

用于表格事实检测的图神经网络模型^{*}

邓哲也, 张 铭

(北京大学 信息科学技术学院 计算机科学技术系, 北京 100871)

通讯作者: 张铭, E-mail: mzhang@net.pku.edu.cn



摘要: 在自然语言理解和语义表征的研究中, 往往需要验证一句文本陈述是否基于给定的事实证据, 这就是事实检测任务. 现有的研究主要局限于处理文本事实验证, 而结构化证据下的验证还有待探索, 比如基于表格等形式的事实验证. TabFact 作为最新的基于表格的事实验证数据集, 基线方法并没有很好地利用表格的结构性特征. 结合表格的结构特征, 设计了以行、单元格为单位的基于图神经网络的事实验证模型 Row-GVM 和 Cell-GVM, 比基线模型的准确率分别提高了 2.62% 和 2.77%. 实验结果表明, 这两种利用了表格特征的方法确实是有效的.

关键词: 表格文本; 事实检测; 图神经网络

中图分类号: TP181

中文引用格式: 邓哲也, 张铭. 用于表格事实检测的图神经网络模型. 软件学报, 2021, 32(3): 753-762. <http://www.jos.org.cn/1000-9825/6184.htm>

英文引用格式: Deng ZY, Zhang M. Graph neural networks for table-based fact verification. Ruan Jian Xue Bao/Journal of Software, 2021, 32(3): 753-762 (in Chinese). <http://www.jos.org.cn/1000-9825/6184.htm>

Graph Neural Networks for Table-based Fact Verification

DENG Zhe-Ye, ZHANG Ming

(Department of Computer Science and Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

Abstract: In the study of natural language understanding and semantic representation, the fact verification task is very important to verify whether a textual statement is based on given factual evidence. Existing research is mainly limited to dealing with textual fact verification, while verification under structured evidence has yet to be explored, such as fact verification based on forms. TabFact is the latest table-based fact verification data set, but the baseline methods do not make good use of the structural characteristics of the table. This study takes advantage of the structural characteristics of the table and designs two models, Row-GVM (Row-level GNN-based verification model) and Cell-GVM (cell-level GNN-based verification model). They have achieved performances of 2.62% and 2.77% higher than the baseline model respectively. The results prove that these two methods using table features are indeed effective.

Key words: table-based; fact verification; graph neural network

随着如今文本数据越来越多, 并且可以轻松地通过互联网传播, 当人们面对错综复杂的信息时, 为了辨别信息的真假, 需要对这些信息进行验证, 这就体现出了事实检测任务的重要性.

在自然语言理解和语义表征的研究中, 验证一句文本陈述是否基于给定的事实证据非常重要, 这是自然语言理解中的一个基础任务. 现有的研究主要局限于处理非结构化证据的事实验证, 如自然语言的句子、文档、

* 基金项目: 国家重点研发计划(2018AAA0101900, 2018AAA0101902); 国家自然科学基金(91646202, 61772039)

Foundation item: Key Research and Development Program of China (2018AAA0101900, 2018AAA0101902); National Natural Science Foundation of China (91646202, 61772039)

本文由“支撑人工智能的数据管理与分析技术”专刊特约编辑陈雷教授、王宏志教授、童咏昕教授、高宏教授推荐.

收稿时间: 2020-08-23; 修改时间: 2020-09-03; 采用时间: 2020-11-06; jos 在线出版时间: 2021-01-21

新闻等,在这一类研究中,细分下来有很多不同的任务,如检查文本关联性^[1]、自然语言推理^[2]、基于维基百科的事实验证^[3]。它们用到的证据都是纯文本信息,而结构化证据下的验证还有待探索,比如基于表格、图表、数据库等形式的事实验证。在这一类研究中,有基于图片证据的数据集 NLVR^[4]/NLVR2^[5],还有基于表格证据的数据集 TabFact^[6]。

TabFact 数据集是一个包含 1.6 万个维基百科中的表格作为事实证据和 11.8 万条人工标注的自然语言陈述的事实验证数据集。不仅形式上与纯文本数据不同,采用的是表格形式的数据,而且它与一般的事实验证数据集不同的点在于,它要求模型同时具有语义推理和符号推理的能力。图 1 就是该数据集的一个实例。

Home team	Home team score	Away team	Away team score	Venue	Crowd	Date
Geelong	13.6 (84)	North Melbourne	4.3 (27)	Kardinia Park	21,555	28-Jun-80
Footscray	10.13 (73)	Melbourne	6.7 (43)	Western Oval	11,785	28-Jun-80
Hawthorn	21.17 (143)	Fitzroy	8.7 (55)	Princes Park	8,209	28-Jun-80
Richmond	9.19 (73)	Collingwood	7.22 (64)	MCG	64,054	28-Jun-80
South Melbourne	5.9 (39)	Carlton	10.18 (78)	Lake Oval	12,362	28-Jun-80
St Kilda	7.11 (53)	Essendon	6.12 (48)	VFL Park	20,485	28-Jun-80

Table: 1980 VFL Season

Claim: Western Oval venue records a lower crowd participation that of the Kardinia park venue.

