计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(12): 3013−3027, 2025

面向心理健康检测的心理支持异质图模型

刘德喜[1,2] 张子靖[1,2] 邹志峰[1,2] 万齐智[1,2] 刘喜平[1,2] 廖国琼[2] 朱廷劭[3]
[1](江西财经大学计算机与人工智能学院 南昌 330013)
[2](数据与知识工程江西省高校重点实验室(江西财经大学) 南昌 330013)
[3](中国科学院心理研究所 北京 100101)
(dexi.liu@163.com)

Heterogeneous Graph Model for Psychological Support Towards Mental Health Detection

Liu Dexi[1,2], Zhang Zijing[1,2], Zou Zhifeng[1,2], Wan Qizhi[1,2], Liu Xiping[1,2], Liao Guoqiong[2], and Zhu Tingshao[3]
[1](*School of Computer and Artificial Intelligence*, *Jiangxi University of Finance and Economics*, *Nanchang* 330013)
[2](*Jiangxi Key Laboratory of Data and Knowledge Engineering* (*Jiangxi University of Finance and Economics*), *Nanchang* 330013)
[3](*Institute of Psychology*, *Chinese Academy of Sciences*, *Beijing* 100101)

Abstract Online mental health forums have become an important carrier of mental health services. Detecting psychological distress from a vast number of posts is the basis for psychological intervention. Fully utilizing the social relationships of seekers is conducive to judging their mental health status. However, most existing methods rely on explicit social relationships. The existing models fail to pay attention to the psychological support relationships between patients and doctors (seekers and supporters). These relationships are based on the patient's personal experiences, symptom causes, self-cognition, and psychological support expertise. Thus, we take suicidal ideation as the detection target and propose post-user psychological support heterogeneous graph (PU-PSHG). PU-PSHG is used to represent the semantics of posts and the doctor-patient relationships between seekers and supporters in online mental health forums. According to PU-PSHG, graph-enhanced suicide ideation detection (GSID) model is proposed. Firstly, based on the definition of psychological support relationships, the semantics of two meta-paths (user-to-user and user-to-post) are defined, and PU-PSHG containing users and posts are constructed. DeepWalk algorithm is used to learn doctor-patient relationships or community relationships from PU-PSHG. Then, the representation of psychological support relationships is learned through relational features. Besides the post semantics and doctor-patient relationships are fused based on heterogeneous relationships. Finally, suicide ideation detection is performed based on the representation of posts. Experimental results on CLPsych2017 shared task show that GSID has better performance compared with existing methods. Compared with C-GraphSAGE, GSID are improved by 7.8%, 4.8%,1.4% in Non-green *F*1, All *F*1, All *Acc*, respectively. Ablation experiments find that removing three different types of relationships (the reply relationship between posts, the psychological support relationship between users and posts, the psychological support relationship between users and users) from PU-PSAG leads to decreases in Non-green *F*1 of 3.04%, 3.80%, 6.17% respectively.

收稿日期: 2024-04-03; **修回日期**: 2025-04-25
基金项目: 国家自然科学基金项目(62272206, 62272205); 江西省主要学科学术和技术带头人培养计划领军人才项目(20213BCJL22041); 江西省自然科学基金项目(20212ACB202002, 20242BAB25119)
This work was supported by the National Natural Science Foundation of China (62272206, 62272205), the Training Plan for Academic and Technical Leaders of Major Disciplines in Jiangxi Province (20213BCJL22041), and the Natural Science Foundation of Jiangxi Province (20212ACB202002, 20242BAB25119).

Key words online mental health forum; suicide ideation detection; psychological support heterogeneous graph; graph attention networks; GSID model

摘 要 在线心理健康论坛已经成为心理健康服务的重要载体,从海量帖子中检测出有心理健康问题的帖子是心理干预的基础.充分利用求助者的社交关系有利于判断其心理健康状态,然而,现有模型大都依赖显式的社交关系,没有关注医患(支持者和求助者)之间基于患者经历、症状成因、自我认知以及心理支持专长所形成的心理支持关系.以自杀意念为检测对象,提出帖子-用户交互心理支持异质图(post-user psychological support heterogeneous graph, PU-PSHG)来表示在线心理健康论坛中求助者和支持者发布的帖子语义和医患关系.基于PU-PSHG提出一种图增强的自杀意念检测(graph-enhanced suicide ideation detection, GSID)模型.首先,基于心理支持关系定义用户对用户、用户对帖子的2种元路径语义,构建包含用户和帖子的PU-PSHG,并采用DeepWalk算法从PU-PSHG中学习医患关系或社群关系.然后,通过关系表征学习心理支持关系的表示,基于异质关系融合帖子语义和医患关系.最后根据帖子的表示进行自杀意念强度分类.在CLPsych2017共享任务上的实验结果表明,GSID模型与现有的方法相比具有更好的性能.在Non-green $F1$, All $F1$, All Acc指标上相比于基准模型C-GraphSAGE提高7.8%, 4.8%, 1.4%.消融实验发现,去除PU-PSHG中帖子与帖子的回复关系、用户对帖子的心理支持关系、用户对用户的心理支持关系,Non-green $F1$分别下降3.04%, 3.80%, 6.17%.

关键词 在线心理健康论坛;自杀意念检测;心理支持异质图;图注意力网络;GSID模型

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202440247 **CSTR:** 32373.14.issn1000-1239.202440247

心理健康问题已成为全球范围内最普遍、最严重的议题之一[1].根据2022年世界卫生组织报道[2],全球大约1/8的人患有心理健康问题.及时发现有心理危机的患者,有助于尽早地为他们提供专业的心理辅导和治疗,从而避免出现极端行为.

随着互联网技术与社交媒体的兴起,在线心理健康论坛蓬勃发展,成为心理健康服务的重要辅助手段[3].用户可以实时、便捷、匿名地披露情绪,寻求心理支持.在线心理健康论坛中用户数量众多,而具有心理支持或心理咨询资质的用户相对有限.因此,自动发现具有心理困扰的帖子并根据需要支持的紧急程度对它们分类,有助于充分利用有限的心理服务资源,及时为最需要的患者提供服务.

目前,基于社交媒体的心理健康检测已取得了丰富的成果[4-11].在信息利用方面,检测模型使用的信息主要包括5类,分别是:1)求助者发布的某条帖子;2)求助者在一段时间内发布的历史帖子集合;3)其他相关用户发布的帖子;4)求助者在社交网络上的行为和属性,如发帖时间、互动等;5)求助者在社交网络上的社交关系.

早期,文献[4−6]主要利用求助者发布的某条帖子判断其心理健康状况.由于心理危机的形成通常需要一个过程,且用户对心理危机的表达常分散在不同的帖子中,求助者在一段时间内发布的所有帖子可以提供更丰富的信息.因此,文献[7−10]基于注意力机制编码所有帖子中蕴含的语义信息,但是它们忽略了帖子之间的关联语义,且未利用来自其他用户的回复.文献[11−19]利用用户在社交网络上的行为、属性和社交关系等信息,旨在从中捕获更多的语义.然而,在线心理健康论坛上的用户极少存在“好友”“关注”等显式社交关系[20],例如,他们没在相关社交网络注册,或者用户之间只是存在共同兴趣,不能较好地反映彼此之间心理健康层面的关联关系.

由于求助者和心理咨询专家(心理支持者)之间属于一种特殊的医患关系[21],不是简单的社交关系.在开放的心理健康论坛上,心理咨询专家通常也会选择自己较为擅长的领域或问题进行心理支持.为此,文献[15−18]通过用户互动构建用户之间的“隐式”关系.Mishra等人[15]通过回复、引用、提及分别构建用户之间的3类社交关系图.Sinha等人[16]和Sawhney等人[17-18]以用户和待评估的帖子为节点,以用户之间的交互(回复、引用、提及)关系以及用户与帖子之间的发帖关系为边,构建异质图.然而,这些方法存在3个局限:1)只考虑了用户之间的关联,将用户的表示直接拼接给帖子,没有将帖子作为节点并采用图机制进行帖子语义更新;2)未利用帖子之间以及

用户与帖子之间的心理支持关联语义；3)异质图只包含当前帖子，未包含历史帖子，不能较好地在异质图中捕获同一个用户所有帖子之间的关联语义，进而迭代影响其他有关联的用户节点和帖子节点的信息.

为了解决上述局限，本文首先构建了帖子-用户交互心理支持异质图(post-user psychological support heterogeneous graph, PU-PSHG)，捕获帖子与帖子、帖子与用户、用户与用户之间的心理支持关联语义. 然后，考虑到用户节点没有较为丰富的语义用于初始化表示(通常随机初始化，由模型调整)，本文基于构建的异质图，利用 DeepWalk 算法[22]和 Word2Vec[23]预训练框架学习用户关联中蕴含的医患关系，作为用户节点的初始化表示. 最后，基于构建的异质图，提出了图增强的自杀意念检测(graph-enhanced suicide ideation detection, GSID)模型.

