神经网络来度量序列对  $(x, y)$  的匹配度。考虑到卷积神经网络具有逐层卷积和池化的功能,在该项任务中则有着独特的优势。卷积神经网络可以准确地捕获  $(x, y)$  在不同空间中的对应关系,给出给定序列对  $(x, y)$ ,先通过简单地连接标记为  $x$  和  $y$  的嵌入向量构造  $2D$  表示。对于  $x$  中的第  $i$  个单词  $x_i$  和  $y$  中的第  $j$  个单词  $y_j$ ,有以下特性映射:

$$Z_{i, j}^{(0)} = [x_i^T, y_j^T]^T \quad (9)$$

基于这种 2D 表示, 这里的卷积窗口设置为  $3 \times 3$  大小, 通过以下的类型  $f$  的特征映射来捕获  $x$  与  $y$  的映射关系:

$$\mathbf{Z}_{i,j}^{(1,f)} = \sigma[\mathbf{W}^{(1,f)} \mathbf{Z}_{i,j}^{(0)} + \mathbf{b}^{(1,f)}] \quad (10)$$

其中,  $\sigma(\cdot)$  表示 *sigmoid* 激活函数,  $\sigma(x) = 1/(1 + \exp(-x))$ , 再使用一个  $2 \times 2$  的窗口进行池化操作:

$$\mathbf{Z}_{i,j}(2,f) = \max[\mathbf{Z}_{2i-1,2j-1}^{(1,f)}, \mathbf{Z}_{2i-1,2j}^{(1,f)}, \mathbf{Z}_{2i,2j-1}^{(1,f)}] \quad (11)$$

在此基础上, 将提取的特征输入到一个全连通层, 利用上一层的 *sigmoid* 激活函数, 求得训练数据中  $(x, y)$  的概率。该鉴别器的优化目标是以训练数据  $(x, y)$  为正例, 序列生成器采样数据为负例, 以最小化二值分类的交叉熵损失为目标。

训练目标: 生成器模型试图生成一个高质量的序列  $\hat{y}$  来欺骗鉴别器。对于鉴别器  $D$ , 使用训练集和  $G$  生成的序列对  $D$  进行训练。形式上  $D$  的目标函数是使  $V(D, G)$  最大化:

$$L_D = E_{(x,y) \sim Pd(x,y)} [\log P(x | y)] + E_{x \sim Pd(x), \hat{y} \sim P(y|x)} [\log(1 - P(x | \hat{y}))] \quad (12)$$

对于序列生成器  $G$ ,  $G$  的目标是最大化期望奖励 ( $D$  的概率), 而不是直接最小化  $V(D, G)$ , 函数如下:

$$L_G = E_{x \sim Pd(x), \hat{y} \sim P(y|x)} [P(x | \hat{y})] \quad (13)$$

鉴别器的训练过程与传统模式训练没有区别, 只需要向鉴别器提供发生器的输出和训练数据。发生器的训练过程不同于鉴别器的训练过程, 因为从发生器的离散采样结果  $\hat{y}$  使得从鉴别器直接反向传播误差信号到发生器要较为困难, 使  $V(D, G)$  对  $G$  的参数不可微。为此, 研究中使用了 *enhance* 算法来优化生成器  $G$ 。

## 2 实验

研究中使用了大型基准数据集 OIE2016<sup>[36]</sup>, 包含 3 200 个句子, 共 10 359 个提取, 包含 24 296 个句子, 共 56 662 个提取。为了验证本文提出方法的性能, 仿真中使用了数个最先进的基准测试, 包括 OLLIE、ClausIE、Stanford OpenIE、PropS 和 OPENIE4 来与本文方法进行比较。实验中, 使用了 2 个评价指标, 即: 精确度和查全率。

