

## 融合对抗学习的因果关系抽取

冯冲<sup>1,2</sup> 康丽琪<sup>1,2</sup> 石戈<sup>1,2</sup> 黄河燕<sup>1,2</sup>

**摘要** 因果关系抽取在事件预测、情景生成、问答以及文本蕴涵等任务上都有重要的应用价值。但多数现有的因果关系抽取方法都需要人工定义模式和约束,且严重依赖知识库。为此,本文利用生成式对抗网络(Generative adversarial networks, GAN)的对抗学习特性,将带注意力机制的双向门控循环单元神经网络(Bidirectional gated recurrent units networks, BGRU)与对抗学习相融合,通过重定义生成模型和判别模型,基本的因果关系抽取网络能够与判别网络形成对抗,进而从因果关系解释信息中获得高区分度的特征。实验结果表明,与当前用于因果关系抽取的方法相比较,该方法表现出更优的抽取效果。

**关键词** 因果关系抽取,生成式对抗网络,注意力机制,对抗学习

**引用格式** 冯冲,康丽琪,石戈,黄河燕.融合对抗学习的因果关系抽取.自动化学报,2018,44(5):811-818

**DOI** 10.16383/j.aas.2018.c170481

### Causality Extraction With GAN

FENG Chong<sup>1,2</sup> KANG Li-Qi<sup>1,2</sup> SHI Ge<sup>1,2</sup> HUANG He-Yan<sup>1,2</sup>

**Abstract** Causality extraction is of important practical value in tasks such as event prediction, scenario generation, question answering, and textual implication; but most of the existing causality extraction methods require artificial definition of patterns and constraints and are heavily dependent on knowledge base. In this paper, the bidirectional gated recurrent units networks (BGRU) with attention mechanism are merged with confrontational learning by leveraging the confrontational learning characteristics of generative adversarial networks (GAN). Through redefining the generator and discriminator, the basic causality extraction network can construct a confrontation with the discriminator, and then obtain a high distinguishing feature from the causality interpretation information. Our experiments show that our approach leads to an improved performance over strong baselines.

**Key words** Causality extraction, generative adversarial network (GAN), attention mechanism, adversarial learning

**Citation** Feng Chong, Kang Li-Qi, Shi Ge, Huang He-Yan. Causality extraction with GAN. *Acta Automatica Sinica*, 2018, 44(5): 811-818

因果关系是“原因”和“结果”之间的关系,是引起和被引起的关系。因果关系作为一种重要的关系类型,在许多任务(例如事件预测<sup>[1]</sup>、情景生成<sup>[2]</sup>、问答<sup>[3]</sup>、文本蕴涵等)中起着重要作用。自然语言文本中存在大量的因果关系。例如,新闻文本中经常报道事故及其造成的后果: The  $\langle e1 \rangle$  arrest $\langle /e1 \rangle$  has caused an  $\langle e2 \rangle$  outcry $\langle /e2 \rangle$  of indignation among some industrial and political circles in France. 该例包含了 arrest (逮捕)和 outcry (强烈抗议)之间的因果关系。其中,  $\langle e1 \rangle$ 、 $\langle /e1 \rangle$ 、 $\langle e2 \rangle$  和  $\langle /e2 \rangle$  是用来标识两个实体名词性词的位置指示标签<sup>[4]</sup>。

因果关系的抽取方法主要分为两类:基于模式

匹配的方法和基于机器学习的方法。基于模式匹配的方法包括词典语法模式、语义关系模式等。基于机器学习的方法包括支持向量机(Support vector machine, SVM)、决策树(Decision tree)等。这些方法主要依赖特征工程,但是特征工程涵盖了繁琐的特征选择和不准确的特征提取,计算开销大且存在错误传播的问题。以往的方法利用 WordNet、NomLex-Plus 以及 VerbNet 等知识库抽取特征,对知识库依赖性强。

为避免特征工程的一系列问题,本文在因果关系抽取任务上探索新的方法。自 2014 年以来,Goodfellow 等<sup>[5]</sup>提出的生成式对抗网络(Generative adversarial networks, GAN)在生成式任务上取得巨大进展。GAN 在结构上受到博弈论的启发,系统由一个生成模型和一个判别模型构成,如图 1 所示。生成模型捕捉真实数据样本的潜在分布,并生成新的数据样本;判别模型是一个二分类器,判别输入是真实数据还是生成的样本。判别模型的存在使得 GAN 中的生成模型能够学习去逼近真实数据,

收稿日期 2017-08-31 录用日期 2018-01-08  
Manuscript received August 31, 2017; accepted January 8, 2018  
本文责任编辑 李力  
Recommended by Associate Editor LI Li  
1. 北京理工大学计算机学院 北京 100081 2. 北京理工大学自然语言处理实验室 北京 100081  
1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081 2. Natural Language Processing Laboratory, Beijing Institute of Technology, Beijing 100081