本文的主要贡献包括：

1)构建了帖子-用户交互心理支持异质图. 相对基于社交关系构建的用户关系图，该图能够捕获帖子与帖子、帖子与用户、用户与用户之间的心理支持关联语义，丰富了帖子的语义表示，有利于用户自杀意念检测.

2)采用 DeepWalk 算法和 Word2Vec 预训练框架，基于用户之间的关联关系，学习用户节点的初始化表示. 这个过程捕获了异质图中用户节点之间的结构信息，使得具有相似个人背景的用户以及具有更高医患匹配度的用户之间的表示更接近，学习到的用户节点表示更有针对性.

3)提出了图增强的自杀意念检测模型. 其中，采用了基于 Transformer 的图学习策略，借助图机制更新帖子的表示，使得帖子的语义更加丰富.

4)构造了丰富的实验，使用多种评测方法对模型进行全面评测. 实验结果验证了模型的有效性，Non-green $F1$, All $F1$, All Acc 分别达 55.5%, 63.8%, 73.9%，平均高出 C-GraphSAGE 基线模型 4.7%，高出 HCN 基线模型 15.1%.

1 相关工作

基于社交媒体的心理健康评估旨在利用机器学习和自然语言处理等技术对社交媒体上各类数据进行分析和建模，从而评估或检测社交媒体上用户或帖子中的心理健康状况. 在早期研究中，大多采用传统特征工程和机器学习的方法. De Choudhury 等人[11]证明 Twitter 等社交媒体上的数据内容可用来检测个体的抑郁程度，提出了一系列用户在社交媒体上的行为和属性特征，包括用户参与、用户情绪、语言风格、自我中心社交图、发帖时间等，并利用这些行为和属性特征来构建 SVM 分类器. Malmasi 等人[12]考虑上下文的词汇特征、语法特征和其他元数据特征，采用基本分类器构建元分类器，用于预测心理健康论坛上帖子所表达的自杀风险程度. 刘德喜等人[13]提出多种特征融合策略，融合帖子作者的行为与属性特征、帖子语言与用词风格特征、帖子内容与上下文特征，对帖子所反映的心理危机程度进行自动评估. Shrestha 等人[14]将帖子的语言风格特征和基于网络的特征相结合，预测论坛用户的抑郁程度，使用了包括 PageRank、互易性、聚类系数、Node2Vec 网络节点嵌入等多种网络特征. 实验表明，网络特征是抑郁用户的强预测因子，可以提高预测的平均精度.

随着 Glove, ELMo, Word2Vec, BERT 等词嵌入算法的不断发展，以及 CNN、LSTM 等深度学习模型对文本建模能力的提升，学者开始将深度学习模型用于心理健康检测任务.

在 CNN 与 LSTM 模型方面，Yates 等人[4]使用 CNN，提出了一个能对 Reddit, ReachOut 等在线论坛帖子进行抑郁或自我伤害倾向检测的神经网络模型. 模型只依赖于帖子内容，避免了繁琐的特征工程. 此外，Sawhney 等人[5]的实验结果显示，C-LSTM 模型与其他模型相比具有一定的优势. Tadesse 等人[6]提出 LSTM-CNN 的组合模型则大幅度提高了检测的准确性.

随着注意力机制的提出，为增强模型对相关词汇的敏感性并捕捉帖子的情感，Cao 等人[7]分析了用户在树洞(社交网络上某个特定的位置)中发布的帖子，提出了一种基于树洞帖子内容的自杀词嵌入方法. 考虑到用户历史情绪和发帖时间的不规则性，Sawhney 等人[8]提出一个基于时间感知的 Transformer 模型，验证了情绪和时间线索在心理危机检测中的效用. 为了解不同类别自杀风险水平之间的关系，Sawhney 等人[9]将自杀风险评估重新表述为有序回归问题，提出了一种分层注意力模型 SISMO. 刘德喜等人[10]构建了多层局部信息融合模型 MPIF，考虑词语、短语、句子 3 个层面的局部信息，使用分层 BiLSTM 和注意力机制相结合的方式提取帖子中词语层面和句子层面的局部信息；同时采用 CNN 网络获取帖子中短语层面的局部信息，并将这 3 个层面的局部信息进行融合.

图神经网络可用于学习图中的各种关系，能够建模在线论坛上用户在交流和互动过程中所形成的社交网络. 近年来，基于图神经网络的心理健康检测方法受到广泛关注. Mishra 等人[15]提出了一种特征堆叠的 SNAP-BATNET 架构，手动提取了一组新颖的特征，包括帖子元数据特征、用户历史贴特征和社交图特征. 实验证明，这些特征有利于心理危机检测任务. Sinha 等人[16]提出了一种集成模型框架，分别训练代表自杀意念帖子不同方面的分类器，包括文本分类模型、用户历史帖子的时间序列模型以及用户社交关系的图神经网络模型，将这 3 个模型加权后的预测结果作为最后的预测结果. 基于用户社交网络关系的图神经网络模型利用了用户与帖子的发帖关系以及用户与用户的互动关系，提高了准确率.

构建图神经网络的节点和边时，所利用的信息不尽相同，例如：发帖与互动信息[17]、背景信息[18]、知识图谱[19]以及词典信息[24]等. 在发帖与互动信息以及背景信息方面，Sawhney 等人[17]提出的 Hyper-SOS 框架利用了发帖关系和互动关系. 该框架可以同时建模用户的历史情感以及来自社交网络中用户的邻居信息. Sawhney 等人[18]认为在线对话可作为理解帖子的重要背景，使用来自在线社区的会话树代替用户的历史帖子，提出了一种基于双曲对话网络的自杀意念检测方法.

在知识图谱与词典信息方面，Cao 等人[19]构建了一种面向自杀意念检测的个人知识图谱，并设计了 2 层注意力机制，用于判断知识图谱中不同属性的重要性以及用户社交图中好友关系的重要性. Lee 等人[24]构建了一组与自杀相关的词典，提出了一种基于图神经网络的自杀倾向检测模型，通过构建帖子与自杀词以及自杀词与自杀词这 2 类关系，让模型更加关注帖子中自杀词的含义.

网络特征和网络关系是心理健康检测任务中的强预测因子，基于帖子内容和网络特征结合的方法均优于仅仅使用帖子内容的方法. 然而，现有方法未充分挖掘和利用用户与用户以及用户与帖子之间的心理支持关系，导致模型无法将用户之间的医患关系或社群关系融入帖子的语义表示中. 具体来说，Cao 等人[19]通过用户之间的关注关系构建用户邻居图，可是在线心理健康论坛中用户极少存在显式的关注关系. Mishra 等人[15]考虑用户的互动关系，并编码用户在互动网络中的结构信息，没有对用户与帖子联合学习. Sinha 等人[16]和 Sawhney 等人[17]在用户互动关系的基础上，增加用户与帖子的发帖关系，但是没有区别异质图中不同节点的类型，也没有考虑帖子之间的心理支持关系. Lee 等人[24]采用的异质图以帖子和帖子中的自杀词作为节点，考虑了自杀词与自杀词以及自杀词与帖子之间的关系，但没有考虑帖子之间、用户之间、以及帖子与用户之间的关系. Sawhney 等人[18]通过帖子之间的回复关系构建会话树，同样没考虑帖子与用户以及用户与用户之间的关系.

2 帖子-用户交互心理支持异质图构建

在线论坛或社区上的数据集 $D=\langle P,U,R\rangle$. 其中，$P=\{p_1,p_2,\cdots,p_{|D|}\}$表示$|D|$个帖子集合，包含帖子 ID、发帖用户 ID、帖子内容；$U=\{u_1,u_2,\cdots,u_M\}$表示 M 个用户集合；$R=\{R^{\mathrm{UP}},R^{\mathrm{PP}}\}$为关系集合，包括用户与帖子的发帖关系 R^{UP} 以及帖子与帖子的回复关系 R^{PP}.

帖子-用户交互心理支持异质图 PU-PSHG 以用户和帖子为节点，节点之间的连边按照以下步骤构建.

步骤 1. 以用户与帖子之间的发帖关系在用户节点与帖子节点之间构建边，如：帖子1 $\overset{\text{发帖}}{\leftarrow}$ 用户1.

步骤 2. 以帖子之间的回复关系在帖子节点与帖子节点之间构建边，如：帖子2 $\overset{\text{回复}}{\rightarrow}$ 帖子1.