Label: ENTAILED

Fig.1 An instance of the TabFact dataset

图 1 TabFact 数据集的一个实例

TabFact^[6]中提出了两个基线模型,分别是基于自然语言推理的 Table-BERT 和基于搜索和数据库查询的 LatentProgram 算法。这两个做法都没有取得非常好的效果,因为 Table-BERT 简单地把表格中的内容拼接在一起组成了文本,转化为了基于文本的事实验证问题,忽略了表格中行与行、列与列之间的关系;而 LatentProgram 算法是基于人工规则的搜索,泛化效果并不好。

本文将围绕 TabFact 数据集,针对表格数据的特殊的结构特征,设计基于表格数据的事实验证模型和算法,完成代码实现,并在基于表格的事实检测数据集上验证了效果。

本文的贡献可以概括为以下两点:

- (1) 提出了以表格的行为单位的 Row-GVM 模型,利用表格的行特征和图注意力网络对该任务建模,准确率相比基线模型提升了 2.62%;
- (2) 提出了以表格的单元格为单位的 Cell-GVM 模型,利用表格的单元格特征和图卷积网络对该任务建模,准确率相比基线模型提升了 2.77%。

1 相关工作

1.1 事实检测任务数据集

FEVER(fact extraction and verification,事实抽取和检测)^[3]数据集是一个公开的事实检测数据集,它包含 185 445 条陈述,陈述是通过改编维基百科中的句子生成的。FEVER 数据集关注的是验证文本陈述的真伪,依据是维基百科中词条里出现的文本信息。这与本文关注的基于表格的事实验证任务非常相似。在本文的任务中,陈述也是文本的形式,但是依据的形式是维基百科中词条里出现的表格。

TabFact^[6]数据集是一种全新的基于表格的事实验证数据集,难点在于需要模型能够完成语义推理和符号推理。还有一些其他的基于表格的数据集,如基于表格的问答数据集 WikiTableQuestion^[7],但本质上和 TabFact 数据集有很大的区别,因为 WikiTableQuestion 数据集中的问题对答案的类型有强烈的指向性,对于模型的推理能力要求不高,因此,在这个数据集上的做法很难迁移到 TabFact 数据集上。

1.2 图神经网络

图神经网络^[8]可以比一般的神经网络更好地挖掘出用图表示的数据.在图神经网络中,节点可以通过与它相连的边共享邻居节点的信息,学到更好的表示.最近的一些研究在利用不同的神经网络结构来优化整张图中节点信息的交互.

图神经网络在基于文本的事实检测任务数据集 FEVER 上有不错的表现.GEAR^[9]使用了图注意力网络(GAT)^[10]来对陈述-依据对进行推理和整合,它把陈述和依据的文本拼在一起当作节点,构建了一张完全图,利用 GAT 在图上传递节点信息.KGAT^[11]也采用了 GAT,它把依据的文本当作节点,引入了称为边核和点核的高斯核池化层,其中,边核用来在依据图上传递信息,点核用来把所有节点的信息进行合并得到整个图的表示.这种方法得到的依据和陈述的注意力相对而言更为集中,使得在 FEVER 数据集上的表现要好过 GEAR.不同于直接把依据文本传进预训练语言模型的其他模型,DREAM^[12]关注到了文本中的语义层面的信息,比如人物、地点、时间等不同实体之间的关系,认为这类语义信息对理解依据中的结构关系非常重要,于是对陈述和依据建出了语义图,在语义图上,利用图卷积网络(GCN)^[13]进行推理,在 FEVER 数据集上取得了目前最好的结果.

图神经网络在阅读理解任务上也有很好的表现,比如,Entity-GCN^[14]在建图过程中,把文本中出现的所有实体都当作了图中的节点.根据节点之间的关系连接了不同种类的边,如:不同文档中描述的同一体之间连边,同一文档中的所有点互相连边等.之后,运用 GCN 对文本内容进行多轮推理,取得了目前最好的效果.

1.3 自然语言推理

对语言的推理建模,是自然语言理解的重要的第 1 步.自然语言推理的目标是,判断一个自然语言假设可否由一个自然语言前提推理得到.具体地说,是判断假设和前提的关系,可以是支持、反驳或中立中的任何一种.