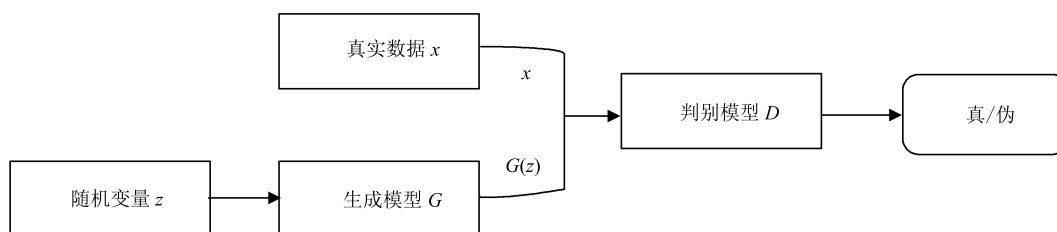


图1 GAN 结构

Fig. 1 The structure of GAN

最终让其生成的数据达到以假乱真的地步. 与传统基于模式匹配和机器学习的抽取方法相比, GAN 能够利用生成模型和判别模型之间的对抗, 获取逼近训练目标的生成模型.

针对已有因果关系抽取方法在特征工程中存在的问题, 本文提出了一个包含因果关系解释信息的对抗学习框架, 分别将原始语句和包含因果关系解释信息的语句输入到基本模型和因果关系增强模型, 利用两个相同结构的神经网络模型进行因果关系分类. 通过基本模型与判别模型的对抗训练, 基本模型能模仿和学习因果关系增强模型生成的特征向量, 进而捕获模型间共同特征, 提升因果关系抽取效果.

本文的主要贡献包括: 1) 提出一种融合对抗学习的因果关系抽取方法, 避免因采用特征工程而引起的一系列问题; 2) 针对因果关系抽取任务的特点, 提出“因果关系解释语句”的概念, 尝试将句子级别的信息作为关系抽取的先验知识. 通过糅合原句和因果关系解释语句, 借助生成对抗网络的对抗学习方法, 得到更加有效的因果关系抽取模型; 3) 为了让因果关系解释信息发挥更大的效用, 提出在融合对抗学习的基础上加入注意力机制, 充分发挥关键信息在最终分类中的作用.

本文内容安排如下: 第 1 节介绍相关工作; 第 2 节给出模型的具体架构; 第 3 节进行实验分析; 第 4 节总结全文.

## 1 相关工作概述

### 1.1 因果关系抽取

因果关系是关系类型中的一种, 对预测未来事件有重要作用. Radinsky 等<sup>[1]</sup> 通过从大型新闻语料库中提取因果关系, 预测未来事件; Hashimoto 等<sup>[2]</sup> 提出了一种从网络中提取因果关系的监督方法, 产生未来情景. 因果关系在问答系统的研究中也具有重要地位, Oh 等<sup>[4]</sup> 等针对问答系统中询问原因的一类问题, 探究词或子句之间的因果关系. 因果关系抽取是关系抽取中备受关注的一部分. 以往的方法多采用基于模式匹配或机器学习的方法, 利用词典语法模式<sup>[6]</sup>、上下文中的单词<sup>[4]</sup>、单词之间的关联关系<sup>[7]</sup>

以及谓词和名词的语义<sup>[8]</sup> 进行因果关系抽取. 相对于因果关系抽取, 关系抽取是一个更加广泛的概念, 通常关注的是多种关系类型的总体抽取效果, 没有针对因果关系类型进行深入研究.

多年来, 关系抽取任务中的方法层出不穷. Rink 和 Harabagiu<sup>[9]</sup> 利用外部语料库生成特征, 在 SVM 分类器上进行研究. Zeng 等<sup>[10]</sup> 提出用卷积神经网络 (Convolutional neural networks, CNN) 进行关系抽取, 采用词汇向量和词的位置向量作为卷积神经网络的输入, 通过卷积层、池化层和非线性层得到句子表示. 实体的位置向量和其他相关的词汇特征使得句子中的实体信息能够被较好地考虑到关系抽取中. 但是由于 CNN 不适合学习长距离的语义信息<sup>[11]</sup>, 所以很多学者尝试在循环神经网络 (Recurrent neural networks, RNN) 上进行关系抽取. Zhang 和 Wang<sup>[12]</sup> 使用双向循环神经网络抽取文本中的关系, 虽然该模型能利用上下文信息, 但是上下文的范围受限于梯度消散问题. 为此, Zhang 等<sup>[13]</sup> 提出双向长短期记忆网络 (Bidirectional long short-term memory networks, BLSTM), 该模型利用自然语言处理工具和词汇资源来获取特征, 利用 LSTM 单元实现了较为理想的结果. Zhou 等<sup>[14]</sup> 提出了基于注意力机制的双向长短期记忆网络 (Bidirectional long short-term memory networks, BLSTM), 该模型仅将带有实体位置指示词的 SemEval-2010-Task8 数据集作为输入, 不依赖任何词汇资源或自然语言处理系统, 简化了关系抽取的过程, 并且取得了可观的结果. 因果关系抽取任务关注的是长序列中两个实体名词性词之间的关系. 考虑到上下文信息以及训练速度, 本文采用构造更简单且训练更快的 BGRU<sup>[15]</sup> 网络.