受元路径[25-26]的启发，本文提出 2 种反映心理支持关系的元路径，分别是用户对用户的心理支持关系以及用户对帖子的心理支持关系. 具体构建步骤为步骤 3 和步骤 4.

步骤 3. 通过式(1)所示元路径，构建用户 2 对用户 1 的心理支持关系.

$$\text{用户2} \overset{\text{发帖}}{\rightarrow} \text{帖子2} \overset{\text{回复}}{\rightarrow} \text{帖子1} \overset{\text{发帖}}{\leftarrow} \text{用户1}. \tag{1}$$

步骤 4. 通过式(2)所示元路径，构建用户 2 对帖子 1 的心理支持关系.

$$\text{用户2} \overset{\text{发帖}}{\rightarrow} \text{帖子2} \overset{\text{回复}}{\rightarrow} \text{帖子1}. \tag{2}$$

基于上述步骤，可得到针对帖子和用户的心理支持异质图 PU-PSHG. 其中，发帖关系与回复关系能够捕获节点之间的基础语义信息；元路径则表示 2 个节点的复合关系，不同元路径包含不同的语义信息，可以加强节点之间的语义联系. PU-PSHG 中，用户对用户的心理支持关系集合表示为 R^{UU}，用户对帖子的心理支持关系集合表示 R^{PU}.

为了更好地理解 PU-PSHG，图 1 展示了构建示例，其中，虚线实心箭头为用户对用户的心理支持关系(心理支持关系 1)，虚线空心箭头为用户对帖子的心理支持关系(心理支持关系 2).

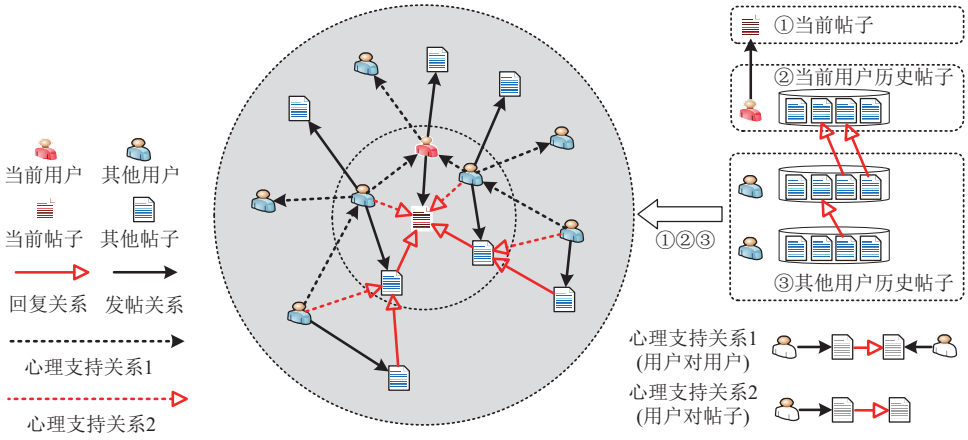


Fig. 1 Construction of PU-PSHG

图 1 PU-PSHG 的构建

3 GSID 模型

对于自杀意念检测任务，本文将其定义为 4 标签的分类任务，即对于每个帖子，预测帖子所反映的自杀意念强度标签.

基于 PU-PSHG 图，本文提出了一个图增强的自杀意念检测模型(GSID). 模型框架如图 2 所示.

该模型主要包括 4 个模块：1)帖子表示模块. 初始化帖子的向量表示得到. 2)用户表示模块. 从 PU-PSHG 中析取用户关联图，并初始化用户节点表示. 3)异质图注意力转换网络. 基于 PU-PSHG，更新学习

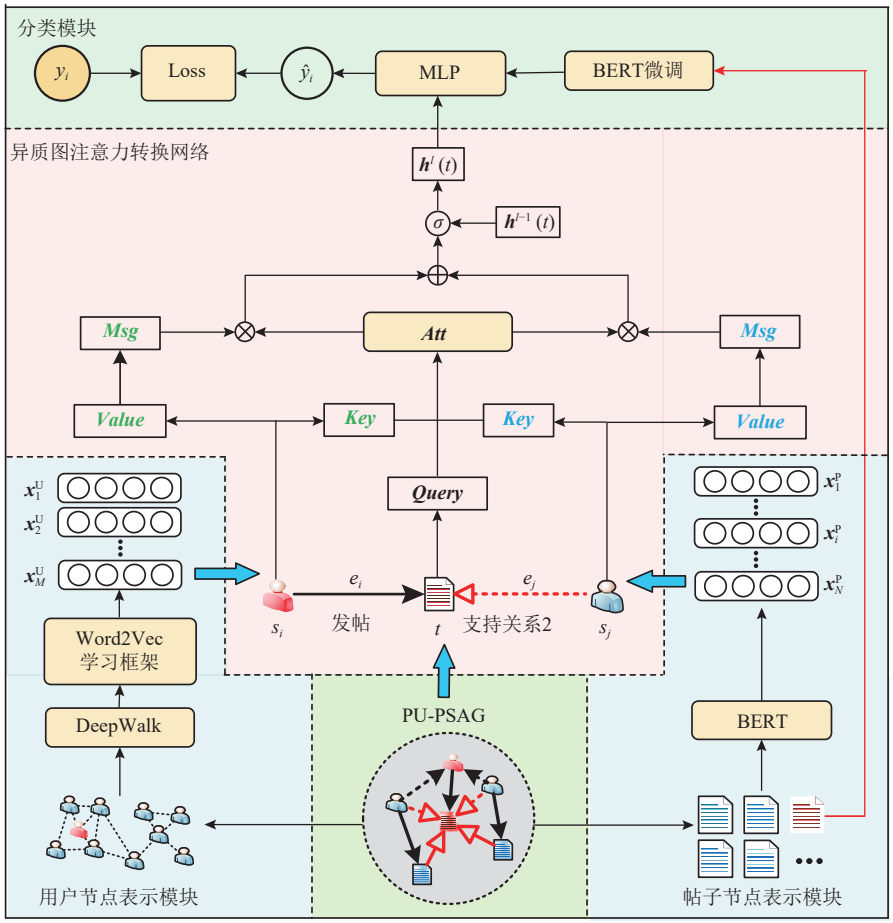


Fig. 2 GSID model

图 2 GSID 模型

帖子节点和用户节点的表示. 4)分类模块. 针对更新后的帖子节点, 预测其自杀意念强度标签.

3.1 帖子节点表示模块

对于数据集 D 中的第 i 个帖子 p_i, 采用 BERT[27] 初始化其表示, 即把帖子 p_i 输入到 BERT 模型中, 并将 [CLS] 符号对应的输出向量作为帖子表示 $\boldsymbol{x}_i^{\mathrm{P}}$, 如式(3)所示.

$$\boldsymbol{x}_i^{\mathrm{P}} = BERT(p_i) \in \mathbb{R}^{1\times d}, \tag{3}$$

其中 $1\leqslant i\leqslant |D|$, $|D|$为帖子数量, $\boldsymbol{x}_i^{\mathrm{P}}$的向量维度为 d.

3.2 用户节点表示模块

为了获取蕴含了医患关系和社群关系等结构信息的用户表示, 从 PU-PSHG 中析取出仅包含用户节点以及用户之间连边的用户心理支持关系子图, 采用 DeepWalk 算法和 Word2Vec 预训练框架, 编码学习用户节点的表示.

DeepWalk 算法是一种网络节点潜在表征方法, 可以用来学习网络中的节点关系以及社群关系. 其思路是, 使用随机游走的方法获得节点在“用户心理支持关系子图”中的局部信息, 并将这些局部信息视为自然语言处理中的句子, 即, 以节点为单词, 以一次随机游走得到的节点序列为一个句子, 将得到的“句子”集合送入 Word2Vec 学习框架, 学习得到用户节点的初始化向量. 第 j 个用户 u_j 的初始化向量 $\boldsymbol{x}_j^{\mathrm{U}}$, 如式(4)所示.

$$\boldsymbol{x}_j^{\mathrm{U}} = Word2Vec(u_j) \in \mathbb{R}^{1\times d}, \tag{4}$$

其中 $1\leqslant j\leqslant M$, M 为用户数量, $\boldsymbol{x}_j^{\mathrm{U}}$的向量维度为 d.

3.3 异质图注意力转换网络

为了编码 PU-PSHG 中节点的表示, 受 Hu 等人[28] 的启发, 本文基于图神经网络框架[29-30] 设计了异质图注意力转换网络. 在构建图注意力聚合模块时, 通过引入可学习的参数, 建模不同的关系类型. 该聚合模块一共包含计算源节点和目标节点之间的注意力、计算源节点和目标节点之间的聚合内容、更新目标节点的向量表示 3 部分.