随着深度学习的兴起,涌现出了很多解决这类问题的模型,如 SNLI^[2],Decomposed Model^[15],Enhanced-LSTM^[16],Multi-NLI^[17],BERT^[18].本文的任务虽然也是自然语言推理,但是本文的前提不是完全以文本形式呈现的,而是由文本组成的半结构化的表格形式呈现的,所以该任务可以看作是半结构化领域的自然语言推理问题.

1.4 预训练语言模型

如今在很多自然语言处理的任上,预训练的语言表示模型,如 ELMo^[19],OpenAI GPT^[20],都被证明是非常高效的.BERT^[18]是一种预训练语言表示的新方法,通过联合调节所有层中的双向 Transformer 来预训练深度双向表示.在本文的工作中,使用 BERT 来对文本信息进行编码.

2 方法

2.1 任务定义

在介绍本文的方法之前,首先给这个基于表格的事实检测任务进行一个正式的定义.

本文把数据集中的每个实例表示成 (T, C, L) ,其中:表格 $T = \{T_{ij} | 0 \leq i \leq R_T, 1 \leq j \leq C_T\}$, R_T 和 C_T 分别为表格的行列数, T_{ij} 表示第 i 行第 j 列的单元格的内容;陈述 $C = \{c_1, c_2, \dots, c_n\}$ 是一个 n 个单词组成的基于表格内容的陈述;标签 $L \in \{0, 1\}$, $L=1$ 表示陈述 C 被表格 T 支持, $L=0$ 表示陈述 C 被表格 T 反驳.本文的目标是给定 (T, C) ,预测正确的标签 L .

2.2 方法

本文结合表格的结构特征,提出了两种用来对基于表格信息的陈述进行事实检测的模型,分别是:

- (1) Row-GVM:以表格中的每一行为单位的(Row-level)基于图神经网络的事实检测模型(GNN-based verification model);
- (2) Cell-GVM:以表格中的每一个单元格为单位的(Cell-level)基于图神经网络的事实检测模型.

2.2.1 Row-GVM

如图 2 所示,Row-GVM 模型分为 3 个部分.

- (1) 信息编码:对行信息进行编码,把每一行当作图中的一个节点;
- (2) 信息传递:建立一个全连接图,在图上利用 GAT 的原理进行信息传递;
- (3) 整合预测:整合每个节点的表示得到一个最终表示,进行二分类预测。

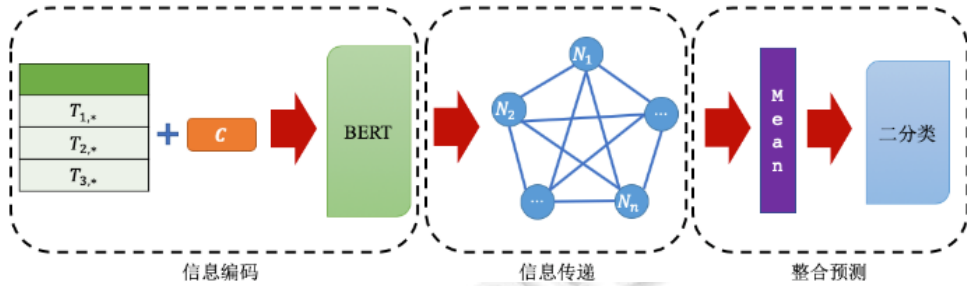


Fig.2 A flowchart of Row-GVM

图2 Row-GVM 的流程图

(1) 编码

为了得到第 i 行的表示,本文把第 i 行的信息 $T_{i,1}, T_{i,2}, \dots, T_{i,c_r}$ 拼接起来,得到 $T_{i,*}$ 。为了在后续的信息传递中更好地捕捉和陈述相关的信息,本文将第 i 行的信息 $T_{i,*}$ 和陈述 C 拼接在一起后,通过 BERT 获得第 i 个节点的表示 $N_i = \text{BERT}(T_{i,*}, C)$ 。 $N_i \in \mathbb{R}^F$, 其中, F 是每个节点的特征维数。

(2) 信息传递

在信息传递模块,本文引入了 GAT,它是由若干个图注意力层堆叠而成的。

第 1 层的输入是一系列节点的表示 $\mathbf{h}^0 = \{\mathbf{h}_1^0, \mathbf{h}_2^0, \dots, \mathbf{h}_n^0\}$, $\mathbf{h}_i^0 \in \mathbb{R}^F$ 。每一个节点的最初表示由每一行的信息拼接上陈述后,通过 BERT 后得到,即 $\mathbf{h}_i^0 = N_i$ 。