### 1.2 生成式对抗网络

自从 2014 年 Goodfellow 等<sup>[5]</sup> 提出 GAN 以来, GAN 在图像领域和自然语言处理领域展现了极大潜力. 在自然语言处理领域, GAN 的应用主要体现在文本生成任务上. Zhang 等<sup>[16]</sup> 以 LSTM 作为 GAN 的生成模型进行文本的生成. Yu 等<sup>[17]</sup> 将序列生成过程当作一个序列决策的过程, 将误差作为

一种增强学习的奖励, 以一种前馈的方式训练, 并采用增强学习的探索模式更新生成模型. Li 等<sup>[18]</sup> 利用对抗性训练进行开放式对话生成, 将对话生成任务看作强化学习的问题, 联合训练生成模型和判别模型. 虽然 GAN 被应用到了诸多领域, 但在我们掌握的文献范围内本文是第一个将其应用于因果关系抽取任务的尝试与研究.

GAN 从出现以来就存在训练困难、生成样本缺乏多样性等问题. 近日, Facebook 人工智能研究中心 (Facebook Artificial Intelligence Research Center, FAIR) 提出 Wasserstein GAN (WGAN)<sup>[19]</sup>, 引入 Wasserstein 距离, 既解决了训练不稳定的问题, 也提供了一个可靠的训练进程指标. 而 WGAN 仍旧存在训练困难和收敛速度慢等问题. 因此, Gulrajani 等<sup>[20]</sup> 提出了相应的改进方案, 修改了原设计中 Lipschitz 的施加方式, 将权重剪裁 (Weight clipping) 改为梯度惩罚 (Gradient penalty). 实验表明 WGAN-GP 能够显著提高训练速度, 解决了原始 WGAN 收敛缓慢的问题. 鉴于 WGAN-GP 的明显优势, 本文采用 WGAN-GP 的训练方法.

## 2 模型

### 2.1 模型架构

针对因果关系抽取任务, 引入外部资源解释补充因果关系有助于提升模型在因果关系分类上的效果. 通过利用 WGAN-GP 能够让两个数据分布靠近的特性, 本文尝试在另一数据分布中引入外部资源, 使得关系分类模型能够在对抗学习中学习到外部资源提供的高区分度的分类特征, 从而提高关系分类效果.

在本文提出的模型中, 两个数据分布均为通过 BGRU 抽取后得到的特征向量表示. 由于因果关系

解释语句能提供更有区分度的特征, 因而在本文的方法中, 除了需要与 GAN 判别模型对抗的关系分类模型 (基本模型), 还需设计一个包含因果关系解释信息的增强模型 (因果关系增强模型). 因果关系增强模型相当于人工注释器, 为每个输入的句子添加关系解释语句, 使得句子的分类特征更加明显. 因果关系增强模型拥有和基本模型相同的网络结构, 经过对其输入的额外处理, 该模型得到的特征向量会包含更多利于分类的信息. 基本模型会向因果关系增强模型学习句子的特征向量表示, 尤其是特征向量中具有高区分度的部分, 使得两个模型抽取的潜在特征尽可能地接近, 从而获得更好的分类结果.

该学习过程由基本模型与判别模型通过对抗的方式推动进行. 在对抗学习中, 判别模型试图增大基本模型与因果关系增强模型所得特征向量的差异, 而基本模型希望两个模型得到的特征向量越来越接近. 最终, 在优化基本模型和判别模型的过程中, 两个数据分布会逐渐靠近, 即基本模型的特征向量会尽量趋近于关系增强模型的特征向量, 使最终分类器的关系抽取结果得到提升. 本文模型的整体架构见图 2.

#### 2.1.1 关系抽取网络

关系抽取网络由基本模型、因果关系增强模型以及分类器组成. 基本模型的输入是带有实体名词性词标签的句子, 因果关系增强模型的输入要在此基础上加入因果关系解释语句. 因果关系解释语句是依据因果关系的相关定义及解释, 预先准备的多种解释语句模板. 使用模板时需要依据原句将两个实体名词性词填充其中, 从而显式表达出原句的因果关系. 在因果关系增强模型的输入部分, 针对每条待抽取关系的句子, 每次随机选择一个解释说明语

