

面向前提选择的新型图约简表示与图神经网络模型

兰咏琪, 何星星, 李莹芳, 李天瑞

引用本文

兰咏琪, 何星星, 李莹芳, 李天瑞. [面向前提选择的新型图约简表示与图神经网络模型](#)[J]. 计算机科学, 2024, 51(5): 193-199.

LAN Yongqi, HE Xingxing, LI Yingfang, LI Tianrui. [New Graph Reduction Representation and Graph Neural Network Model for Premise Selection](#) [J]. Computer Science, 2024, 51(5): 193-199.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于深度多视图网络的政务事件分拨方法](#)

Government Event Dispatch Approach Based on Deep Multi-view Network

计算机科学, 2024, 51(5): 216-222. <https://doi.org/10.11896/jsjcx.230300034>

[基于特征注意力提纯的显著性目标检测模型](#)

Salient Object Detection Based on Feature Attention Purification

计算机科学, 2024, 51(5): 125-133. <https://doi.org/10.11896/jsjcx.230300018>

[基于渐进式多尺度Transformer的图像去雾算法](#)

Multi Scale Progressive Transformer for Image Dehazing

计算机科学, 2024, 51(5): 117-124. <https://doi.org/10.11896/jsjcx.230300049>

[集成全尺度融合和循环注意力的医学图像分割网络](#)

Medical Image Segmentation Network Integrating Full-scale Feature Fusion and RNN with Attention

计算机科学, 2024, 51(5): 100-107. <https://doi.org/10.11896/jsjcx.230400114>

[基于跨模态信息过滤的视觉问答网络](#)

Cross-modal Information Filtering-based Networks for Visual Question Answering

计算机科学, 2024, 51(5): 85-91. <https://doi.org/10.11896/jsjcx.230300202>

面向前提选择的新型图约简表示与图神经网络模型

兰咏琪¹ 何星星¹ 李莹芳² 李天瑞³

1 西南交通大学数学学院 成都 611756

2 西南财经大学计算机与人工智能学院 成都 611130

3 西南交通大学计算机与人工智能学院 成都 611756

(302941597@qq.com)

摘要 自动定理证明器在证明问题时其搜索空间通常会呈现爆炸式增长,前提选择为该问题提供了新的解决思路。针对现有前提选择方法中逻辑公式图、图神经网络模型难以捕捉到公式图潜在信息的问题,提出了一种基于删除重复量词的简化逻辑公式图表示和具有注意力机制的项游走图神经网络模型,充分利用逻辑公式的语法和语义信息提高前提选择问题的分类精度。首先,将一阶逻辑猜想和前提公式转化为基于删除重复量词的简化一阶逻辑公式图;其次,利用消息传递图神经网络对节点和节点的项游走特征信息进行聚合和更新,随后使用注意力机制为图上的节点分配权重,进而调整图节点嵌入信息;最后,将前提图向量和猜想图向量拼接并输入二元分类器中实现前提分类。实验结果表明,所提方法在 MPTP 数据集和 CNF 数据集上的准确率分别达到了 88.61% 和 84.74%,超越现有最优的前提选择方法。

关键词 图神经网络;前提选择;注意力机制;一阶逻辑公式;图约简表示方法

中图分类号 TP391

New Graph Reduction Representation and Graph Neural Network Model for Premise Selection

LAN Yongqi¹, HE Xingxing¹, LI Yingfang² and LI Tianrui³

1 School of Mathematics, Southwest Jiaotong University, Chengdu 611756, China

2 School of Computer and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu 611130, China

3 School of Computer and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

Abstract The search space of an automatic theorem prover usually grow explosively when it proves problems. Premise selection provides a new solution to this problem. Aiming at the problem that the logical formula graph and graph neural network models in the existing premise selection methods are difficult to capture the potential information of the formula graph, this paper proposes a simplified logical formula graph representation based on removing repeated quantifiers and a term-walk graph neural network with attention mechanism, which makes full use of the syntactic and semantic information of logical formulas to improve the classification accuracy of premise selection problems. Firstly, the conjecture formulas and premise formulas are transformed into simplified first-order logic formula graphs based on removing repeated quantifiers. Secondly, the message passing graph neural network is used to aggregate and update nodes and their term walk feature information, and then the attention mechanism is used to assign weights to nodes on the graph, so as to adjust the graph nodes embedding information. Finally, the premise vector and the conjecture vector are concatenated and input into the binary classifier to realize the classification. Experimental results show that the accuracy of the proposed method on MPTP dataset and CNF dataset reaches 88.61% and 84.74%, respectively, which surpasses the existing premise selection method.

Keywords Graph neural network, Premise selection, Attention mechanism, First order logical formula, Graph reduction representation

1 引言

自动定理证明是人工智能领域一个核心且前沿的方向,

作为人工智能系统的重要组成部分,它被广泛应用于专家系统^[1]、电路设计^[2-3]、编译器^[4]和软件验证^[5]等领域。自动定理证明(Automated Theorem Provers, ATPs)首先将猜想与

到稿日期:2023-03-24 返修日期:2023-08-10

基金项目:中央高校基本科研业务费专项资金(2682020ZT107);国家自然科学基金(62106206);教育部人文社科项目(19YJCZH048, 20XJCZH016);四川省科技计划(2023YFH0066)

This work was supported by the Fundamental Research Funds for the Central Universities of Ministry of Education of China(2682020ZT107), National Natural Science Foundation of China(62106206), Humanities and Social Sciences Project of Ministry of Education(19YJCZH048, 20XJCZH016) and Science and Technology Program of Sichuan Province(2023YFH0066).