1)计算源节点和目标节点之间的注意力

在 PU-PSHG 中, 用户节点融合了医患关系或社群关系, 帖子节点包含求助内容和心理支持内容. 不同类型的节点所表达的信息不同, 不同类型的边所表达的关系也不同. 针对 PU-PSHG 中的邻居节点对, 本文将有向边所指向的节点看成目标节点 t, 并映射为向量 ***Query***; 将有向边的起始节点看成源节点 s, 映射为向量 ***Key***. 由于不同类型节点所表达的信息不同, 为不同类型的节点都单独分配线性投影矩阵 $\boldsymbol{W}^{\mathrm{Q}}$ 与 $\boldsymbol{W}^{\mathrm{K}}$, 如式(5)(6)所示.

$$\boldsymbol{Query}^{l}(t) = \boldsymbol{W}_{Type(t)}^{\mathrm{Q}}\boldsymbol{h}^{l-1}(t), \tag{5}$$

$$\boldsymbol{Key}^{l}(s) = \boldsymbol{W}_{Type(s)}^{\mathrm{K}}\boldsymbol{h}^{l-1}(s), \tag{6}$$

其中 $Type\in\{\text{user, post}\}$表示节点 t 或 s 的类型, $\boldsymbol{W}^{\mathrm{Q}}$, $\boldsymbol{W}^{\mathrm{K}}$ 为权重矩阵, l 为图神经网络聚合的迭代层数, $\boldsymbol{h}^0(t)$ 等于 $\boldsymbol{x}_t^{\mathrm{P}}$或者 $\boldsymbol{x}_t^{\mathrm{U}}$, $\boldsymbol{h}^0(s)$等于 $\boldsymbol{x}_s^{\mathrm{P}}$或者 $\boldsymbol{x}_s^{\mathrm{U}}$.

由于不同类型的边表达的关系不同, 例如, 发帖关系 R^{UP} 反映的是作者自身的心理状态, 帖子对帖子的回复关系 R^{PP} 反映了帖子内容上的相关性, 用户对用户的支持关系 R^{UU} 反映的是用户之间的医患关系或社群关系, 用户对帖子的支持关系 R^{PU} 反映了用户擅长的领域或用户具有类似的经历或经验. 因此, 本文为不同类型的关系分配了不同的注意力权重 $\boldsymbol{W}_e^{\mathrm{Att}}$. 对于邻居集合为 $N(t)$ 的目标节点 t, 源节点 $s\in N(t)$ 在目标节点 t 上的多头注意力如式(7)所示.

$$\boldsymbol{Att}_k(s,e,t) = \underset{\forall s\in N(t)}{softmax}\left(\boldsymbol{Key}^{l}(s)\boldsymbol{W}_e^{\mathrm{Att}}\boldsymbol{Query}^{l}(t)^{\mathrm{T}}\frac{\mu_e}{\sqrt{d}}\right), \tag{7}$$

其中 k 为多头注意力, e 用于区分关系的类型, μ_e是可学习的关系因子, $\sqrt{d}$为缩放因子.

2)计算源节点和目标节点之间的聚合内容

与向量 ***Key***, ***Query*** 类似, 将源节点 $s\in N(t)$ 映射为向量 ***Value***. 由于不同类型节点的差异, 对不同类型的节点使用不同的线性投影矩阵. 此外, 考虑到不同心理支持关系消息传递内容的差异, 为不同类型的心理支持关系都单独分配一个消息传递权重 $\boldsymbol{W}_e^{\mathrm{Msg}}$, 以便区分不同心理支持关系类型的消息传递的信息, 如式(8)(9)所示.

$$\boldsymbol{Value}^{l}(s) = \boldsymbol{W}_{Type(s)}^{\mathrm{V}}\boldsymbol{h}^{l-1}(s), \tag{8}$$

$$\boldsymbol{Msg}_k(s,e,t) = \boldsymbol{Value}^{l}(s)\boldsymbol{W}_e^{\mathrm{Msg}}, \tag{9}$$

其中 $\boldsymbol{W}_{Type(s)}^{\mathrm{V}}$为权重矩阵.

3)更新目标节点的向量表示

为得到聚合了各种关系类型语义的节点表示, 本文利用注意力权重信息和邻居节点所传递的内容更新目标节点的向量表示, 如式(10)(11)所示.

$$\tilde{\boldsymbol{h}}^{l}(t) = \Big\Vert_{k=1}^{K}\bigoplus_{\forall s\in N(t)}(\boldsymbol{Att}_k(s,e,t)\otimes\boldsymbol{Msg}_k(s,e,t)), \tag{10}$$

$$\boldsymbol{h}^{l}(t) = \sigma(\tilde{\boldsymbol{h}}^{l}(t))\boldsymbol{W}^{\mathrm{H}}\oplus\boldsymbol{h}^{l-1}(t), \tag{11}$$

其中$\Vert$为拼接操作, $\oplus$为向量的相加, $\otimes$为向量的数乘, $\boldsymbol{W}^{\mathrm{H}}$为权重矩阵, σ为 GELU 激活函数, K 为多头注意力个数.

3.4 分类模块

基于 PU-PSHG 图获取帖子表示时融合了帖子、

用户以及心理支持关系，弱化了帖子原本的语义信息，本文利用 BERT 微调，再次获取第 i 个帖子节点的表示 $\boldsymbol{h}_i^{\mathrm{B}}$，如式(12)所示.

$$\boldsymbol{h}_i^{\mathrm{B}} = BERT_{\text{fine-tuning}}(p_i) \in \mathbb{R}^{1\times d}, \tag{12}$$

其中 d 为向量的维度. 然后，将语义编码 $\boldsymbol{h}_i^{\mathrm{B}}$ 和异质图注意力转换网络最后一层中帖子节点 p_i 的编码 $\boldsymbol{h}^l(p_i)$ 进行拼接，送入全连接层进行解码，如式(13)所示.

$$\hat{\boldsymbol{y}}_i = (\boldsymbol{h}_i^{\mathrm{B}} \| \boldsymbol{h}^l(p_i))\boldsymbol{W}^{\mathrm{O}} + \boldsymbol{b}, \tag{13}$$

其中 $\boldsymbol{W}^{\mathrm{O}}$ 为权重矩阵，$\boldsymbol{b}$ 为偏置值. 本文用交叉熵作为训练损失，如式(14)所示.

$$L = -\frac{1}{|D|}\sum_{i=1}^{|D|}\sum_{c=1}^{C} y_{ic} \lg \hat{y}_{ic}, \tag{14}$$

其中 $|D|$ 和 C 分别为数据集中的样本个数和类别标签个数，$\hat{y}_{ic}$，y_{ic} 分别为第 i 个帖子的预测标签和黄金标签.

3.5 模型算法描述与复杂度分析

整个异质图注意力转换网络的节点表示计算和更新过程如算法 1 所示.

算法 1. 异质图注意力转换网络节点表示计算和更新算法.

输入：$G_{\text{PU-PSHG}} = \langle V, E\rangle$，输入节点特征 $\{\boldsymbol{x}_t: t\in V\}$，网络深度 l；权重矩阵 $\boldsymbol{W}_e^{\mathrm{Att}}$，$\boldsymbol{W}_{Type(s)}^{\mathrm{Value}}$，$\boldsymbol{W}_e^{\mathrm{Msg}}$；可学习的关系因子 μ_e，缩放因子 $\sqrt{d}$；

输出：$G_{\text{PU-PSHG}}$ 中节点 $t\in V$ 表示 $\boldsymbol{r}_t$.

① $\boldsymbol{h}^0(t) \leftarrow \boldsymbol{x}_t$，$\boldsymbol{h}^0(s) \leftarrow \boldsymbol{x}_s$，$t$ 表示目标节点，s 表示源节点；

② for $l = 1, 2, \cdots, L$ do

③ for $t\in V$ do

④ $\boldsymbol{Att}_k(s,e,t) \leftarrow \underset{\forall s\in N(t)}{softmax}\left(\boldsymbol{h}^l(s)\boldsymbol{W}_e^{\mathrm{Att}}\boldsymbol{h}^l(t)^{\mathrm{T}}\dfrac{\mu_e}{\sqrt{d}}\right)$；

⑤ $\boldsymbol{Value}^l(s) \leftarrow \boldsymbol{W}_{Type(s)}^{\mathrm{Value}}\boldsymbol{h}^{l-1}(s)$；

⑥ $\boldsymbol{Msg}_k(s,e,t) \leftarrow \boldsymbol{Value}^l(s)\boldsymbol{W}_e^{\mathrm{Msg}}$；

⑦ end for

⑧ $\tilde{\boldsymbol{h}}^l(t) \leftarrow \Big\|_{k=1}^{K} \underset{\forall s\in N(t)}{\oplus} (\boldsymbol{Att}_k(s,e,t)\otimes \boldsymbol{Msg}_k(s,e,t))$；

⑨ $\boldsymbol{h}^l(t) \leftarrow \sigma(\tilde{\boldsymbol{h}}^l(t))\boldsymbol{W}^{\mathrm{H}} \oplus \boldsymbol{h}^{l-1}(t)$；

⑩ end for

⑪ $\boldsymbol{r}_t \leftarrow \boldsymbol{h}_t^L$，$t\in V$.