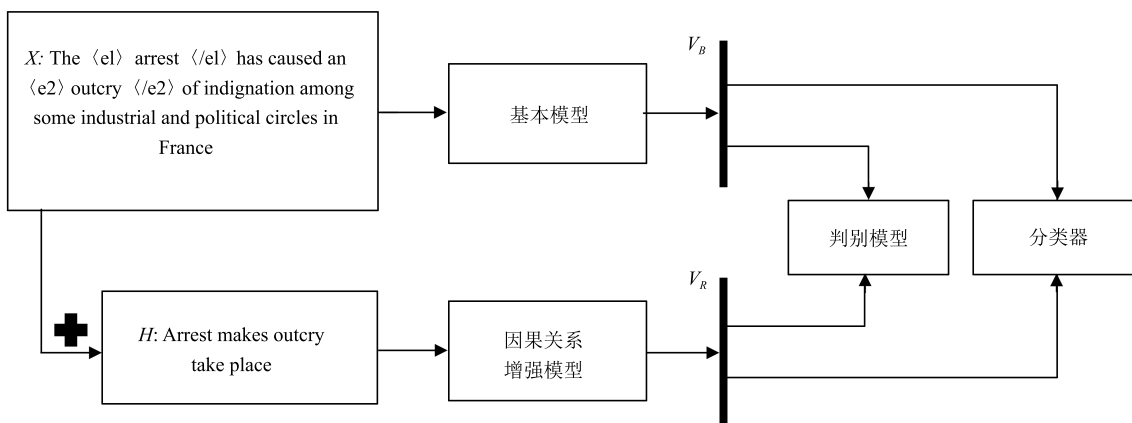


图 2 模型整体架构

Fig. 2 The overall architecture of the model

句加入到原句之后,组成新的句子.例如:

原句: The <e1> arrest </e1> has caused an <e2> outcry </e2> of indignation among some industrial and political circles in France.

因果关系解释语句: Arrest makes outcry take place.

新句: The <e1> arrest </e1> has caused an <e2> outcry </e2> of indignation among some industrial and political circles in France. Arrest makes outcry take place.

关系: Cause-Effect (e1, e2).

根据关系的方向性,因果关系类型又可以分为 Cause-Effect (e1, e2) 和 Cause-Effect (e2, e1) 两种,括号内前者为原因,后者为结果.本文的因果关系抽取网络中能够同时对关系及其方向性作出判断.

基本模型和因果关系增强模型均采用双向门控循环神经网络,整个关系抽取网络的架构见图 3. 网络由输入层、词向量层、BGRU 层、注意力层以及输出分类层构成. 其中,最后一层分类层被两个模型共享,又称分类器. 输入层将句子转化成单词序列输入到网络中. 词向量层利用预训练的词向量,把每个单词转化成对应的词向量,句子中的实体名词性词位置标签同样作为单词转为词向量. 对于 BGRU 层,一方面,GRU 是 LSTM 的一个变种,与 LSTM 一样能够缓解梯度消散的问题,同时能够提高训练速度;另一方面,BGRU 包括两个 GRU 的子网络,能够获取序列中上下文的信息. 所以通过该层的处理,网络可以得到更高层次的特征.

鉴于注意力机制在序列学习任务上具有巨大的提升作用,本文在模型中加入注意力机制. 注意力层

使得对句子分类的注意力更加集中在对因果关系分类有决定性作用的词上. 为解释注意力层的处理过程,用  $H$  表示包含 BGRU 层输出向量  $[h_1, h_2, \dots, h_T]$  的矩阵,其中  $T$  表示句子长度. 经注意力层得到的句子表示  $r$  由这些输出向量的加权和得到.

$$M = \tanh(H) \quad (1)$$

$$\alpha = \text{softmax}(w^T M) \quad (2)$$

$$r = H\alpha^T \quad (3)$$

其中,  $H \in \mathbf{R}^{d^w \times T}$ ,  $d^w$  是词向量的维度,  $w$  是训练好的向量,  $w^T$  是  $w$  的转置. 在本文模型中,基本模型与因果关系增强模型得到的特征向量为

$$h^* = \tanh(r) \quad (4)$$

最后,两个网络都通过 softmax 层给句子输出一个分类.

### 2.1.2 判别模型

判别模型是 GAN 中最重要的构成之一,与生成模型一同完成 GAN 的竞争对抗. 针对因果关系抽取任务,本文的判别模型被建模为多层感知机 (Multi-layer perceptron, MLP),由 4 个全连接层组成. 判别模型近似拟合基本模型与因果关系增强模型所得特征向量之间的 EM (Earth mover) 距离.