在第 t 层图注意力层中,输入是 \mathbf{h}^{t-1} , 输出是 \mathbf{h}^t 。为了更好地捕捉节点之间的信息,本文先用一个矩阵 $\mathbf{W}^t \in \mathbb{R}^{F' \times \mathbb{R}^F}$ 把 \mathbf{h}^{t-1} 映射到高维空间,再对每个节点计算自注意力,用映射 $\beta \in \mathbb{R}^{F' \times \mathbb{R}^{F'}} \rightarrow \mathbb{R}$ 来计算注意力系数 $e_{ij} = \beta(\mathbf{W}^t \mathbf{h}_i^{t-1}, \mathbf{W}^t \mathbf{h}_j^{t-1})$ 。这表示第 j 个点对第 i 个点的重要程度。本文希望每个节点的信息都可以传达到图中其他的每一个点,所以对每一对节点都要计算注意力系数。对于点 i ,本文要把其他点对它的重要程度做一个正则化,这里使用 softmax 函数:

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k=1}^{c_r} \exp(e_{ik})} \quad (1)$$

在本文的实验中,自注意力模块 β 中,和第 i 个节点有关的重要程度是用一个权值向量 $\beta_i \in \mathbb{R}^{2 \times F'}$ 来实现的,并且引入了 LeakyReLU(负数输入的斜率为 0.1)^[21], α_{ij} 可以重写为

$$\alpha_{ij} = \frac{\exp(\beta_i^T (\text{LeakyReLU}(\mathbf{W}^t \mathbf{h}_i^{t-1} \parallel \mathbf{W}^t \mathbf{h}_j^{t-1})))}{\sum_{k=1}^{c_r} \exp(\beta_i^T (\text{LeakyReLU}(\mathbf{W}^t \mathbf{h}_i^{t-1} \parallel \mathbf{W}^t \mathbf{h}_k^{t-1})))}$$

其中, T 表示转置, \parallel 表示向量的连接操作。

最后,根据计算出的注意力系数把其他点的表示整合起来,得到每个节点的新的表示,作为第 t 层的输出:

$$\mathbf{h}_i^t = \sum_{j=1}^{c_r} \alpha_{ij} \mathbf{h}_j^{t-1} \quad (2)$$

(3) 预测

为了进行最后的预测,本文需要得到一个整张图的表示。这里,本文采用求每一维特征的平均值的方法得到最终表示 $\mathbf{h}^G, \mathbf{h}^G \in \mathbb{R}^F$ 。得到了图的表示 \mathbf{h}^G 后,用一个一层的全连接网络来获得最终的预测结果:

$$p_{\theta} = \text{softmax}(\text{ReLU}(\mathbf{W}^G \mathbf{h}^G + \mathbf{b}^G)),$$

其中, $\mathbf{W}^G \in \mathbb{R}^{2 \times F}$, $\mathbf{b}^G \in \mathbb{R}^{2 \times 1}$. 本文通过优化参数 θ 来最小化交叉熵损失 $\mathcal{L}(p_\theta, L)$, 其中, L 是真实的标签.

2.2.2 Cell-GVM

如图 3 所示, Cell-GVM 模型分为 3 个部分.

- (1) 信息编码: 对单元格的信息进行编码, 把每个单元格当作图中的一个节点;
- (2) 信息传递: 同一行之间的点两两之间互相连边(图 3 中的实线边), 记作 ROW 边; 同一列之间的点两两之间互相连边(图 3 中的点线边), 记作 COL 边. 在图上利用 GCN 的原理进行信息传递;
- (3) 整合预测: 整合每个节点的表示得到一个最终表示, 进行二分类预测.

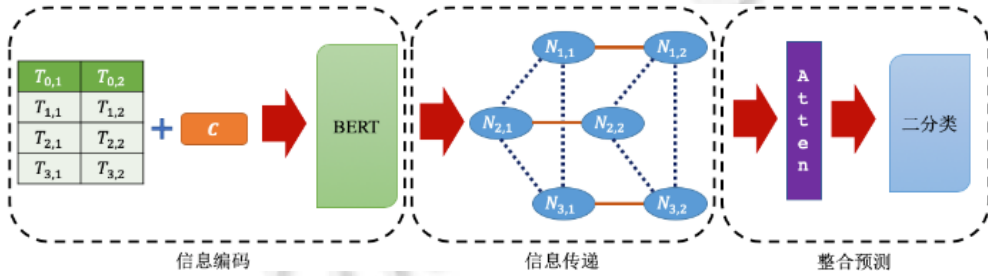


Fig.3 A flowchart of Cell-GVM

图 3 Cell-GVM 的流程图

(1) 编码

为了在后续的信息传递中更好地捕捉和陈述相关的信息, 将第 i 行第 j 列的单元格的信息 $T_{i,j}$ 和陈述 C 拼接在一起后, 通过 BERT 获得第 i 行第 j 列的节点的表示 $N_{i,j} = \text{BERT}(T_{i,j}, C)$. 这里把陈述 C 也通过 BERT, 得到它的表示 $N_C = \text{BERT}(C)$. $N_{i,j}, N_C \in \mathbb{R}^F$, 其中, F 是每个节点的特征维数.