通信作者:李莹芳(liyf@swufe.edu.cn)

前提形式化为逻辑公式,然后将逻辑公式输入自动定理证明器中,从而实现从前提自动演绎出猜想。

ATPs 通过不断迭代搜索问题库中所有待处理的子句集来实现对新问题的证明,它在较小规模的问题库中可以高效且自然地证明问题。然而,当未被证明的问题处于具有成千上万个前提的大规模问题库(如 Mizar 数学库^[6])中时,ATPs 仍需要搜索并处理大规模问题库中的包括大量与问题结论不相关的所有子句,这显然超出了计算机的计算成本和运行能力。因此,为给定的新猜想选择更相关的前提十分关键。前提选择为解决该问题提供了一个新的途径,即在将逻辑公式输入证明器之前确定并选出有助于证明给定问题的结论的公式。

近年来,由于有效的前提选择方法能够大大提升 ATP 性能,学者们提出了许多前提选择方法。早期的前提选择方法主要是一种基于符号比较分析的手工设计的启发式方法^[7],它通过比较分析从输入公式中抽取子句的深度、符号计数等符号和结构特征来筛选前提集合中与结论更相关的前提。通常,启发式算法依靠手工设计的特征,其选取特征的局限性往往会限制前提选择的能力。随着计算机能力的提升,一些机器学习^[8-9]的方法成为前提选择的一种有效的替代方法,它们将问题自然地转化为分类或排序问题如长短期记忆神经网络(Long-Short Term Memory, LSTM)^[10]、门控循环神经网络(Gated Recurrent Neural Network, GRU)^[11]等。

在此基础上,由于逻辑公式可以自然地表示为图,而融合图拓扑结构信息的特征能够充分体现逻辑公式的特征,因此图神经网络^[12]与自动定理证明的结合成为了当前较为热门的研究主题。Wang 等^[13]首次提出带有边序信息的图神经网络模型并将其应用于前提选择问题,该模型仅仅选用了简单的图卷积神经网络(Graph Convolutional Neural Network, GCN)进行节点嵌入。Schlichtkrull 等^[14]引入节点关系权重矩阵,扩展了现有的 GCN,进而提出能够保持逻辑公式节点关系信息的图神经网络模型。Crouse 等^[15]从逻辑公式的有向图表示出发,利用有向图具有的可达性分别使用消息传递神经网络(Message Passing Neural Network, MPNN)和 LSTM 嵌入节点信息。Paliwal 等^[16]在 Wang 等工作的基础上探究不同形式的高阶逻辑公式图表示,包括其子表达式共享、叶子共享、变量重命名和它们之间的不同组合。

本文研究发现,现有的基于图神经网络的前提选择方法虽然能够在一定程度上提高前提选择分类的能力,但仍存在一些不足:1)多数前提选择方法忽略了逻辑公式不同图表示法对图神经网络模型的影响,这将导致图神经网络不能很好捕捉逻辑公式的内部和外部信息;2)现有的图神经网络模型常通过聚合来自邻居节点或其他节点的信息,来生成保留更多逻辑公式信息的特征,这些特征往往包含大量的节点信息,这可能导致生成的公式特征被图上不重要的信息影响,从而无法充分表示逻辑公式信息的图特征。

为了解决上述问题,本文提出了一种删除重复量词的简化一阶逻辑公式图表示法和一种具有注意力机制的项游走图神经网络模型,并将其应用于前提选择问题。首先,通过判断并删除连续重复的量词得到简化的一阶逻辑公式图;其次,基于新的逻辑公式图,提出了一种具有注意力机制的项游走图

神经网络模型(Attention-Term-Walk-based Graph Neural Network, Attention-TW-GNN),该模型按照项游走模式聚合位于项游走三元组上部、中部和下部的节点信息,引入注意力机制计算节点的项游走特征权重,并将该权重与节点信息结合生成新的节点嵌入向量,再通过全局平均池化得到最终的公式图特征向量;最后,将候选前提和给定猜想的图特征向量输入二元分类器,进而实现对候选前提的分类。本文模型分别应用于 MPTP 和 CNF 数据集,其分类精度超越了现有最优的前提选择方法。

2 一种约简的一阶逻辑公式图表示法

一阶逻辑公式图通常按照其语法解析树的结构转换生成。下面给出一阶逻辑公式和图的相关定义。

定义 1^[17] 一阶逻辑中的项常被递归定义为:1)常量符号是项;2)变量符号是项;3)若 f 是 n 元函数符号, $t_1, t_2, \dots, t_n$ 是项,则 $f(t_1, t_2, \dots, t_n)$ 是项;4)所有的项都是有限次使用 1)–3)生成的符号串。

定义 2^[17] 一阶逻辑公式中的公式可被递归定义为:1)原子是公式;2)若 H, G 是公式,则 $(\sim H), (H \wedge G), (H \vee G), (H \rightarrow G), (H \leftrightarrow G)$ 是公式;3)若 G 是公式, x 是 G 中的自由变量,则 $(\forall x)G, (\exists x)G$ 是公式。

定义 3^[18] 给定一个有序二元组 $G = (V, E)$ 表示一个图,其中,非空集合 $V = \{v_1, v_2, \dots, v_n\}$ 包含 G 中所有的顶点,集合 $E = \{\langle v_i, v_j \rangle \mid v_i, v_j \in V\}$ 包含 G 中所有顶点组成的边。

常用的逻辑公式到图的表示法多为被拓展的有向无环图(Directed Acyclic Graphs, DAGs)^[13],其一般步骤为:1)将逻辑公式转化为类似程序语言的语法解析树;2)合并解析树上相同的子表达式和叶子节点;3)重命名逻辑公式中的变量。

为了减小图数据的规模以及让 DAGs 包含更多的逻辑性质,本章提出了基于删除重复量词的有向无环图(Simplified Directed Acyclic Graphs Based on removing repeated quantifiers, Simplified-DAGs)。该 Simplified-DAGs 的操作相当于在原有 DAGs 的基础上将满足相同且连续这一条件的量词进行合并。具体来说,给定一个变量符号集合 V 、一个谓词符号集合 P 、一个函数符号集合 F 、逻辑联结词集合 C 、量词集合 Q 以及一个一阶逻辑公式表达式 s (项、原子或公式),一个基于删除重复量词的一阶逻辑公式图 $G = (V, E)$ 可被递归定义:

- 1)若是项,则有:
 - (1)若是常量符或变量符,则 $V_s = s, E_s = \emptyset$ 。
 - (2)若 $s = f(s_1, s_2, \dots, s_n)$ 是函数符项,其中 f 是 n 元函数符, $s_1, s_2, \dots, s_n$ 为项。对于 $s_i (1 \leq i \leq n)$,则有:
 - ①若 s_i 是常量符项或变量符项,则 $V_{s_i} = s_i, E_{s_i} = \{\langle f, s_i \rangle\}$ 。
 - ②若 s_i 是函数符项,则按照递归方式生成 V_{s_i} 和 E_{s_i} ; 对于 s ,有 $V_s = \bigcup_{i=1}^n V_{s_i}, E_s = \bigcup_{i=1}^n E_{s_i}$ 。
 - (2)若 $s = P(s_1, s_2, \dots, s_n)$,其中 P 是 n 元谓词符, $s_1, s_2, \dots, s_n$ 为项。对于 $s_i (1 \leq i \leq n)$,则有:
 - (1)若 s_i 是常量符项或变量符项,则 $V_{s_i} = s_i, E_{s_i} = \{\langle P, s_i \rangle\}$ 。
 - (2)若 s_i 是函数符项,则按照递归方式生成 V_{s_i} 和 E_{s_i} ; 对于

s , 有 $V_s = \bigcup_{i=1}^n V_{s_i}$, $E_s = \bigcup_{i=1}^n E_{s_i}$ 。

3) 若 $s = \mathcal{O}_1 x_1 \cdots \mathcal{O}_m x_m \hat{s}$, 其中 $\mathcal{O}_i (1 \leq i \leq m)$ 是量词, $\hat{s}$ 是子表达式, 则有:

$$V_s = \bigcup_{i=1}^m \{\mathcal{O}_i, x_i\} \cup V_{\hat{s}}$$

$$E_s = \{\langle \mathcal{O}_1, x_2 \rangle, \langle \mathcal{O}_1, \mathcal{O}_2 \rangle, \langle \mathcal{O}_2, x_2 \rangle, \dots, \langle \mathcal{O}_m, H(\hat{s}) \rangle\} \cup E_{\hat{s}}$$

其中, $\langle \mathcal{O}_i, \mathcal{O}_j \rangle$ 中 $\mathcal{O}_i$ 和 $\mathcal{O}_j$ 为不同的谓词符号, $H(s_i)$ 是表达式 s_i 最外层的符号。

4) 若 s 是逻辑联结词, 则有:

(1) 若 s 为一元逻辑联结词 $\neg$, 不妨设 $s = \neg s_1$, s_1 是子表达式, 则 $V_s = \{\neg\} \cup V_{s_1}$, $E_s = \{\langle \neg, H(s_1) \rangle\} \cup E_{s_1}$, 其中 $H(s_1)$ 是表达式 s_1 最外层的符号。

(2) 若 s 为二元逻辑联结词 $\odot$, 不妨设 $s = s_1 \odot s_2$, s_1, s_2 是子表达式, $V_s = V_{s_1} \cup \{\odot\} \cup V_{s_2}$, $E_s = E_{s_1} \cup \{\langle \odot, H(s_1) \rangle, \langle \odot, H(s_2) \rangle\} \cup E_{s_2}$, 其中, $H(s_1)$ 和 $H(s_2)$ 分别是表达式 s_1 和 s_2 最外层的符号。

满足上述定义的一阶逻辑公式图也就是本章提出的 Simplified-DAGs, 该图在减少 DAGs 数据规模的同时, 使得 DAGs 能够保留一阶逻辑中量词的少部分性质, 即 $\forall x \forall y f(x, y) = \forall y \forall x f(x, y)$, $\exists x \exists y f(x, y) = \exists y \exists x f(x, y)$ 。图 1(b) 给出了公式 $\forall x \forall y \forall z (f(x, y) \rightarrow f(p(x, y), p(z, x)))$ 的简化公式图表示, 图 1(a) 给出了文献[13]中的一阶逻辑公式图表示法。根据上述过程, 一阶逻辑公式可以被转化为——对应的 Simplified-DAGs。

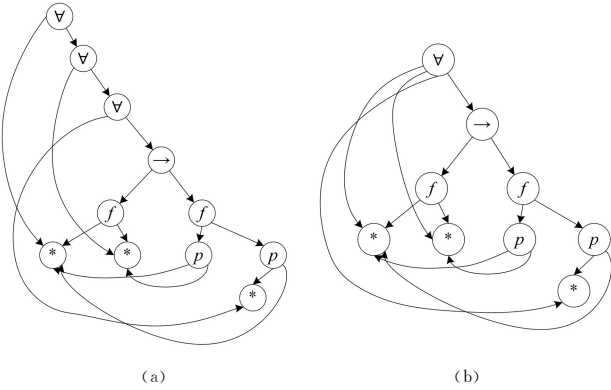


图 1 基于删除重复量词的简化一阶逻辑公式图

Fig. 1 Simplified first-order logic formula graph based on removing repeated quantifiers

3 模型设计

本章提出了一种基于新的逻辑公式图的具有注意力机制的项游走图神经网络模型 (Attention-TW-GNN), 并将其应用于前提选择问题中。一个完整的基于 Attention-TW-GNN 的前提选择模型的总体架构如图 2 所示, 主体部分包括以下 3 个分支: 一阶逻辑公式图表示、图神经网络和二元分类器。针对图表示模块, 模型接收一阶逻辑公式序列, 并构建基于删除重复量词的简化逻辑公式图表示 (详见第 2 章)。针对图神经网络模块, 受到项游走图神经网络^[19]的启发, 该模块采用项游走模式聚集简化逻辑公式图的节点信息, 在获得节点项

游走特征后, 本文的 Attention-TW-GNN 方法引入注意力机制平衡项游走特征信息和节点信息以生成最终一阶逻辑公式图嵌入向量 (详见 3.1 节)。针对二元分类器模块, 模型将前提图向量和猜想图向量拼接并输入一个具有分类能力的多层感知机 (Multi-Layer Perceptrons, MLPs) 中, 输出最终的预测分类得分 (前提有用得分和无用得分), 通过该得分确定前提的标签属性 (0 或 1), 再与已有标签属性进行比较, 从而对前提进行分类 (详见 3.2 节)。