## 2.2 训练过程

采用 GAN 最新的训练方案 WGAN-GP,根据判别模型的输入后向计算出权重梯度,并针对梯度的范数进行惩罚,从而解决原始 GAN 以及 WGAN

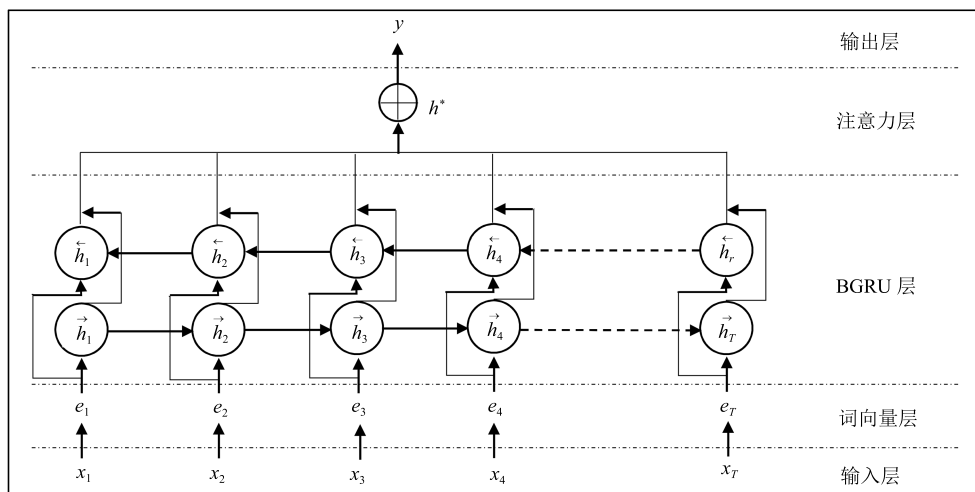


图 3 带注意力机制的双向门控循环神经网络

Fig. 3 Bidirectional GRU model with attention

存在的训练缺陷.

为了更清楚地说明训练过程, 用  $\theta_B$  和  $\theta_R$  分别表示基本模型和因果关系增强模型的参数, 用  $\theta_C$  表示分类器的参数, 用  $\theta_D$  表示判别模型的参数. 设定输入句子为  $x_b$ , 添加关系解释语句后的句子为  $x_r$ , 正确标签为  $y$ . 将基本模型得到的特征向量表示为  $\mathbf{V}_B$ , 因果关系增强模型得到的特征向量表示为  $\mathbf{V}_R$ .

在对抗训练开始之前, 预训练基本模型和因果关系增强模型. 对于基本模型的预训练, 训练目标是利用基本模型将句子中两个名词性词之间的关系正确分类, 即最小化交叉熵损失.

$$Obj^{B,C}(\theta_B, \theta_C) = \mathbb{E}_{(x_b, y) \sim \text{data}} [J(C(H_B(x_b; \theta_B); \theta_C), y)] \quad (5)$$

其中,  $J(C(H_B(x_b; \theta_B); \theta_C), y)$  表示基本模型特征向量通过分类器得到分类标签和真实标签  $y$  的交叉熵损失.

对于因果关系增强模型的参数, 类似于基本模型, 也通过最小化交叉熵损失进行学习.

$$Obj^R(\theta_R) = \mathbb{E}_{(x_r, y) \sim \text{data}} [J(C(H_R(x_r; \theta_R)), y)] \quad (6)$$

预训练基本模型和因果关系增强模型的过程是最小化式 (5) 和式 (6) 的过程. 由于基本模型和因果关系增强模型共享同一个分类器, 所以能形成一个统一的特征空间.

预训练之后, 基本模型和判别模型进行竞争对抗. 在对抗训练中, 基于带梯度惩罚的训练方法, 判别模型的损失函数为

$$Obj^D(\theta_D) = \mathbb{E}_{x_b \sim \text{data}} [D(H_B(x_b; \theta_D))] - \mathbb{E}_{x_r \sim \text{data}} [D(H_R(x_r; \theta_D))] + \gamma \mathbb{E}_{(x_b, x_r) \sim \text{data}} [\|\nabla_x D(\varepsilon x_b + (1 - \varepsilon)x_r)\|_p - 1]^2 \quad (7)$$

判别模型的训练目标是近似拟合两数据分布的 EM 距离. 其中,  $\mathbb{E}_{(x \sim \text{data})} [\cdot]$  表示数据分布的期望. 对于损失函数最后的惩罚项, 是将 Lipschitz 约束以差分形式加入到了惩罚中.

在对抗训练中, 基本模型的训练目标是要生成同因果关系增强模型产生的特征向量相近的特征向量. 为此, 固定因果关系增强模型和分类器的参数, 优化基本模型使得二者的 EM 距离缩小, 同时最小化基本模型的交叉熵损失.

$$Obj^B(\theta_B) = -\lambda \mathbb{E}_{x_b \sim \text{data}} [D(H_B(x_b; \theta_B))] + \mathbb{E}_{(x_b, y) \sim \text{data}} [J(C(H_B(x_b; \theta_B)), y)] \quad (8)$$

其中,  $\lambda$  用于平衡分类损失和调整特征损失权重. 在对抗训练过程中, 利用式 (7) 和式 (8) 分别训练判别模型和基本模型, 使基本模型和判别模型进行竞争, 促进两个模型的共同优化, 直到基本模型与因果关系增强模型得到的特征向量足够接近, 即基本模型能够抽取到高区分度的因果关系分类特征. 整个训练过程的算法如下:

**输入.** 训练数据和因果关系解释信息.

**输出.** 对抗学习后的基本模型及分类器.

**步骤 1.** 预训练基本模型和因果关系增强模型. 通过最小化式 (5) 和式 (6) 来分别初始化  $\theta_B$ ,  $\theta_C$  和  $\theta_R$ .

**步骤 2.** 进行对抗训练. 用式 (7) 训练判别模型.

**步骤 3.** 用式 (8) 训练关系模型.

**步骤 4.** 若未收敛, 转步骤 2.

**步骤 5.** 算法结束.

### 3 实验

本文设计并完成了 BGRU 因果关系抽取模型预训练、对抗训练、带注意力机制的对抗训练以及与已有因果关系抽取方法的对比四部分实验. 在对抗训练之前, 先进行 BGRU 因果关系抽取实验, 即对基本模型和因果关系增强模型进行预训练; 对抗训练实验是在上述实验的基础上加入 GRU 框架, 利用 GRU 的对抗学习特性进行训练学习; 为发挥因果关系解释信息的关键作用, 在上述实验中加入词级别的注意力机制; 最后给出传统因果关系抽取方法 SVM<sup>[10]</sup> 的结果和已知文献中最好的抽取方法<sup>[11]</sup> 与本文提出的方法的对比.

#### 3.1 数据集

使用的数据集来自两个部分. 1) SemEval-2010-Task8 的因果关系数据和非因果关系数据, 但因果关系数据的数量不能满足本文神经网络模型的需求; 2) 是人工标注新闻语料后得到的因果系数数据. 数据集共包含 3 931 条数据, 其中包含因果关系的数据有 2 031 条. 数据集的构造说明见表 1.

表 1 数据集的构造说明

Table 1 Description of the dataset

关系类型	数据来源	数据条数
因果关系	SemEval	1 331
	新闻语料人工标注	700
非因果关系	SemEval	1 900

为了进行评估, 将语料按照 4:1:1 的比例随机

分成训练集、开发集和测试集, 分别用于模型的训练、超参数的调优和模型的评估.

### 3.2 BGRU 因果关系抽取模型预训练

基本模型和因果关系增强模型是本文提出的因果关系抽取方法的重要组成部分, 为评价 GAN 框架发挥的效用, 必须对比分析包含因果关系解释信息的因果关系增强模型与基本模型的关系抽取结果.

基本模型和因果关系增强模型结构相同, 都采用双向门控循环单元神经网络, 包括输入层、词向量层、BGRU 层以及输出层. 输出层输出的是分类器分类后的结果. 使用 word2vec 在包含实体名词性词位置指示标签的新闻语料上训练出 200 维的词向量. 在基本模型和因果关系增强模型的训练中, 使用学习率为 0.005 的 Adam 进行训练, batch 设为 500. 模型参数采用  $10^{-5}$  的 L2 正则化. 本文在 BGRU 层使用 dropout, 将其设为 0.5. 实验结果见表 2.

表 2 BGRU 因果关系抽取结果 (%)

Table 2 Results of BGRU causality extraction (%)

Model	P	R	F1
B-BGRU	92.93	85.98	89.32
R-BGRU	93.74	94.39	94.06

表 2 中的 B-BGRU 是基本模型, 此处的抽取结果是该模型在预训练之后通过分类器得到的数据结果. R-BGRU 是因果关系增强模型, 与 B-BGRU 拥有相同的网络结构. 但 R-BGRU 在句子输入时能够根据句子中的标签随机加入因果关系解释语句. 此处的抽取结果同样是其预训练后的分类结果. 实验结果表明, 因果关系增强模型的抽取结果高于基本模型. 由于训练目标是让 B-BGRU 学习 R-BGRU 的特征向量, 从而在分类结果上靠近 R-BGRU, 因此该实验结果是对抗训练的基础.

### 3.3 对抗训练

在第 3.2 节实验结果的基础上, 将基本模型和因果关系增强模型置于 GAN 框架内, 基本模型与判别模型进行对抗训练, 使得基本模型的因果关系抽取结果得到提升. 实验结果见如表 3.

表 3 展示的是基本模型预训练后的抽取结果和在 GAN 框架下与判别模型对抗学习之后的抽取结果. 实验结果表明, GAN 的对抗训练使得基本模型的因果关系抽取结果得到明显提升. 在基本模型和判别模型的对抗训练中, 使用学习率为  $10^{-5}$  的 Adam 进行训练, 对于式 (8) 中的平衡参数, 设置  $\lambda = 0.2$ .

表 3 GAN 框架下的因果关系抽取 (%)

Table 3 Causality extraction under GAN framework (%)

Model	P	R	F1
B-BGRU	92.93	85.98	89.32
GAN-BGRU	93.75	87.62	90.58

### 3.4 带注意力机制的对抗训练实验

在基本模型和因果关系增强模型中加入词级别的注意力层, 将特征向量的注意力集中在对因果关系抽取有关键性作用的词上. 实验结果见表 4.