### 2.1 模型参数

研究中利用 *keras-transformer* 实现 *Adversarial-OIE*。本次实验中使用了 *Tesla P100 GPU*。该模型

包括 6 个编码器和 6 个解码器、768 维的隐藏状态和 512 维的单词嵌入。词汇量有 55 K 个。仿真时采用 *Relu* 作为注意激活函数和前馈激活, 退出率设为 0.05。同时选用 *Adam* 优化了本文模型。对于对手  $D$ , *CNN* 由 2 个卷积+池化层、一个全连接层和一个 *softmax* 层组成, 卷积窗口大小为  $3 \times 3$ , 池化窗口大小为  $2 \times 2$ , 特征大小为 20, 隐藏层的大小为 20。

### 2.2 实验结果

在 OIE2016 中的脚本作为本次研究的工具来评估精度和召回结果。仿真得到的精确率/召回率曲线如图 2 所示。在基准测试集 OIE2016 中, 将本文提出的模型与许多基准测试方法进行了比较。从实验数据可以看出, 与现有方法相比, 本文方法在精度方面有明显的优势。该模型的主要优点就是端到端神经网络模型, 避免了错误传播的问题。

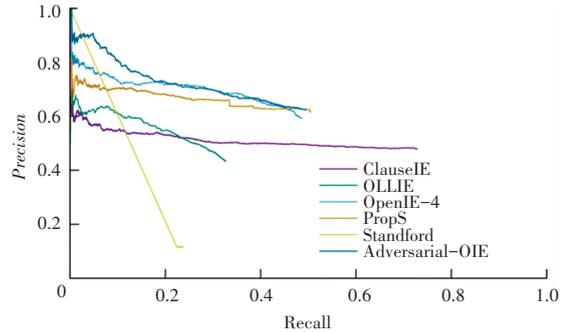


图 2 实验结果

Fig. 2 Experimental results

## 3 结束语

本文创新性地采用生成对抗网络模型来处理开放式信息抽取任务。研究中, 将开放式信息提取任务定义为序列到序列的任务。编码-解码器结构的网络模型处理序列到序列的任务会导致暴露偏差问题(由训练和推理过程中的文本生成不一致引起)。这种不一致性反映在推理和训练中使用的不同输入上。在训练过程中, 每个单词输入都来自于一个真实的样本, 但是当前用于推理的输入来自于上一个预测的输出。采用生成对抗网络模型可以有效地解决这一问题。生成对抗网络模型的再训练阶段给每个单词设定一个相应的奖励, 而非如同极大似然估计一样来增加单词出现的概率, 如此就不会产生曝光偏差问题。本文的模型尽可能地保证了鉴别器的可靠性。利用高质量的鉴别器改进了序列发生器。无需任何手工制作的模式和其他 NLP 工具, 序列生成器就可以生成精度更高的三元组。实验表明, 该模型具有较好的性能。从实验结果不难看出 GAN

模型的不足。也就是说,在一定的时间间隔内,精度会迅速下降。究其原因则在于 GAN 模型的对抗性训练不容易控制。在训练过程中,如果鉴别器模型有偏差,发生器模型就会被误导,产生一个劣质循环,这一偏差也会越来越大。最终,模型的性能会突然下降。