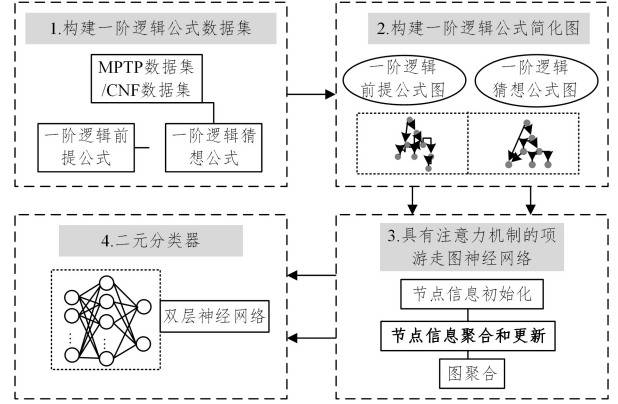


图 2 前提选择模型框架

Fig. 2 Framework of premise selection model

3.1 具有注意力机制的项游走图神经网络模型

3.1.1 项游走特征

Jakubov 等提出的项游走特征^[20]能够应用于自动定理证明的各项任务中, 而基于逻辑公式解析树的项游走特征也可以很轻易地被拓展为基于 (删除重复量词) DAGs 的项游走特征。设 $G=(V, E)$ 是有向无环图, G 中的项游走特征为图 G 上所有长度为 3 的有向路径, 则设 G 中的所有项游走特征都形如三元组 (u, v, w) , 其中, $u, v, w \in V$ 为图 G 上的节点, 节点 u 为节点 v 的父节点, 节点 w 为节点 v 的子节点。例如, 对一阶逻辑公式 $\forall x (f(x) \wedge p(x))$ 的项游走特征集合为 $\{(\forall, \wedge, f), (\forall, \wedge, p), (\wedge, f, x), (\wedge, p, x)\}$ 。

3.1.2 图节点向量初始化

对于输入图 $G=(V, E)$, 其中 V 为图上的所有节点。Attention-TW-GNN 首先为每个节点 $v \in V$ 分配一个初始嵌入 x_v , 再通过 k 轮迭代生成的节点状态向量 $h_v^{(k)}$ 实现消息传递过程, 其中 $k \in \{1, \dots, K\}$ 。

$$h_v^{(0)} = F_V(x_v) \quad (1)$$

其中, $h_v^{(0)} \in \mathbf{R}^{d_{h_v}}$ 是固定大小的初始状态向量; d_{h_v} 为节点嵌入向量的输出维度; F_V 为一个查找表, 存储着同一类型的数据元素; $x_v \in \mathbf{R}^{d_v}$; d_v 是 793, 它表示 793 个节点标记, 在这些标记中, 本文统一地将变量表示为 "Var"。

3.1.3 图节点信息聚合

Attention-TW-GNN 的节点信息聚合模块基于节点的项游走特征信息, 加入注意力机制平衡节点信息与项游走特征信息, 生成最终的节点嵌入信息。

节点的项游走特征信息聚合模块受项游走图神经网络的启发^[19], 按照项游走模式聚集节点信息。具体地, 对输入图 $G=(V, E)$ 的每个节点 v , 设 $T_u(v)$, $T_m(v)$ 和 $T_l(v)$ 分别表示节点 v 位于上部、中部和下部的项游走

特征集合, 其中:

$$T_u(v) = \{(v, u, w) \mid (v, u), (u, w) \in E\}$$

$$T_m(v) = \{(u, v, w) \mid (u, v), (v, w) \in E\}$$

$$T_l(v) = \{(u, w, v) \mid (u, w), (w, v) \in E\}$$

为了区分来自项游走特征不同位置的节点 v , 分别对 $T_u(v)$, $T_m(v)$ 和 $T_l(v)$ 三元组中的向量进行拼接, 然后模型针对 $T_u(v)$, $T_m(v)$ 和 $T_l(v)$ 以及聚合函数 F_u , F_m 和 F_l 聚集来自不同位置的节点 v 信息。

在第 k 次迭代中, 节点 v 来自集合 $T_u(v)$, $T_m(v)$ 和 $T_l(v)$ 的信息 $\mathbf{m}_{vu}^{(k)}$, $\mathbf{m}_{vm}^{(k)}$ 和 $\mathbf{m}_{vl}^{(k)}$ 为:

$$\mathbf{m}_{vu}^{(k)} = \sum \frac{F_u([\mathbf{h}_v^{(k-1)}; \mathbf{h}_u^{(k-1)}; \mathbf{h}_w^{(k-1)}])}{|T_u(v)|} \quad (2)$$

$$\mathbf{m}_{vm}^{(k)} = \sum \frac{F_m([\mathbf{h}_u^{(k-1)}; \mathbf{h}_v^{(k-1)}; \mathbf{h}_w^{(k-1)}])}{|T_m(v)|} \quad (3)$$

$$\mathbf{m}_{vl}^{(k)} = \sum \frac{F_l([\mathbf{h}_u^{(k-1)}; \mathbf{h}_w^{(k-1)}; \mathbf{h}_v^{(k-1)}])}{|T_l(v)|} \quad (4)$$

其中, $[\cdot]$ 表示不同节点状态向量的拼接; $|T_u(v)|$, $|T_m(v)|$ 和 $|T_l(v)|$ 分别表示集合 $T_u(v)$, $T_m(v)$ 和 $T_l(v)$ 中所有三元组个数的总和。

在节点信息聚合中节点本身的信息也是至关重要的, 项游走特征信息聚合了来自节点项游走特征集合的信息, 其中节点自身的信息只占了小部分, 这很有可能导致节点学习到过多其他节点的信息, 进而忽略了节点本身。基于此, 本文的 Attention-TW-GNN 方法引入注意力机制来平衡节点状态信息 $\mathbf{h}_v^{(k)}$ 与节点的项游走特征信息 $\mathbf{m}(k)_{vu}$, $\mathbf{m}_{vm}^{(k)}$ 和 $\mathbf{m}(k)_{vl}$ 。为降低模型结构的复杂度, 使用 $\mathbf{m}(k)_{vu}$, $\mathbf{m}_{vm}^{(k)}$ 和 $\mathbf{m}(k)_{vl}$ 作为项游走特征信息对节点信息的注意力分数 (贡献), 通过 softmax 函数归一化该注意力分数得到注意力权重 α_{vu} , α_{vm} 和 α_{vl} , 最终使用聚合函数 F_u , F_m 和 F_l 得到平衡后的节点信息 $\mathbf{s}_{vu}^{(k)}$, $\mathbf{s}_{vm}^{(k)}$ 和 $\mathbf{s}_{vl}^{(k)}$ 。

$$\alpha_{vu} = \text{softmax}(\mathbf{m}_{vu}^{(k)}) = \frac{\exp(\mathbf{m}_{vu}^{(k)})}{\sum_{vu \in V} \exp(\mathbf{m}_{vu}^{(k)})} \quad (5)$$