为便于直观评估 GSID 模型的效率与性能，本文对模型的复杂度进行简要分析. 假设 $G_{\text{PU-PSHG}} = \langle V, E\rangle$ 中节点的数量、边的数量与边的种类分别为 $|V|$，$|E|$，$|e|$，输入向量的维度与隐藏层向量的维度分别为 F，F'. GSID 模型的核心为异质图注意力转换网络(3.3 节)，故模型的复杂度由输入时的特征转换 $O(|e|\times V\times F\times F')$ 与注意力计算 $O(|e|\times|E|\times F')$ 决定，即模型的时间复杂度为 $O(|e|\times(V\times F\times F' + |E|\times F'))$.

4 实验与结果

4.1 数据集

本文实验使用 CLPsych2017 共享任务数据集. 该数据集[31]来自 ReachOut 在线论坛，发表时间从 2012 年 1 月到 2017 年 3 月，包括 2 606 个用户发表的 143 183 条帖子集合，每条帖子有帖子 ID、匿名用户 ID、帖子内容、帖子之间的回复关系等信息. 其中，1 588 条帖子被人工标注了自杀意念强度，1 188 条帖子用作训练，400 条帖子用作测试. 自杀意念强度反映了帖子需要心理健康专家或版主给予关注的紧急程度，分为 4 个等级：Green, Amber, Red, Crisis, 分别表示“不需要任何干预”“不紧急”“紧急”“非常紧急”，4 类样本的分布情况如表 1 所示. 帖子-用户交互心理支持异质图基于整个数据集构建，包含 2 606 个用户和 143 183 条帖子.

Table 1 Sample Category Distribution

表 1 样本类别分布

类标签	训练数据个数（占比/%）	测试数据个数（占比/%）
Crisis	40（3.4）	42（10.5）
Red	137（11.5）	48（12.0）
Amber	296（24.9）	94（23.5）
Green	715（60.2）	216（54.0）
总数	1 188	400

4.2 评测指标

本文的评测指标采用 CLPsych 共享任务[31]的官方评测指标.

1）Non-green $F1$. 该指标是 Crisis, Red, Amber 三类样本上 $F1$ 的宏观平均. 该指标反映模型对非 Green 类样本的检测能力，是 CLPsych2017 对各参赛系统进行排名的指标.

2）All $F1$. 该指标是 Crisis, Red, Amber, Green 四类样本上 $F1$ 的宏观平均. 该指标反映模型对各类样本的检测能力，是 CLPsych2019 对各参赛系统进行排名的指标.

3）All Acc. 该指标是分类正确的样本数占总样本数的比例，反映模型在所有样本上检测的正确率.

4）Green $F1$, Amber $F1$, Red $F1$, Crisis $F1$. 这些指标是指对 Green, Amber, Red, Crisis 四类样本分类结果的 $F1$，反映模型对单个类别的检测能力.

5）Flagged $F1$. 该指标包括 Flagged 类（Amber, Red,

Crisis 三类样本的并集)和非 Flagged 类(即 Green 类)上 *F*1 的宏观平均,不考虑自杀意念的程度,仅反映模型对有自杀意念的帖子的检测能力.

6)Flagged *Acc*. 该指标是指 Flagged 类和非 Flagged 类(即 Green 类)上的正确率.

7)Urgent *F*1. 该指标是指 Urgent 类(Red 类与 Crisis 类样本的并集)和非 Urgent 类上 *F*1 的宏观平均,反映模型对需要紧急处理和不需要紧急处理的样本的区分能力.

8)Urgent *Acc*. 该指标是指 Urgent 类和非 Urgent 类上的正确率.

4.3 实验环境

实验的环境和模型的超参数设置如表 2 所示.

预训练语言模型采用“BERT-base-uncased”,模型采用了 12 层 Transformer,隐藏大小为 768,多头注意力的参数为 12. DeepWalk 算法中随机游走的步数设为 30,每个节点循环遍历 10 次,Word2Vec 学习框架中窗口大小设置为 5.

Table 2 Experimental Environment and Hyper-Parameter Settings

表 2 实验环境和超参数设置

实验环境	环境配置
操作系统	Linux
处理器	NVIDIA GeForce A40
内存/GB	128
编程语言	Python 3.6
深度学习框架	PyTorch 1.11.0
参数	**参数值**
learning_rate	5E–5
Epochs	20
Dropout	0.1
batch_size	32
K	8
L	1
D	768

4.4 基线模型

为了验证 GSID 模型的有效性,本文选择与 CLPsych2017 共享任务[31]中排名前 4 的模型进行比较,它们的参赛人员或队伍分别是 Xia 等人、Altszyler、Nair 等人、Han 等人. 表 3 中用参赛队伍名代表其模型名.

除此之外,还选择以下 7 个基线模型进行比较:

1)C-LSTM. 该模型由 Sawhney 等人[5]提出,在自杀意念检测任务上有稳健的表现. 该模型利用 Word2Vec 的词向量作为输入,使用 CNN 模型来捕获短语的局部特征,使用 LSTM 模型来捕获句子全局和序列信息.

2)SDM(suicide detection model). 由 Cao 等人[7]提出,学习自杀相关词的嵌入表示,以增强自杀相关词汇和上下文的敏感性,该模型使用微调后的 FastText 编码帖子作为输入.

3)SISMO. 由 Sawhney 等人[9]提出,通过软标签引入自杀风险水平的等级次序,利用 BERT 预训练获

Table 3 Experimental Result on Official Measurement Metrics of CLPsych Shared Task

表 3 在 CLPsych 共享任务官方评测指标上的实验结果

模型	Green *F*1	Amber *F*1	Red *F*1	Crisis *F*1	Non-green *F*1	Flagged		Urgent		All	
						*F*1	*Acc*	*F*1	*Acc*	*F*1	*Acc*
Han 等人*					0.447	0.843	0.858	0.571	0.798		0.683
Nair 等人*					0.461	0.800	0.820	0.571	0.843		0.678
Altszyler*					0.462	0.881	0.883	0.598	0.825		0.695
Xia 等人*					0.467	0.845	0.877	0.591	0.746	0.569	0.708
C-LSTM[5]	0.761	0.404	0.353	0.245	0.334	0.741	0.743	0.718	0.828	0.441	0.580
SDM[7]	0.786	0.396	0.263	0.344	0.334	0.743	0.750	0.654	0.800	0.447	0.608
SISMO[9]	0.888	0.564	0.449	0.346	0.453	0.870	0.873	0.755	0.858	0.562	0.723
BiLSTM-Attention	0.864	0.558	0.520	0.441	0.506	0.854	0.855	0.784	0.863	0.596	0.708
BERT[30]	**0.896**	0.582	0.462	0.488	0.511	**0.887**	**0.888**	0.778	0.857	0.607	0.732
C-GraphSAGE[24]	0.892	0.572	0.451	0.520	0.515	0.880	0.881	0.776	0.856	0.609	0.729
HCN[18]	0.862	0.539	0.473	0.342	0.451	0.850	0.858	0.573	0.833	0.554	0.690
GSID(本文)	0.887	**0.594**	**0.521**	**0.549**	**0.555**	0.878	0.879	**0.795**	**0.870**	**0.638**	**0.739**

注:“*”表示用 CLPsych2017[31]中的参赛队员 (队伍) 名称指代模型,黑体数值表示最优结果.

得的词向量作为输入.

4）BiLSTM-Attention. 双向 LSTM 模型和注意力网络，广泛用于本文分类任务中，SDM 模型与 SISMO 模型均是该模型的变体. 该模型将 BERT 预训练获得的词向量送入双向 LSTM 层，学习单词的隐藏表示，随后将学习到的隐藏信息送入注意力层，以重点关注帖子中的部分单词.

5）BERT. 由 Kipf 等人[30]提出预训练模型 BERT，利用下游任务对模型参数进行微调.

6）C-GraphSAGE. 由 Lee 等人[24]提出基于上下文-图神经网络的自杀检测模型，通过帖子和自杀词之间的关系来学习自杀词汇的上下文语义. 模型利用 BERT 获得帖子向量，利用 Word2Vec 获得自杀词向量，通过构建自杀词与帖子以及自杀词与自杀词之间的关系，并利用 GraphSAGE，使得模型可以融合帖子与自杀词的信息.