(2) 信息传递

在信息传递模块, 与 Row-GVM 不同的是, 本文引入了 GCN.

第 1 层的输入是一系列节点的表示 $\mathbf{h}^0 = \{\mathbf{h}_{1,1}^0, \mathbf{h}_{1,2}^0, \dots, \mathbf{h}_{R,C}^0\}$, $\mathbf{h}_{i,j}^0 \in \mathbb{R}^F$. 每一个节点的最初表示由每一行的信息拼接上陈述后, 通过 BERT 后得到, 即 $\mathbf{h}_{i,j}^0 = N_{i,j}$.

本文把 RC 个节点和 $RC(C-1)/2$ 条 ROW 边组成的子图记作 G_{ROW} , RC 个节点和 $RC(R-1)/2$ 条 COL 边组成的子图记作 G_{COL} .

记图 $G_r (r \in \{ROW, COL\})$ 给每个点加上自环后的邻接矩阵为 A_r , 度数矩阵为 $D_r \left(D_{r(i,i)} = \sum_j A_{r(i,j)} \right)$, 归一化后的对称邻接矩阵为 $\tilde{A}_r = D_r^{-\frac{1}{2}} A_r D_r^{-\frac{1}{2}}$.

第 t 层图卷积层会通过两类边整合每个点的邻居的信息, 得到第 t 层的输出:

$$\mathbf{h}_{i,j}^t = \frac{1}{2} \sum_{r \in \{ROW, COL\}} \sigma(\tilde{A}_r \mathbf{h}_{i,j}^{t-1} \mathbf{W}_r^t) \quad (3)$$

其中, r 表示的是边的种类, \mathbf{W}_r^t 是权值矩阵, σ 是激活函数.

(3) 预测

为了得到图的表示, 先用矩阵 \mathbf{W}_H 和 \mathbf{W}_C 分别把 \mathbf{h}^{t-1} 和 N_C 映射到高维空间, $\mathbf{W}_H, \mathbf{W}_C \in \mathbb{R}^{F' \times F}$, 然后用映射 $\gamma: \mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$ 来计算节点和陈述之间的注意力 $p_{i,j} = \gamma(\mathbf{W}_H \mathbf{h}_{i,j}^t, \mathbf{W}_C N_C)$, 然后进行归一化, 得到:

$$\beta_{i,j} = \text{softmax}_j(p_{i,j}) = \frac{\exp(p_{i,j})}{\sum_{1 \leq x \leq R, 1 \leq y \leq C} \exp(p_{x,y})} \quad (4)$$

在本文的实验中, 映射 γ 采用的是点积运算, 所以 $\beta_{i,j}$ 可以重写为

$$\beta_{i,j} = \frac{\exp((W_H h_{i,j}^L)^T W_C N_C)}{\sum_{1 \leq x \leq R, 1 \leq y \leq C} \exp((W_H h_{x,y}^L)^T W_C N_C)} \tag{5}$$

根据注意力系数加权相加,可以得到图的表示 $h_G = \sum_{i=1}^R \sum_{j=1}^C \beta_{i,j} h_{i,j}^L$, 这里的求和是向量按位求和。

得到了图的表示 h^G 后,把它和陈述的表示 N_C 拼接在一起,通过一个一层的全连接网络来获得最终的预测结果:

$$p_{\theta} = \text{softmax}(\text{ReLU}(W^G(h^G || N_C) + b^G)) \tag{6}$$

其中, $W^G \in \mathbb{R}^{2 \times F}$, $b^G \in \mathbb{R}^{2 \times 1}$, $||$ 表示向量的连接操作。

本文通过优化参数 θ 来最小化交叉熵损失 $\mathcal{L}(p_{\theta}, L)$, 其中, L 是真实的标签。

3 实验

3.1 实验设置

3.1.1 数据集

本文的实验全部在 TabFact 数据集^[6]的子集 TabFact-small 上进行,其中每一条陈述都是针对某一个表格的,并且被人工打上了“支持”或者“反驳”的标签。