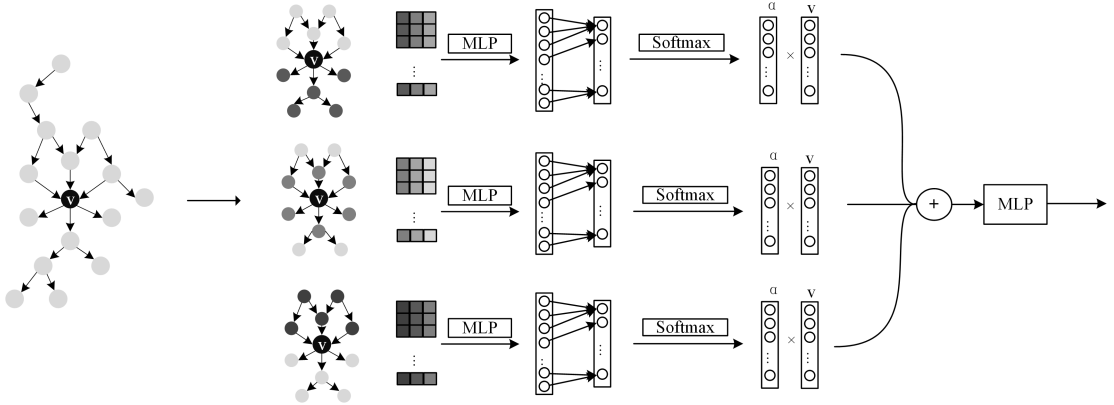


图 3 模型节点嵌入和传递过程

Fig. 3 Model node embedding and transfer process

3.1.5 图池化

在 k 次迭代结束后, Attention-TW-GNN 对逻辑公式图上的所有节点进行平均池化操作 (AvgPool)。最终的公式图嵌入向量 $\mathbf{h}_G$ 为:

$$\mathbf{h}_G = \text{AvgPool}\{\mathbf{h}_v^{(k)}\} = \frac{1}{V} \sum_{v_i \in V} \mathbf{h}_{v_i}^{(k)} \quad (14)$$

$$\mathbf{s}_{vu}^{(k)} = F_u(\alpha_{vu} \cdot \mathbf{h}_v^{(k-1)}) \quad (6)$$

$$\alpha_{vm} = \text{softmax}(\mathbf{m}_{vm}^{(k)}) = \frac{\exp(\mathbf{m}_{vm}^{(k)})}{\sum_{vm \in V} \exp(\mathbf{m}_{vm}^{(k)})} \quad (7)$$

$$\mathbf{s}_{vm}^{(k)} = F_m(\alpha_{vm} \cdot \mathbf{h}_v^{(k-1)}) \quad (8)$$

$$\alpha_{vl} = \text{softmax}(\mathbf{m}_{vl}^{(k)}) = \frac{\exp(\mathbf{m}_{vl}^{(k)})}{\sum_{vl \in V} \exp(\mathbf{m}_{vl}^{(k)})} \quad (9)$$

$$\mathbf{s}_{vl}^{(k)} = F_l(\alpha_{vl} \cdot \mathbf{h}_v^{(k-1)}) \quad (10)$$

其中, α_{vu} , α_{vm} 和 α_{vl} 分别表示节点项游走特征信息 $\mathbf{m}_{vu}^{(k)}$, $\mathbf{m}_{vm}^{(k)}$ 和 $\mathbf{m}_{vl}^{(k)}$ 对节点信息 $\mathbf{h}_v^{(k)}$ 的注意力权重, vu , vm 和 vl 分别表示位于项游走三元组上部、中部和下部的节点。

本节中, F_u , F_m , F_l , F_u , F_m 和 F_l 被设计为一个简单的多层 MLPs:

$$F_* = \text{ReLU}(\mathbf{W}_*^{(k)} + b_*^{(k)}) \quad (11)$$

其中, $*$ 为 u, m, l, U, M, L ; $\mathbf{W}_u^{(k)}, \mathbf{W}_m^{(k)}, \mathbf{W}_l^{(k)} \in \mathbf{R}^{d_{h_v} \times 3d_{h_v}}$ 和 $\mathbf{W}_U^{(k)}, \mathbf{W}_M^{(k)}, \mathbf{W}_L^{(k)} \in \mathbf{R}^{d_{h_v} \times d_{h_v}}$ 是可学习的参数矩阵, $b_*^{(k)} \in \mathbf{R}^{d_{h_v}}$ 是可学习的偏置; ReLU 是激活函数。

因此, 在第 k 次迭代中, 节点 v 聚集来自集合 $T_u(v)$, $T_m(v)$ 和 $T_l(v)$ 的平衡节点信息 $\mathbf{s}_{vu}^{(k)}$, $\mathbf{s}_{vm}^{(k)}$ 和 $\mathbf{s}_{vl}^{(k)}$ 可被汇总为 $\mathbf{s}_v^{(k)}$:

$$\mathbf{s}_v^{(k)} = \mathbf{s}_{vu}^{(k)} + \mathbf{s}_{vm}^{(k)} + \mathbf{s}_{vl}^{(k)} \quad (12)$$

3.1.4 图节点信息传递

Attention-TW-GNN 利用来自集合 $T_u(v)$, $T_m(v)$ 和 $T_l(v)$ 的调整后的节点总信息 $\mathbf{s}_v^{(k)}$ 和上一步的节点 v 状态向量 $\mathbf{h}_v^{(k-1)}$ 对节点向量 $\mathbf{h}_v^{(k)}$ 进行传递和更新。在第 k 次迭代中, 节点 v 的状态向量更新为:

$$\mathbf{h}_v^{(k)} = \mathbf{h}_v^{(k-1)} + F_{\text{sum}}(\mathbf{s}_v^{(k)}) \quad (13)$$