## 参考文献

- [1] WINGFIELD A, STINE-MORROW A L E. Language and speech [M]// Craik F I M, Salthouse T A. The Handbook of Aging and Cognition. Mahwah, Nj:Lawrence Erlbaum Associates, 2000; 359-416.
- [2] AGICHTEN E, GRAVANO L. Snowball: Extracting relations from large plain-text collections [C]//Proceedings of the Fifth ACM Conference on Digital Libraries. San Antonio, TX, USA: ACM,2000;85-94.
- [3] BRIN S. Extracting patterns and relations from the world wide web [C]//The World Wide Web, Databases, International Workshop WebDB'98. Valencia, Spain ;Springer,1998,1590;172-183.
- [4] RILOFF E, JONES R. Learning dictionaries for information extraction by multi-level boot-strapping [C]//Proceedings of The 16<sup>th</sup> National Conference on Artificial Intelligence. Orlando, Florida, USA: AAAI Press, 1999;474-479.
- [5] BANKO M, CAFARELLA M J, SODERLAND S, et al. Open information extraction from the Web[J]. Communications of the ACM,2008, 51(12) :68-74.
- [6] ETZIONI O, FADER A , CHRISTENSEN J, et al. Open information extraction: The second generation [C]//Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. Barcelona, Spain: AAAI Press, 2011,1;3-10.
- [7] CETTO M, NIKLAUS C, FREITAS A, et al. Graphene: Semantically-linked propositions in open information extraction [J]. arXiv preprint arXiv:1807.11276,2018.
- [8] KYUNGHYUN C, MERRIENBOER B V, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint arXiv:1406.1078,2014.
- [9] CHRISTENSEN J, MAUSAM, SODERLAND S. Semantic role labeling for open information extraction [C]// Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms, Methodology for Learning by Reading. Los Angeles, CA, USA: Association for Computational Linguistics, 2010;52-60 .
- [10] CORRO L D, GEMULLA R. Clause: Clause-based open information extraction [C]//Proceedings of the 22<sup>nd</sup> International Conference on World Wide Web. Rio de Janeiro:dblp, 2013;355-366.
- [11] FADER A, SODERLAND S, ETZIONI O. Identifying relations for open information extraction [C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, UK ;ACL, 2011;1535-1545
- [12] GASHTEOVSKI K, GEMULLA R, CORRO L D. Minie: Minimizing facts in open information extraction [C]// Proceedings of The 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark; Association for Computational Linguistics ,2017;2630-2640.
- [13] MAUSAM M. Open information extraction systems, downstream applications [C]// Proceedings of the twenty-fifth International Joint Conference on Artificial Intelligence. New York, NY, USA: AAAI, 2016;4074-4077.
- [14] HARINDER P, MAUSAM. Demonyms compound relational nouns in nominal open IE [C]//Proceedings of the 5<sup>th</sup> Workshop on Automated Knowledge Base Construction. San Diego, CA: Association for Computational Linguistics,2016;35-39.
- [15] GUARASCI R, DAMIANO E, MINUTOLOO A, et al. Lexicon-grammar based open information extraction from natural language sentences in Italian [J]. Expert Systems with Applications, 2020, 143;112954.
- [16] HERMANN K M, KOCISKY T, GREFFENSTETTE E, et al. Teaching machines to read, comprehend [C]//Advances in Neural Information Processing Systems. Montreal, Canada;NIPS, 2015: 1693-1701.
- [17] HILL F, BORDES A. The goldilocks principle: Reading children's books with explicit memory representations [J]. arXiv preprint arXiv:1511.02301,2015.
- [18] SAHA S, HARINDER P, MAUSAM. Boot-strapping for numerical open IE [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Vancouver; Association for Computational Linguistics, 2017;1-8.
- [19] YATES A, CAFARELLA M, BANKO M, et al. Texrunner: open information extraction on the web [C]//Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). Rochester, New York, USA: ACM, 2007: 25-26.
- [20] SCHMITZ M, SODERLAND M, BART S, et al. Open language learning for information extraction [C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing, Computational Natural Language Learning. Jeju Island, Korea; Association for Computational Linguistics, 2012: 523-534.
- [21] GABOR A, PREMKUMAR M J, MANNING C. Leveraging linguistic structure for open domain information extraction [C]// Proceedings of the 53<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics, the 7<sup>th</sup> International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Beijing, China:ACL, 2015;344-354.
- [22] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014. Montreal, Quebec, Canada; NIPS, 2014;2672-2680.
- [23] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align, translate [J]. arXiv preprint arXiv:1409. 0473,2014.
- [24] CUI L, WEI F, ZHOU M. Neural open information extraction [C]//Proceedings of the 56<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia; Association for Computational Linguistics, 2018;407-413.
- [25] GONG X, CHANG S, JIANG Y, et al. Autogan: Neural architecture search for generative adversarial networks [C]//2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019. Seoul, Korea (South) : IEEE, 2019;3223-3233.