由于计算资源的限制,本文从原数据集中移除了最大的一些表格,这部分表格占到原数据集的 8.12% 左右。本文对所有模型做的实验都是在 TabFact-small 数据集上进行。TabFact-small 数据集的信息见表 1。

Table 1 Statistics for the TabFact-small dataset

表 1 TabFact-small 数据集的统计数据

数据集划分	陈述的数量	表格的数量
训练集	84 363	13 175
验证集	11 579	1 696
测试集	11 499	1 695

3.1.2 基线模型

本次实验中的基线模型一共有 3 个。其中,第 1 个模型 Bert-Concat 是 TabFact 数据集^[6]中提出的表现最好的模型;后面两个模型 Row-Mean 和 Cell-Atten 是受第 1 个模型的启发,并结合本文提出的模型,实现的两个基线模型。

- Bert-Concat^[6]

如图 4 右上角图所示,把每一个单元格的信息拼起来,然后和陈述 C 拼接在一起,通过 Bert 和一个多层感知器 f_{MLP} 对结果进行分类,得到一个预测 $p_{\theta}(T, C) = \sigma(f_{MLP}(BERT(T, C)))$, 其中, σ 是 sigmoid 函数。

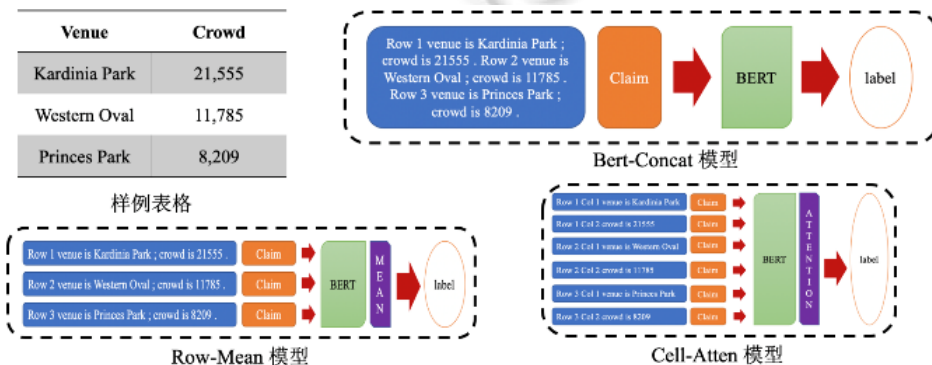


Fig.4 Flowcharts of baseline models

图 4 基线模型的流程图

- Row-Mean

如图 4 左下角图所示,把每行的所有单元格的信息拼起来,一共得到 R 个句子.每一行的句子都和陈述 C 拼接在一起,通过 Bert 后先对所有表示取平均值,然后用一个多层感知器 f_{MLP} 对结果进行分类,得到一个预测:

$$p_{\theta}(T,C)=\sigma(f_{MLP}(\text{Mean}(\text{BERT}(T_i,C)))) \quad (7)$$

其中, σ 是 sigmoid 函数.

- Cell-Atten

如图 4 右下角图所示,对每一个单元格都得到一个句子,一共有 RC 个句子.每一个单元格的句子都和陈述 C 拼接在一起,通过 Bert 后对所有表示计算和它和陈述 C 之间的注意力,然后用一个多层感知器 f_{MLP} 对结果进行分类,得到一个预测:

$$p_{\theta}(T,C)=\sigma(f_{MLP}(\text{Atten}(\text{BERT}(T_{ij},C)))) \quad (8)$$

其中, σ 是 sigmoid 函数.

3.1.3 评价标准

本文通过比较模型的预测输出和数据集中人工标注的真实标签计算准确率来评价模型,即:

$$Acc = \frac{1}{K'} \sum_{k=1}^{K'} \mathbb{I}(\hat{L}_k = L_k),$$

其中, K' 是测试集的大小, \hat{L}_k 是模型预测的标签, L_k 是真实的标签.

3.2 实现细节

在所有的模型中,本文使用的都是用 104 种语言预训练的 12 层、隐藏层维数 768、12 头的 BERT 模型^[18].所有的参数都是默认参数.优化器是 BERTAdam^[18],学习率是 $1e-5$.对于 Row-GVM,最大序列长度设为 128,批大小设为 8;对于 Cell-GVM,最大序列长度设为 64,批大小设为 2.热身比例是 0.4.

4 结果与分析

4.1 实验结果

在表 2 中列出了每个模型在 TabFact-small 数据集的验证集和测试集上的表现.在这个表中有如下发现.