其中, F_{sum} 为节点信息传递函数。与 $F_{U, M, L}$ 相同, F_{sum} 被设计为一个简单的单层 MLPs。

图 3 给出了具有注意力机制的项游走图神经网络模型的节点信息嵌入过程。从中可以看出: Attention-TW-GNN 能够利用项游走特征信息调整节点信息, 而 TW-GNN 只利用了节点的项游走特征信息, 忽略了节点本身信息的重要性。

其中, $\mathbf{h}_G \in \mathbf{R}^{d_{h_v}}$, AvgPool 为平均池。

3.2 二元分类器

受到文献[19]的启发, 基于 Attention-TW-GNN 的前提选择模型的二元分类器同样使用一个简单的多层 MLPs 实现, 其分类过程为有监督二元分类。模型将候选前提和猜想的图嵌入向量对 $(\mathbf{h}_p, \mathbf{h}_c)$ 输入分类函数 F_{class} 中, 得到候选前提

在该猜想下的有用性得分。

$$\mathbf{z}=F_{\text{class}}([\mathbf{h}_p;\mathbf{h}_c])$$

(15)

其中, $\mathbf{z}\in\mathbf{R}^2$ 表示候选前提对猜想有用和无用的得分。

模型使用 softmax 函数归一化候选前提对猜想的有用和无用得分,并按照得分大小对候选前提进行划分:

$$\hat{\mathbf{y}}=\text{softmax}(\mathbf{z})$$

(16)

$$\hat{y}_i=\frac{\exp(z_i)}{\sum_j\exp(z_j)}$$

(17)

其中, $\hat{\mathbf{y}}$ 表示归一化后的前提有用得分和无用得分, z_i 为 $\mathbf{z}$ 中的第 i 个元素,候选前提有用得分和无用得分对应着不同的标签属性(1 或 0),则根据划分结果(得分最大值),确定候选前提的标签属性,并与已有标签进行比较,进而实现分类。本节中, F_{class} 被设计为一个简单的双层 MLPs。

$$F_{\text{class}}=\mathbf{W}_{c2}(\text{ReLU}(\mathbf{W}_{c1}(\cdot)+b_1))+b_2$$

(18)

其中, $\mathbf{W}_{c1}\in\mathbf{R}^{d_{h_v}\times2d_{h_v}}$, $\mathbf{W}_{c2}\in\mathbf{R}^{2\times d_{h_v}}$ 为可训练的参数矩阵; $b_1\in\mathbf{R}^{d_{h_v}}$, $b_2\in\mathbf{R}^2$ 为可训练的偏置。

3.3 损失函数

前提选择问题为分类问题,因此本文模型选用预测值 $\hat{\mathbf{y}}$ 和真实值 $\mathbf{y}$ 的交叉熵损失作为其损失函数 $L(\hat{\mathbf{y}},\mathbf{y})$:

$$L(\hat{\mathbf{y}},\mathbf{y})=\sum_i(y_i\ln(\hat{y}_i)+(1-y_i)\ln(1-\hat{y}_i))$$

(19)

其中, $y_i,\hat{y}_i$ 分别为真实值和预测值在第 i 类的值, y 为真实值的独热编码。

4 实验设置与结果分析

本文提出的基于 Attention-TW-GNN 的前提选择模型使用 Pytorch 框架^[21]实现,所有的实验在 GPU 为 NViDIA RTX 2080Ti、内存为 256 GB 和 Linux 版本为 Center OS 7.6 的超微 4029GP-TRT 服务器上进行。

4.1 数据集

本文基于 MPTP2078 题库^[22]建立了两个数据集,分别为原始(MPTP)数据集、合取范式(CNF)数据集,其中 MPTP 数据集为 MPTP2078 题库中的逻辑公式, CNF 数据集为 MPTP2078 题库中逻辑公式对应的合取范式。在命题逻辑中,由有限个简单析取式构成的合取式称为合取范式。一阶逻辑公式的合取范式与命题逻辑类似,首先通过消除蕴涵连接词、否定符号内移、变量标准化和 skolem 标准化等操作将逻辑公式转换为不包含存在量词和蕴涵符号的一阶逻辑公式;再运用命题逻辑结合律和分配律生成最终的一阶逻辑合取范式。本文的 CNF 数据集为 MPTP2078 题库中的一阶逻辑公式通过上述操作转化而来,转化后的一阶逻辑公式形如 $\forall x_1\forall x_2\cdots\forall x_n(Atom\vee Atom)\wedge (Atom\vee Atom)$,其中 $Atom$ 为原子公式。

本文在上述两个数据集上评估了基于 Attention-TW-GNN 的前提选择模型的实验结果,该题库包含 1469 个猜想和 24087 个用于证明这些猜想的前提。本文构建的用于训练、验证和测试(40996、13990 和 14068 个样本)的 MPTP 数据集与文献^[19]相同,其中每个数据集都形如三元组(前提、猜想、标签),前提是给定猜想的候选前提,标签是二元分类中

的 0 或 1(1 表示前提有用,0 表示前提无用)。同时,本文构建的 CNF 数据集分布与 MPTP 数据集相同,值得注意的是, CNF 数据集仅修改了 MPTP 数据集中一阶逻辑公式的表达形式。

4.2 实验设置

本文的基于 Attention-TW-GNN 的前提选择模型使用自适应矩估计 Adam 优化器^[24]的默认设置来训练模型,并使用 Pytorch 库中的 ReduceLROnPlateau 策略自动调整学习率。模型所有参数配置如表 1 所列。

表 1 模型参数设置

Table 1 Parameter setting

模型参数	设置
节点向量维数 d_{h_v}	128,256,512
迭代次数 K	1,2,3
训练轮数	100,150,200
学习率	0.01
学习率衰减系数	0.1
批量大小 Batchsize	16,32,64

为了评估基于 Attention-TW-GNN 的前提选择模型的性能,本文采用机器学习中常用的准确率(Accuracy)、召回率(Recall)、精确率(Precision)和 F1 来反映模型分类的正确性和准确性。指标越高,说明模型分类的效果越好。

4.3 实验结果和分析

4.3.1 基于 Attention-TW-GNN 的前提选择模型与一些主流方法的比较

本文在 MPTP 数据集和 CNF 数据集上评估基于 Attention-TW-GNN 的前提选择模型,并将该模型与一些主流方法进行比较。表 2 和表 3 列出了模型的分类能力。