表 4 带注意力机制的 GAN 框架下的因果关系抽取 (%)

Table 4 Causality extraction under GAN framework with attention (%)

Model	P	R	F1
B-Att-BGRU	92.91	88.79	90.80
R-Att-BGRU	94.63	94.63	94.63
GAN-Att-BGRU	93.17	89.25	91.17

表 4 展示的是在加入注意力层的情况下, 基本模型和因果关系增强模型预训练后的抽取结果以及在 GAN 框架下基本模型与判别模型对抗学习之后的抽取结果. 实验结果表明, 最终模型 F1 值达到了 91.17%, 明显高于表 3 中的 90.58%, 说明加入的注意力机制进一步优化了抽取结果.

### 3.5 与已有因果关系抽取方法的对比实验

在因果关系抽取任务上的研究多采用人工设计特征的传统方法, 本文首先借鉴以往工作<sup>[10]</sup>, 利用模型 SVM 在构造的数据集上进行因果关系抽取. 对于已有文献中最好的关系抽取方法, 根据论文<sup>[11]</sup> 中网络结构及相关参数的设置, 对其关系抽取模型进行复现, 最终通过在数据集上的实验得到抽取结果.

图 4 汇总了不同模型的因果关系抽取结果. 其中 Att-BLSTM 是本文复现的关系抽取网络, 在本文数据集上的 F1 值达到了 89.21%, 而本文最终的模型 GAN-Att-BGRU 的 F1 值达到了 91.17%, 因果关系抽取结果有明显提升.

实验结果说明, 基本模型在 GAN 的对抗训练中, 从包含因果解释信息的因果关系增强模型中学习了高区分度的特征, 并通过注意力机制更好地发挥作用, 最终, 本文提出的在 GAN 框架下的带注意力机制的双向门控循环单元神经网络超越了传统方法和已有文献中的最好方法, 获得了更优的抽取效果.

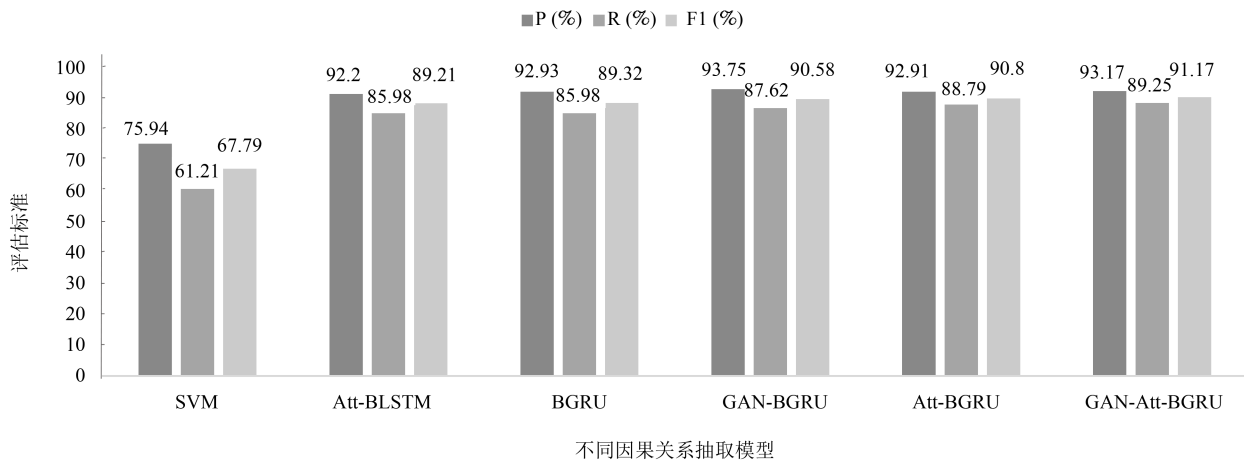


图4 不同模型的对比实验

Fig.4 Comparative experiment of different models

## 4 结论

本文提出一种 GAN 对抗框架下的因果关系抽取方法, 将带注意力机制的 BGRU 与对抗学习相融合, 在因果关系增强模型中引入因果关系解释语句. 基本模型通过模仿学习因果关系增强模型生成的特征向量表示, 得到高区分度的特征用于因果关系分类, 从而提高因果关系抽取效果. 同时, 加入的注意力机制在此基础上进一步优化了抽取结果. 通过在 SemEval-2010-Task8 扩展之后的数据集上进行对比实验, 说明 GAN 框架下带注意力机制的双向门控循环单元神经网络在因果抽取任务上的有效性和优越性.

因果关系抽取是一项基本任务, 除了提高因果关系抽取结果外, 还有很多有意义的工作值得深入研究. 同时, 因果关系抽取及其应用研究仍旧存在诸多挑战, 需要更多的关注和探索.