7）HCN. 由 Sawhney 等人[18]提出的 HCN 模型认为在线对话可作为理解帖子的重要背景. HCN 使用来自在线社区的会话树代替用户的历史帖子，并基于会话树对帖子进行自杀意念检测.

4.5 实验结果

评测结果如表 3 所示. 总体上，本文提出的 GSID 模型在各个评测指标上都显著优于其他模型，Non-green $F1$ 和 All $F1$ 分别为 0.555 和 0.638，比 BERT 模型分别高出 8.6% 和 5.1%，比 C-GraphSAGE 模型分别高出 7.8% 和 4.8%. 对表 3 中实验结果分析可得出以下结论：

1）心理支持关系的利用有助于提升自杀意念检测的性能. 同为图神经网络模型的 C-GraphSAGE，通过融合相关帖子和自杀词来丰富当前帖子的信息表示，其 Crisis $F1$，Non-green $F1$，All $F1$ 分别为 0.520，0.515，0.609，比 BERT 模型分别提升了 5.6%，0.8%，0.3%. 这表明发现和利用帖子与自杀词的关系可以辅助帖子的预测，其对非常紧急帖子的预测效果提升明显，但却无法更好地区分细粒度的自杀意念强度. 相比 C-GraphSAGE 模型可知，GSID 在 Crisis $F1$，Non-green $F1$，All $F1$ 上分别提升 5.6%，7.8%，4.8%，说明挖掘和利用心理支持关系，不仅有助于检测出非常紧急帖子，也有助于判断自杀意念的强度等级.

2）融合了预训练语言模型和 PU-PSHG 的 GSID 具有明显优势. 从 CLPsych2017 共享任务的官方评测指标 Non-green $F1$ 和 All $F1$ 看，不使用预训练语言模型的深度学习模型并不比基于特征工程的模型更好. 在 CLPsych2017 共享任务中排名靠前的模型大都采用传统特征工程的方法，其中 Xia 等人提出的模型在 Non-green $F1$ 和 All $F1$ 上分别为 0.467 和 0.569，超过采用深度学习的模型 C-LSTM、SDM 以及使用 BERT 预训练获得词向量的 SISMO. 这说明，样本数量较少时，采用传统特征工程的方法也可以获得不错的效果. 预训练语言模型 BERT 在各项指标上较深度学习模型和基于特征工程的模型都有大幅度的提升，其 Non-green $F1$ 和 All $F1$ 达到 0.511 和 0.607. 而在 BERT 上引用帖子-用户交互心理支持异质图后，GSID 模型在 Non-green $F1$ 和 All $F1$ 进一步提升，提升幅度分别为 8.6% 和 5.1%，说明在自杀意念检测任务上，基于心理支持关系建立起的用户关系是帖子内容的有益补充.

表 3 还反映了一个有趣的现象，尽管 SDM 和 SISMO 基于 BiLSTM-Attention，但它们在自杀意念检测上却不如 BiLSTM-Attention. 进一步观察发现，SISMO 在 Green $F1$ 和 Amber $F1$ 上分别达到 0.888 和 0.564，较 BiLSTM-Attention 的 0.864 和 0.558 要高，但在 Red $F1$ 和 Crisis $F1$ 上只有 0.449 和 0.346，远低于 BiLSTM-Attention 的 0.520 和 0.441. 原因在于，自杀意念强度检测数据集是严重不均衡的，训练集中 Red 类和 Crisis 类仅占 11.5% 和 3.4%，而 Green 类和 Amber 类占 60.2% 和 24.9%. SISMO 模型引入的软标签期望错分的标签在自杀意念强度上要与真实标签更接近，当 Green 类和 Amber 类占有近 85% 的样本时，这种软标签的效果很难在 Red 类和 Crisis 类上体现出来. SDM 模型表现不佳的原因在于它通过增强自杀相关词汇的影响来提升模型的能力，其效果极大地依赖于自杀词典，本文采用 SDM 方法时使用的自杀词典不一定适合 CLPsych2017 数据，干扰了词嵌入的表示能力.

3）BERT 在 Green 与 Flagged 指标上表现好于 GSID. 从 4.2 节的评测指标说明中可以看出，Green 与 Flagged 指标均为 2 分类任务，分类的粒度不够细腻，仅关注通用语义也能够得到较准确的分类结果，因此，帖子-用户交互心理支持异质图的领域优势无法体现. 细化分类的粒度后，GSID 的表现都要明显好于 BERT. 例如在 Non-green 中，由于任务变为更细粒度的 3 分类，虽然是对 Amber，Red，Crisis 的分类，BERT 的表现却显著下降.

4）HCN 的表现相较于 C-GraphSAGE，BERT，GSID 有明显差距. 通过分析模型以及 CLPsych2017 数据集的特点，HCN 表现不佳的原因可归纳为以下 2 点：首先，HCN 旨在利用以目标帖子为根节点的对话树进行自杀意念检测，然而 CLPsych2017 数据集中有大量

待检测的目标帖子并非处于对话树中根节点的位置,反而处在叶节点位置,因此,没有充分的辅助信息可被 HCN 模型利用. 其次,HCN 仅对“在线心理健康论坛”中的帖子进行建模,没考虑用户节点、节点间边的类型等信息.

5 实验分析

为了进一步分析和确定影响自杀意念检测效果的具体因素,我们做了消融实验和模型冷启动验证实验,并对 GSID 模型部分实验结果进行实例分析.

5.1 消融实验

1)心理支持关系的影响

在帖子-用户交互心理支持异质图中分别去除帖子与帖子的回复关系 R^{PP}、用户对帖子的心理支持关系 R^{PU} 和用户对用户的心理支持关系 R^{UU},分析不同类型的心理支持关系对帖子自杀意念检测的影响. 实验结果如表 4 所示.

Table 4 Influence of Psychological Support Relationships on GSID Model

表 4 心理支持关系对 GSID 模型的影响 %

模型	Green Δ*F*1	Amber Δ*F*1	Red Δ*F*1	Crisis Δ*F*1	Non-green Δ*F*1	Flagged		Urgent		All	
						Δ*F*1	Δ*Acc*	Δ*F*1	Δ*Acc*	Δ*F*1	Δ*Acc*
w/o R^{PP}	0.11	–0.04	–7.95	–1.64	–3.04	–0.03	0.00	–0.50	–0.26	–1.95	–0.41
w/o R^{PU}	–0.33	–2.68	–8.52	–0.54	–3.80	–0.62	–0.57	0.44	0.20	–2.59	–1.05
w/o R^{UU}	0.74	–2.43	–7.73	–8.72	–6.17	0.56	0.60	–1.26	–0.95	–3.76	–0.85
平均值	0.17	–1.72	–8.07	–3.63	–4.34	–0.03	0.01	–0.44	–0.34	–2.77	–0.77

注:Δ 表示对应指标的变化.

从异质图中移除 R^{PP} 后,模型的 Non-green *F*1 和 All *F*1 分别下降 3.04% 和 1.95%;移除 R^{PU} 后,Non-green *F*1 和 All *F*1 分别下降 3.80% 和 2.59%;移除 R^{UU} 后,模型的性能下降最多,Non-green *F*1 和 All *F*1 分别下降 6.17% 和 3.76%. 实验说明,帖子-用户交互心理支持异质图中的 3 类关系对模型都有帮助. 其中,R^{UU} 对模型效果影响最大,用户之间心理支持行为能够反映在线心理健康论坛上用户之间独特的社交关系,这种关系对自杀意念检测有较大帮助.

此外,观察表 4 可知,心理支持关系在不同强度的自杀意念检测上表现出明显的差异. 总体上看,各类心理支持关系对 Green *F*1 没有太大影响,而对 Amber *F*1, Red *F*1, Crisis *F*1 的影响较大. 分别去掉 3 类心理支持关系后,GSID 在 Green *F*1 上平均提高 0.17%,而在 Amber *F*1, Red *F*1, Crisis *F*1 上分别下降 1.72%, 8.07%, 3.63%,进一步说明心理支持关系对非 Green 类帖子的检测帮助较大,特别是对自杀意念较强的 Red 和 Crisis 类帖子.

2)医患关系或社群关系挖掘的影响

GSID 模型使用 DeepWalk 挖掘帖子-用户交互心理支持异质图中的用户关系网络,尝试发现用户之间形成的医患关系或社群关系,并将其编码在用户节点的初始向量中. 为了验证医患关系挖掘和社群关系挖掘对模型的影响,不使用 DeepWalk 算法初始化用户节点向量,而是通过用户发布的所有帖子对用户进行初始化(采用帖子向量的平均值),实验结果如表 5 所示.