从表 2 和表 3 可以看出:

1)基于主流图神经网络的前提选择模型结果较差。例如 GCN,GAT,SGC 等模型在 MPTP 数据集上的 Accuracy 指标分别为 86.25%,85.38%,85.67%。这说明主流的神经网络只能简单捕捉逻辑公式图的拓扑结构,无法捕捉逻辑公式的深层次信息。

2)在其他基线方法中,基于手工设计特征的图神经网络 PC-GCN 和 TW-GNN 明显优于主流的图神经网络。例如 PC-GCN 和 TW-GNN 模型在 CNF 数据集上的 F1 指标分别为 83.98%和 83.72%。一个合理的解释是,这些图神经网络模型除了聚合邻居节点信息外,还聚合了来自更远节点的信息。

3)本文的基于 Attention-TW-GNN 的前提选择模型的分类精度超越了现有的其他基于图神经网络的前提选择模型。例如:在 MPTP 数据集上,Attention-TW-GNN 相比主流图神经网络至少提高 2%,相比其他基线方法提高 0.5%;在 CNF 数据集上,Attention-TW-GNN 相比主流图神经网络模型提高 3%。这说明在加入注意力机制调整后的图神经网络能够更好地表征一阶逻辑公式的语法和语义信息,同时一阶逻辑公式的图表示也在一定程度上影响模型的分类效果。本文模型不仅能捕捉更远节点的信息,还考虑到模型的过拟合现象,通过简化节点嵌入信息来提高模型的拟合能力。

表 2 MPTP 数据集的模型对比实验结果

Table 2 Model experimental results comparison on MPTP dataset (%)

模型	Acc	Recall	Precision	F1
GCN ^[24]	86.25	85.47	86.69	86.08
GAT ^[25]	85.38	85.56	85.11	85.34
Graph-SAGE ^[26]	86.15	85.01	86.86	85.92
SGC ^[27]	85.67	85.12	85.92	85.52
Cheb ^[28]	86.14	85.21	86.68	85.94
PC-GCN ^[16]	87.77	88.64	87.01	87.82
TW-GNN ^[19]	88.12	86.95	88.15	88.04
Attention-TW-GNN	88.61	87.73	89.19	88.45

表 3 CNF 数据集的模型对比实验结果

Table 3 Model experimental results comparison on CNF dataset (%)

模型	Acc	Recall	Precision	F1
GCN	80.51	78.73	81.44	80.07
GAT	81.80	81.25	81.99	81.62
GraphSAGE	82.96	81.62	83.69	82.64
SGC	80.08	79.86	80.05	79.95
Cheb	81.32	79.08	82.60	80.80
PC-GCN	84.17	83.40	84.56	83.98
TW-GNN	84.24	81.42	86.15	83.72
Attention-TW-GNN	84.74	83.67	85.32	84.49

4.3.2 一阶逻辑公式不同图表示对模型的影响

一阶逻辑公式的不同图表示法也可能会对前提选择模型产生影响,表 4 和表 5 列出了 DAGs 和 Simplified-DAGs 在不同数据集中用于不同基于图神经网络前提选择模型的实验结果,从中可以看出:

1) Simplified-DAGs 能够在大多数基于图神经网络的前提选择任务上取得更好的结果。例如:在 CNF 数据集上,基于 Simplified-DAGs 的 GraphSAGE, SGC, Attention-TW-GNN 的 Accuracy 指标分别提高了 0.57%,0.85%和0.26%。这表明 Simplified-DAGs 能够在不损失 DAGs 的大部分信息的情况下减小图的规模,同时前提选择模型分类精度的提升也在一定程度上反映了 Simplified-DAGs 的有效性。

2)少部分基于 Simplified-DAGs 的图神经网络模型的效果较差,原因可能是这些模型已经进行一部分简化,导致不能很好地学习到逻辑公式图上的信息。

表 4 MPTP 数据集上逻辑公式图表示法对前提选择模型的影响

Table 4 Influence of logical formula graph representation on premise selection model on MPTP dataset (%)

模型	DAGs	Sim-DAGs	DAGs	Sim-DAGs
	Acc	Acc	F1	F1
GCN	86.25	86.57	86.08	86.46
GAT	85.38	86.04	85.34	85.91
GraphSAGE	86.15	86.11	85.92	85.96
SGC	85.57	86.11	85.52	85.96
Cheb	86.14	85.90	85.94	85.65
PC-GCN	87.77	88.19	87.82	88.04
TW-GNN	88.12	88.29	88.04	88.05
Attention-TW-GNN	88.27	88.61	88.03	88.45

表 5 CNF 数据集上逻辑公式图表示法对前提选择模型的影响

Table 5 Influence of logical formula graph representation on premise selection model on CNF dataset (%)

模型	DAGs	Sim-DAGs	DAGs	Sim-DAGs
	Acc	Acc	F1	F1
GCN	80.51	80.97	80.07	80.48
GAT	81.80	81.80	81.62	81.48
GraphSAGE	82.96	83.53	82.64	83.36
SGC	80.08	80.93	79.95	80.62
Cheb	81.32	80.03	80.80	79.70
PC-GCN	84.17	84.40	83.98	84.37
TW-GNN	84.24	84.64	83.72	84.54
Attention-TW-GNN	84.48	84.74	84.24	84.49

3)现有的一阶逻辑公式的图表示法仍然不能完全捕捉到逻辑公式潜在的语义信息。同时,DAGs 虽然能够完全表示整个逻辑公式,但逻辑公式中可能存在的无关的部分会导致最终输出公式图向量中存在错误的信息,这也可能是 Simplified-DAGs 能够提高前提选择分类准确性的原因。

4.3.3 消融实验

基于 Attention-TW-GNN 的前提选择模型包括两个模块:一阶逻辑公式图表示模块和图神经网络模块。本小节设计了 4 个基于 Attention-TW-GNN 的前提选择模型变体,其中 R-,A-,T-分别表示删除图表示简单化、注意力机制和项游走特征的前提选择模型。表 6 列出了前提选择模型的消融实验结果,从中可以看出:1) Attention_TW_GNN_{AR-} 的效果仅次于 Attention_TW_GNN_{R-} 和 Attention_TW_GNN_{A-},这表明逻辑公式图表示和注意力机制可影响前提选择任务的分类精度;2) Attention_TW_GNN_{T-} 的效果最差,这说明项游走特征在前提选择模型中是十分关键的;3) Attention_TW_GNN 模型的分类结果始终最优,表明了将图表示和注意力机制相融合的前提选择模型能够得到更好的分类精度,缺少任意模块都无法得到最佳结果。