## References

- Radinsky K, Davidovich S, Markovitch S. Learning causality for news events prediction. In: Proceedings of the 21st International Conference on World Wide Web. Lyon, France: ACM, 2012. 909–918
- Hashimoto C, Torisawa K, Kloetzer J, Sano M, Varga I, Oh J H, et al. Toward future scenario generation: extracting event causality exploiting semantic relation, context, and association features. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, MD, USA: ACL, 2014. 987–997
- Girju R. Automatic detection of causal relations for question answering. In: Proceedings of the 2003 ACL Workshop on Multilingual Summarization and Question Answering. Sapporo, Japan: ACL, 2003. 76–83
- Oh J H, Torisawa K, Hashimoto C, Sano M, De Saeger S, Ohtake K. Why-question answering using intra-and inter-sentential causal relations. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria: ACL, 2013. 1733–1743
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 28th Annual Conference on Neural Information Processing Systems. Montreal, QC, Canada: NIPS, 2014. 2672–2680
- Abe S, Inui K, Matsumoto Y. Two-phased event relation acquisition: coupling the relation-oriented and argument-oriented approaches. In: Proceedings of the 22nd International Conference on Computational Linguistics. Manchester, United Kingdom: ACM, 2008. 1–8
- Do Q X, Chan Y S, Roth D. Minimally supervised event causality identification. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, United Kingdom: ACM, 2011. 294–303
- Hashimoto C, Torisawa K, De Saeger S, Oh J H, Kazama J. Excitatory or inhibitory: a new semantic orientation extracts contradiction and causality from the web. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island, Korea: ACM, 2012. 619–630
- Rink B, Harabagiu S. UTD: classifying semantic relations by combining lexical and semantic resources. In: Proceedings of the 5th International Workshop on Semantic Evaluation. Los Angeles, California, USA: ACM, 2010. 256–259
- Zeng D J, Liu K, Lai S W, Zhou G Y, Zhao J. Relation classification via convolutional deep neural network. In: Proceedings of the 25th International Conference on Computational Linguistics. Dublin, Ireland: ACL, 2014. 2335–2344
- Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S. Recurrent neural network based language model. In: Proceedings of the 11th Annual Conference of the International Speech Communication Association. Makuhari, Chiba, Japan: DBLP, 2010. 1045–1048
- Zhang D X, Wang D. Relation classification via recurrent neural network. arXiv: 1508.01006, 2015.

- 13 Zhang S, Zheng D Q, Hu X C, Yang M. Bidirectional long short-term memory networks for relation classification. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. Shanghai, China: PACLIC, 2015. 73–78
- 14 Zhou P, Shi W, Tian J, Qi Z Y, Li B C, Hao H W, et al. Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: ACL, 2016. 207–212
- 15 Chung J Y, Gulcehre C, Cho K H, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv: 1412.3555, 2014.
- 16 Zhang Y Z, Gan Z, Carin L. Generating text via adversarial training. In: Proceedings of the 2016 Workshop on Adversarial Training. Barcelona, Spain: NIPS, 2016.
- 17 Yu L T, Zhang W N, Wang J, Yu Y. SeqGAN: sequence generative adversarial nets with policy gradient. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, California, USA: AAAI, 2017. 2852–2858
- 18 Li J W, Monroe W, Shi T L, Jean S, Ritter A, Jurafsky D. Adversarial learning for neural dialogue generation. arXiv: 1701.06547, 2017.
- 19 Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. arXiv: 1701.07875, 2017.
- 20 Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved training of Wasserstein GANs. In: Proceedings of the 31st Annual Conference on Neural Information Processing Systems. Long Beach, CA, USA: NIPS, 2017. 5769–5779



冯 冲 北京理工大学计算机学院副研究员。2005 年获中国科学技术大学计算机科学系博士学位。主要研究方向为自然语言处理, 信息抽取, 机器翻译。本文通信作者。

E-mail: fengchong@bit.edu.cn

(FENG Chong Associate professor at the College of Computer Science and

Technology, Beijing Institute of Technology. He received

his Ph.D. degree from the Department of Computer Science, University of Science and Technology of China in 2005. His research interest covers natural language processing, information extraction, and machine translation. Corresponding author of this paper.)



康丽琪 北京理工大学计算机学院硕士研究生。主要研究方向为信息抽取, 关系抽取。E-mail: klq207@126.com

(KANG Li-Qi Master student at the College of Computer Science and Technology, Beijing Institute of Technology. Her research interest covers information extraction, relation extraction.)



石 戈 北京理工大学计算机学院博士研究生。主要研究方向为信息抽取, 事件抽取。E-mail: shige@bit.edu.cn

(SHI Ge Ph.D. candidate at the College of Computer Science and Technology, Beijing Institute of Technology. His research interest covers information extraction, event extraction.)



黄河燕 北京理工大学计算机学院教授。1989 年获中国科学院计算技术研究所计算机科学与技术博士学位。主要研究方向为自然语言处理和机器翻译社交网络与信息检索, 智能处理系统。

E-mail: hhy63@bit.edu.cn

(HUANG He-Yan Professor at the College of Computer Science and Technology, Beijing Institute of Technology. She received her Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences. Her research interest covers natural language processing, machine translation, social network, information retrieval, and intelligent processing system.)