去掉医患挖掘或社群关系挖掘后,GSID 的性能显著下降,特别是在 Non-green 类帖子的检测上,Non-green *F*1 下降 2.02%,其中 Amber *F*1 和 Red *F*1 分别下降 1.29% 和 4.51%. 原因有 2 点:1)Non-green 类帖子样本数量较少,医患关系或社群关系的引入弥补了训练样本不足的问题;2)由有自杀意念用户所形成的医患关系或社群关系更紧密可靠,因此对 GSID 模型的帮助更大. 这证明了采用 DeepWalk 挖掘医患关系或社群关系并初始化用户节点的方法是有效的.

3)训练集用户对测试集帖子的影响

本文中自杀意念检测的对象是用户发布的某条

Table 5 Influence of Initializing User Nodes

表 5 初始化用户节点的影响 %

模型	Green Δ*F*1	Amber Δ*F*1	Red Δ*F*1	Crisis Δ*F*1	Non-green Δ*F*1	Flagged		Urgent		All	
						Δ*F*1	Δ*Acc*	Δ*F*1	Δ*Acc*	Δ*F*1	Δ*Acc*
w/o DeepWalk	–0.32	–1.29	–4.51	–0.44	–2.02	–0.76	–0.68	–0.23	0.26	–1.43	–0.24

注:Δ 表示对应指标的变化.

帖子，在构建帖子-用户交互心理支持异质图时使用了采集到的全部用户和帖子，因此存在某个用户发布的多条帖子分别出现在训练集和测试集中的情况. 统计显示，在 400 条测试数据中，与训练集中某条帖子来自相同用户的有 95 条，称这部分帖子的集合为 T_1，其余的 305 条帖子的集合为 T_2.

GSID 模型在 T_1 和 T_2 上的评测结果如图 3 所示. 总体上看，T_1 上的 All $F1$ 和 All Acc 分别为 0.623 和 0.740，T_2 上的 All $F1$ 和 All Acc 分别为 0.640 和 0.738. 对于没有自杀意念的 Green 类帖子，知晓其用户发布的某条帖子的自杀意念强度，能够显著提升检测的效果，Green $F1$ 从 0.887 提升到 0.906，提升了 2.15%. 这主要是因为一个没有自杀意念的用户，其发布的帖子大都是没有自杀意念的. 对于 Amber 类帖子，知晓其用户发布的某条帖子的自杀意念强度，并不能提升检测的效果，All $F1$ 从 0.594 下降到 0.566，下降了 4.71%. 这主要是因为 Amber 类帖子中的自杀意念比较模糊，知晓其用户发布的某条帖子的自杀意念强度，反而干扰了判断.

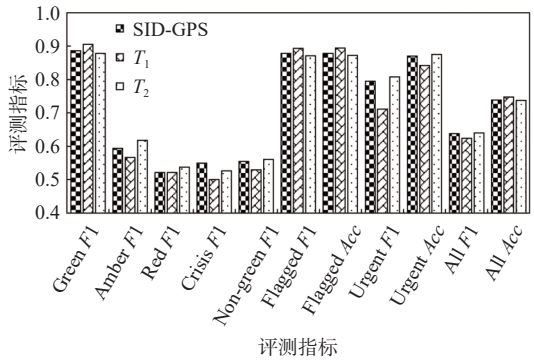


Fig. 3 Influence of the intensity of suicidal ideation of a known post on the detection of other posts of the user

图 3 已知某条帖子的自杀意念强度对该用户其他帖子检测的影响

图 3 还显示，对于有自杀意念的非 Green 类帖子，即使不知道其用户发布的某条帖子的自杀意念强度，模型也能表现出较好的性能，All $F1$，Red $F1$，Crisis $F1$ 分别从 0.566，0.522，0.500 提升到 0.618，0.537，0.526，提高了 9.21%，2.99%，5.26%. 这主要是因为一个有自杀意念的用户，其自杀意念会反映在所发布的多条帖子中. 此外，心理支持关系依赖于论坛内的支持行为，即使不知道用户的历史信息，模型也能较好地区分帖子的自杀意念.

5.2 GSID 模型的冷启动问题

GSID 模型在构建帖子-用户交互心理支持异质图时，使用了整个数据集中的全部用户和帖子. 在模型的实际使用时，对于新来的用户或者某个用户新发布的帖子，重新构建帖子-用户交互心理支持异质图并学习帖子的表示代价太大. 对于该冷启动问题，GSID 模型采用如下朴素的策略.

1）新帖子问题（$G_{\text{PU-PSHG}}$ 中的用户所发布的新帖子）. 对于用户 u_j（$u_j \in G_{\text{PU-PSHG}}$）发布的帖子 p_i（$p_i \notin G_{\text{PU-PSHG}}$），聚合节点 u_j 的 1 层（l=1）邻居节点，得到节点 u_j 的编码 $\boldsymbol{h}_j^l$，通过微调后的 BERT 模型获取的帖子 p_i 的语义编码 $\boldsymbol{h}_i^{\text{B}}$，拼接 $\boldsymbol{h}_j^l$ 和 $\boldsymbol{h}_i^{\text{B}}$，送入到解码层中.

2）新用户问题（不在 $G_{\text{PU-PSHG}}$ 中的新用户所发布的帖子）. 对于用户 u_j（$u_j \notin G_{\text{PU-PSHG}}$）发布帖子 p_i（$p_i \notin G_{\text{PU-PSHG}}$），首先通过 BERT 获取帖子 p_i 的向量表示；再基于用户全部帖子向量的平均值，利用余弦相似度从 $G_{\text{PU-PSHG}}$ 中筛选出 10 个最相似的用户作为 u_j 的邻居；最后将 p_i 视为老用户发布的新帖子，预测其标签.

实验时，分别将测试集中的帖子和用户（含用户的全部帖子）从 $G_{\text{PU-PSHG}}$ 中删除，用于模拟 GSID 模型应对“新帖子”和“新用户”的冷启动能力. 如表 6 所示，删除测试集中的帖子后，$G_{\text{PU-PSHG}}$ 中包含 2 597 个用户和 142 783 条帖子；删除测试集中的用户以及其发布的全部帖子后，$G_{\text{PU-PSHG}}$ 中包含 2 447 个用户和 71 554 条帖子. 需要说明的是，在模拟“新帖子”类的冷启动问题时，从 $G_{\text{PU-PSHG}}$ 中删除测试集中的帖子后，如果 $G_{\text{PU-PSHG}}$ 中某用户发布的帖子数量为 0，则从 $G_{\text{PU-PSHG}}$ 中删除该用户，此时冷启动问题退化为“新用户”问题，这类帖子在测试集中共有 9 条.

Table 6 Cold Start Test Data Set

表 6 冷启动测试数据集

策略	$G_{\text{PU-PSHG}}$ 中用户数量	$G_{\text{PU-PSHG}}$ 中帖子数量	测试集中帖子数量
GSID	2 606	143 183	400
新帖子问题	2 597	142 783	391（9 条退化为“新用户”问题）
新用户问题	2 447	71 554	400

GSID 模型针对“新帖子”和“新用户”的表现如表 7 所示. 实验结果显示，模型对冷启动问题上也有较好的表现.

尽管老用户发布的新帖子或新用户发布的帖子给 GSID 带来了一定的挑战，但 GSID 的表现总体上依然优于 BERT 模型. 在“新帖子”问题上，GSID 比 BERT 在 Non-green $F1$ 和 All $F1$ 上分别高出 1.57% 和 0.66%，在“新用户”问题上分别高出 1.76% 和 0.82%. GSID 在 Urgent 类帖子的冷启动问题上优势明显，其在“新帖子”问题和“新用户”问题上的 Urgent Acc 分

Table 7 Experimental Results on Cold Start Issues

表 7 冷启动问题的实验结果

模型	Green *F*1	Amber *F*1	Red *F*1	Crisis *F*1	Non-green *F*1	Flagged		Urgent		All	
						*F*1	*Acc*	*F*1	*Acc*	*F*1	*Acc*
BERT（只用目标帖，无冷启动问题）	**0.896**	0.582	0.462	0.488	0.511	**0.887**	**0.888**	0.778	0.857	0.607	0.732
GSID（应对“新帖子”问题）	0.889	0.587	0.482	0.488	0.519	0.875	0.877	0.784	0.868	0.611	0.736
GSID（应对“新用户”问题）	0.888	0.588	0.435	0.537	0.520	0.879	0.880	0.784	0.863	0.612	0.728
GSID	0.887	**0.594**	**0.521**	**0.549**	**0.555**	0.878	0.879	**0.795**	**0.870**	**0.638**	**0.739**

注：黑体数值表示最优结果.