表 6 消融实验

Table 6 Ablation experiment (%)

模型	MPTP 数据集		CNF 数据集	
	Acc	F1	Acc	F1
Att_TW_GNN _{R-}	88.27	88.03	84.48	84.24
Att_TW_GNN _{A-}	88.29	88.05	84.64	84.54
Att_TW_GNN _{AR-}	88.12	88.04	84.24	83.72
Att_TW_GNN _{T-}	86.04	85.91	81.80	81.62
Att_TW_GNN	88.61	88.45	84.74	84.49

结束语 本文提出了一个基于简化图表示和图神经网络的前提选择模型 Attention-TW-GNN,在图神经网络中加入具有注意力机制的项游走特征信息以调整节点的嵌入信息。针对一阶逻辑公式特性,提出了删除重复量词的简化一阶逻辑公式图表示法。针对该简化逻辑公式图表示法以及项游走特征,设计了一个具有注意力机制的项游走图神经网络。该模型能够综合调整项游走特征信息和节点嵌入信息,进而生成保留一阶逻辑公式更多特性的嵌入公式图向量。在两个数据集上的实验结果表明,本文模型在前提选择任务上能够取得更好的精度。针对一阶逻辑公式的特点,我们计划在下一

步的工作中提出更多保留一阶逻辑公式信息的图表示法以及更多表征一阶逻辑公式的图神经网络模型,同时探索针对不同一阶逻辑公式图表示的评估方法,进而解释其合理性。

参 考 文 献

- [1] LE SERGENT T. SCADe: A comprehensive framework for critical system and software engineering[C]// Integrating System and Software Modeling: 15th International SDL Forum Toulouse (SDL 2011). France, 2011. Revised Papers 15. Berlin Heidelberg: Springer, 2012: 2-3.
- [2] LARRABEE T. Test pattern generation using Boolean satisfiability[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1992, 11(1): 4-15.
- [3] EGGERSGLUSS S, DRECHSLER R. Efficient data structures and methodologies for SAT-based ATPG providing high fault coverage in industrial application [J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2011, 30(9): 1411-1415.
- [4] DARWIVHE A. Compiling knowledge into decomposable negation normal form[C]// IJCAI. 1999, 99: 284-289.
- [5] JACKSON D, SCHECHTER I, SHLYAHTER H. Alcoa: the alloy constraint analyzer[C]// Proceedings of the 22nd International Conference on Software Engineering. 2000: 730-733.
- [6] KALISZYK C, URBAN J. Mizar 40 for mizar 40[J]. Journal of Automated Reasoning, 2015, 55(3): 245-256.
- [7] MENG J, PAULSON L C. Lightweight relevance filtering for machine-generated resolution problems[J]. Journal of Applied Logic, 2009, 7(1): 41-57.
- [8] KALISZYK C, URBAN J. Learning-assisted automated reasoning with flyspeck [J]. Journal of Automated Reasoning, 2014, 53(2): 173-213.
- [9] IRVING G, SZEGEDY C, ALEMI A A, et al. Deepmath-deep sequence models for premise selection[J]. arXiv: 1606. 04442, 2016.
- [10] LOOS S, IRVING G, SZEGEDY C, et al. Deep network guided proof search[J]. arXiv: 1701. 06972, 2017.
- [11] CHVALOVSKY K, JAKUBUV J, SUDA M, et al. ENIGMA-NG: efficient neural and gradient-boosted inference guidance for E[C] // Automated Deduction-CADE 27: 27th International Conference on Automated Deduction. Natal, Brazil: Springer International Publishing, 2019: 197-215.
- [12] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.
- [13] WANG M, TANG Y, WANG J, et al. Premise selection for theorem proving by deep graph embedding[C]// Advances in Neural Information Processing Systems. 2017: 2786-2796.
- [14] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// The Semantic Web: 15th International Conference (ESWC 2018). Heraklion, Crete, Greece, Springer International Publishing, 2018: 593-607.
- [15] CROUSE M, ABDELAZIZ I, CORNELIO C, et al. Improving graph neural network representations of logical formulae with subgraph pooling[J]. arXiv: 1911. 06904, 2019.
- [16] PALIWAL A, LOOS S, RABE M, et al. Graph representations for higher-order logic and theorem proving[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 2967-2974.
- [17] LIU X H. Automatic Reasoning Based on Conclusion Method [M]. Science Press, 1994: 15-16.
- [18] BOLLOBAS B. Modern graph theory[M]. Springer Science & Business Media, 1998.
- [19] LIU Q H. Research on First-order Logic Premise Selection for Grand Theory [D]. Chengdu: Southwest Jiaotong University, 2021.
- [20] JAKUBUV J, URBAN J. ENIGMA: efficient learning-based inference guiding machine[C]// 10th International Conference Intelligent Computer Mathematics (CICM 2017). Edinburgh, UK, 2017, Springer International Publishing, 2017: 292-302.
- [21] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. arXiv: 1912. 01703, 2019.
- [22] RUDNICKI P. An overview of the Mizar project [C]// Proceedings of the 1992 Workshop on Types for Proofs and Programs. 1992: 311-330.
- [23] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv: 1412. 6980, 2014.
- [24] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv: 1609. 02907, 2016.
- [25] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv: 1710. 10903, 2017.
- [26] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[J]. arXiv: 1706. 02216, 2017.
- [27] WU F, SOUZA A, ZHANG T, et al. Simplifying graph convolutional networks [C] // International Conference on Machine Learning. PMLR, 2019: 6861-6871.
- [28] DEFFERRARD M, BRESSION X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016: 3844-3852.



LAN Yongqi, born in 1998, postgraduate. Her main research interests include graph neural networks for premise selection.



LI Yingfang, born in 1985, Ph.D, associate professor. Her main research interests include intelligent information processing, intelligent decision making and control, reasoning and problem solving and so on.