- (1) Row-Mean 的表现比 Bert-Concat 稍好,但是 Cell-Atten 的表现没有超过 Bert-Concat.这可能是因为独立地检查单个单元格很难得出正确的判断,需要一行或所有单元格联合起来才能提供更有效的信息;
- (2) Row-GVM 和 Cell-GVM 都显著地超过了原文献的基线模型 Bert-Concat,它们在验证集上的提升分别是 2.47%和 3.08%,在测试集上的提升分别是 2.62%和 2.77%.这说明利用表格的行与列的相关性特征可以显著提高模型的表现;
- (3) Cell-GVM 在该数据集上取得了最好的效果.

Table 2 Main results

表 2 主要实验结果

模型	验证集准确率	测试集准确率
Bert-Concat	0.665 0	0.663 4
Row-Mean	0.673 4	0.669 2
Cell-Atten	0.642 3	0.638 7
Row-GVM	0.6814(+2.47%)	0.6808(+2.62%)
Cell-GVM	0.6855(+3.08%)	0.6818(+2.77%)

4.2 消融实验与分析

4.2.1 Row-GVM 中 GAT 模块的有效性

为了验证 Row-GVM 中 GAT 的有效性,本文调整了 GAT 的层数,分别设置成了 0 层~3 层,每种情况的实验结果都列在表 3 中.其中,0 层的 GAT 等价于从 Row-GVM 中删除了 GAT 模块,退化为 Row-Mean 模型.

Table 3 Comparison of different layers of GAT
with respect to their effectiveness

表 3 不同的 GAT 层数对模型效果的影响

Row-GVM 中 GAT 的层数	验证集准确率
0	0.673 4
1	0.678 6
2	0.679 7
3	0.681 4

本文发现:有 GAT 模块的 Row-GVM 的表现始终比删去 GAT 模块的 Row-Mean 要好,表现最好的由 3 层 GAT 组成的 Row-GVM 在结果上提升了 0.80%,证实了 GAT 模块的有效性,同时也说明了 Row-GVM 有能力处理需要多轮推理的事实检测问题。

4.2.2 Cell-GVM 中 GCN 模块的有效性

为了验证 Cell-GVM 中的有效性,本文调整了 GCN 的层数,分别设置成了 0 层~3 层,每种情况的实验结果都列在表 4 中.其中,0 层的 GCN 等价于从 Cell-GCN 中删除了 GCN 模块,退化为 Cell-Atten 模型。

Table 4 Comparison of different layers of GCN
with respect to their effectiveness

表 4 不同的 GCN 层数对模型效果的影响

Cell-GVM 中 GCN 的层数	验证集准确率
0	0.642 3
1	0.676 9
2	0.681 1
3	0.685 5

值得注意的是,Cell-Atten 的表现要低于基线模型 Bert-Concat.本文分析:这是因为本文对所有单元格的内容都求出了经过 Bert 的向量表示,然后直接对所有表示进行了简单融合,因此不能捕捉单元格与单元格之间的联系.相比之下,把所有单元格拼成一起经过 Bert 得到的表示还是可以捕捉到一部分单元格之间的联系,所以 Cell-Atten 的表现会低于 Bert-Concat.

本文发现:有 GCN 模块的 Cell-GVM 的表现始终比删去 GCN 模块的 Cell-Atten 要好,表现最好的由 3 层 GCN 组成的 Cell-GVM 在结果上提升了 4.32%,证实了 GCN 模块的有效性,同时也说明了 Cell-GVM 有能力处理需要多轮推理的事实检测问题。

4.2.3 Cell-GVM 中区分边的种类的有效性

为了验证 Cell-GVM 中区分 ROW 边和 COL 边的有效性,本文设置了一组对比实验.在这组实验中,不区分 ROW 边和 COL 边,建图时把这两类边标记为同一种边,信息传递时,节点的信息不再通过两种边传递,而是只通过这一种边传递.在该模型的 GCN 的第 l 层中,直接整合第 i 行第 j 列的单元格节点的邻居的信息:

$$h'_{i,j} = \sigma(\tilde{A}h_{i,j}^{-1}W^l).$$

记该模型为 Cell-GVM-SameEdge.

Cell-GVM-SameEdge 和 Cell-GVM 的实验结果列在表 5 中.可以看到:在不区分边的情况下,Cell-GVM 的表现下降了 0.42%,证实了区分边的种类对提升实验效果是有效的。

Table 5 Comparison of different types of edges
with respect to their effectiveness

表 5 边的种类对模型效果的影响

模型	验证集准确率
Cell-GVM-SameEdge	0.682 6
Cell-GVM	0.685 5

本文分析认为:3 层 GCN 是最有效的,因为 3 层的 GCN 可以把一个单元格(如 $h_{i,j}$)的信息通过 ROW 边传

到同一行的另一个单元格($h_{i,k}$),再通过 COL 边传到同一列的另一个单元格($h_{l,k}$),最后通过 ROW 边传到同一行,且和起点单元格的列数一样的单元格($h_{i,j}$).这样就可以在 3 步之内整合另一列的信息传递到这一列中其他的节点上.

5 结 论

本文针对基于表格的事实验证数据集 TabFact,利用了表格的结构特征,结合图注意力网络和图卷积神经网络,分别设计了以表格的行为单位的 Row-GVM 和以表格的单元格为单位的 Cell-GVM,并且都在 TabFact-small 数据集上取得了最好的结果,比基线模型分别提高了 2.62%和 2.77%.这两个模型通过实验被证明都是高效的,这说明利用表格的行与列的相关性特征确实可以提高模型的表现.

从目前的结果可以看到,预测的准确率还有很大的提升空间.分析了错误的案例后,发现本文的模型在捕捉和陈述相关的行与单元格方面的表现有较好的提升,但是符号推理方面表现稍有不足.在未来的研究中,将设计更加侧重于符号推理的模块,融入到本文的模型中,相信会有更好的表现.

References:

- [1] Dagan I, Glickman O, Magnini B. The PASCAL recognising textual entailment challenge. In: Proc. of the Machine Learning Challenges Workshop. Berlin, Heidelberg: Springer-Verlag, 2005. 177–190.
- [2] Bowman S, Angeli G, Potts C, *et al.* A large annotated corpus for learning natural language inference. In: Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing. 2015. 632–642.
- [3] Thorne J, Vlachos A, Christodoulopoulos C, *et al.* FEVER: A large-scale dataset for fact extraction and verification. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1: Long Papers. 2018. 809–819.
- [4] Suhr A, Lewis M, Yeh J, *et al.* A corpus of natural language for visual reasoning. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol.2: Short Papers. 2017. 217–223.
- [5] Suhr A, Zhou S, Zhang A, *et al.* A corpus for reasoning about natural language grounded in photographs. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. 6418–6428.
- [6] Chen W, Wang H, Chen J, *et al.* TabFact: A large-scale dataset for table-based fact verification. arXiv preprint arXiv:1909.02164, 2019.
- [7] Pasupat P, Liang P. Inferring logical forms from denotations. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics, Vol.1: Long Papers. 2016. 23–32.
- [8] Scarselli F, Gori M, Tsoi AC, *et al.* The graph neural network model. IEEE Trans. on Neural Networks, 2008,20(1):61–80.
- [9] Zhou J, Han X, Yang C, *et al.* GEAR: Graph-based evidence aggregating and reasoning for fact verification. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. 892–901.
- [10] Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. arXiv preprint arXiv:1710.10903, 2017.
- [11] Liu Z, Xiong C, Sun M. Fine-Grained fact verification with kernel graph attention network. arXiv preprint arXiv:1910.09796, 2019.
- [12] Zhong W, Xu J, Tang D, *et al.* Reasoning over semantic-level graph for fact checking. arXiv preprint arXiv:1909.03745, 2019.
- [13] Kipf TN, Welling M. Semi-Supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.
- [14] De Cao N, Aziz W, Titov I. Question answering by reasoning across documents with graph convolutional networks. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1: Long and Short Papers. 2019. 2306–2317.
- [15] Parikh A, Täckström O, Das D, *et al.* A decomposable attention model for natural language inference. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. 2016. 2249–2255.
- [16] Chen Q, Zhu X, Ling ZH, *et al.* Enhanced LSTM for natural language inference. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol.1: Long Papers. 2017. 1657–1668.
- [17] Williams A, Nangia N, Bowman SR. A broad-coverage challenge corpus for sentence understanding through inference. arXiv preprint arXiv:1704.05426, 2017.

- [18] Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1: Long and Short Papers. 2019. 4171–4186.
- [19] Peters M, Neumann M, Iyyer M, *et al.* Deep contextualized word representations. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1: Long Papers. 2018. 2227–2237.
- [20] Radford A, Narasimhan K, Salimans T, *et al.* Improving language understanding by generative pre-training. 2018. <https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/languageunderstandingpaper.pdf>
- [21] Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. Proc. of the ICML, 2013,30(1):3.



邓哲也(1997—),男,学士,主要研究领域为自然语言处理,数据挖掘.



张铭(1966—),女,博士,教授,博士生导师,CCF 高级会员,主要研究领域为信息检索,推荐系统,数据挖掘,机器学习.

www.jos.org.cn

www.jos.org.cn