别高出 BERT 模型 1.28% 和 0.70%.

此外，由表 7 可知，对于完全没见过的新用户，通过其发布的帖子内容找到相似的用户作为其邻居，这种方式得到的总体效果与老用户发布的“新帖子”类似，这也说明发布相似内容的用户之间具有潜在的心理支持关系.

5.3 实例分析

为直观展示帖子-用户交互心理支持异质图对 GSID 模型的帮助，本节给出 BERT 与 GSID 模型对某帖子自杀意念检测的结果，如图 4 所示.

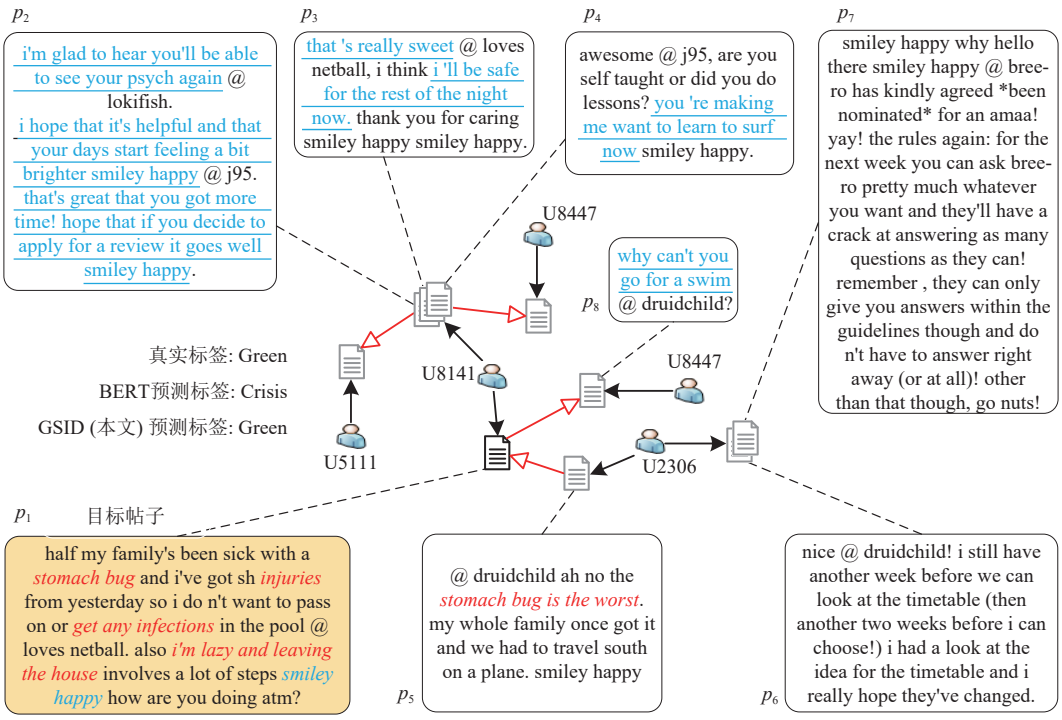


Fig. 4 Case study

图 4 实例分析

图 4 的例子中，BERT 错误地将目标帖子的标签判断为 Crisis，而 GSID 则做出了正确的判断，将目标帖子 p_1 分类为 Green. 观察 p_1 的内容可以发现，p_1 中包含大量消极词语，例如：stomach bug（胃痛）、injuries（受伤）、get any infections（感染）、lazy（懒惰）、leaving the house（出门/离家出走）等. 如果模型（例如 BERT）仅关注目标帖子的语义，那么很容易将其分类为较为严重的自杀意念类别.

然而，观察图 4 可以发现：首先，p_1 是对帖子 p_8 内容（“why can’t you go for a swim @ druidchild?”）的解释，而不是对自身心理状况的描述，所以其中的 get any infections, lazy, leaving the house 并不能反映其具有自杀意念；其次，p_1 的发帖用户 u8141 发布的帖子大都是对其他用户进行心理支持的（对 U5111 以及对 U8447），且帖子内容较为积极（p_2, p_3, p_4）；再次，回复 p_1 的 p_5 也不涉及自杀意念的内容，说明 p_1 中有自杀意念的可能性不大；最后，p_5 的发帖用户 U2306 发布的其他帖子与心理支持的关系不大，说明他们之

间不存在心理支持上的社群关系或医患关系，进一步降低 p_1 中有自杀意念的程度.

综上，对于 GSID 模型而言，因为模型成功建模并融合了帖子-用户交互心理支持异质图中关于用户的心理支持关系与帖子的回复关系，所以 GSID 模型能够对图 4 中的目标帖子类型进行正确的判断.

6 结论和局限性

本文以自杀意念为心理健康检测对象，提出了面向帖子-用户交互心理支持异质图的一种图增强的自杀意念检测模型(GSID)，通过构建并学习用户与帖子之间的心理支持关系，使得模型能够捕获并融入用户的医患关系或社群关系，从而提升模型检测自杀意念帖子的能力. 在 CLPsych2017 共享任务上对 GSID 模型进行全面的评测，实验表明该模型明显优于现有方法，Non-green $F1$, All $F1$, All Acc 分别为 0.555, 0.638, 0.739，平均高出 C-GraphSAGE 基线模型 4.7%.

尽管取得了具有竞争力的效果，但 GSID 模型也存在以下局限性. 1)复杂度受到帖子-用户交互心理支持异质图规模的影响：GSID 模型需要通过利用用户之间的心理支持关系图挖掘医患关系或社群关系，相比于直接利用帖子的模型，初始化用户节点需要更多的计算资源；2)存在冷启动问题：尽管采用朴素的冷启动方式可以解决新用户和新帖子的问题，但效果会显著下降；3)没有区分不同的用户角色：心理健康论坛上，用户有不同的类型，充当着不同的角色，因此在自杀意念检测时应该被区别对待，但 GSID 模型未对用户进行区分，笼统地将所有的心理支持用户同等对待；4)没有考虑医患或社群关系的变化：求助者的心理、支持者的专长会随着时间而变化，GSID 只考虑了静态信息，没考虑异质图的动态变化. 未来工作将从以上 4 个方面考虑.

作者贡献声明：刘德喜提出了模型思路和实验方案；张子靖补充实验并撰写和修改论文；邹志峰开展实验和撰写论文；万齐智修改了论文框架；刘喜平与廖国琼提供了论文的修改思路和实验的改进方向；朱廷劭提出与心理健康相关的建议. 刘德喜和万齐智为共同通信作者.

参 考 文 献

[1] Li Jing, Liu Dexi, Wan Changxuan, et al. A survey of automatic evaluation of mental health of social network users[J]. Journal of Chinese Information Processing, 2021, 35(2): 19−32 (in Chinese)
(李静，刘德喜，万常选，等. 社会网络用户心理健康自动评估研究综述[J]. 中文信息学报，2021，35(2)：19−32)

[2] Freeman M. The world mental health report: Transforming mental health for all[J]. World Psychiatry, 2022, 21(3): 391−392

[3] Zhao Yun, Liu Dexi, Wan Changxuan, et al. Mental health text matching model integrating characters' mental portrait[J]. Journal of Computer Research and Development, 2024, 61(7): 1812−1824 (in Chinese)
(赵芸，刘德喜，万常选，等. 融合角色心理画像的心理健康文本匹配模型[J]. 计算机研究与发展，2024，61(7)：1812−1824)

[4] Yates A, Cohan A, Goharian N. Depression and self-Harm risk assessment in online forums[C]//Proc of the Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2017: 2968−2978

[5] Sawhney R, Manchanda P, Mathur P, et al. Exploring and learning suicidal ideation connotations on social media with deep learning[C]//Proc of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Stroudsburg, PA: ACL, 2018: 167−175

[6] Tadesse M M, Lin Hongfei, Xu Bo, et al. Detection of suicide ideation in social media forums using deep learning[J]. Algorithms, 2019, 13(1): 7−25

[7] Cao Lei, Zhang Huijun, Feng Ling, et al. Latent suicide risk detection on microblog via suicide-oriented word embeddings and layeRed attention[C]//Proc of the 2019 Conf on Empirical Methods in Natural Language Processing and the 9th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2019: 1718−1728

[8] Sawhney R, Joshi H, Gandhi S, et al. A time-aware Transformer based model for suicide ideation detection on social media[C]//Proc of the 2020 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2020: 7685−7697

[9] Sawhney R, Joshi H, Gandhi S, et al. Towards ordinal suicide ideation detection on social media[C]//Proc of the 14th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2021: 22−30

[10] Liu Dexi, Bao Liping, Wan Changxuan, et al. Identification of online forum users' psychological crises based on multi-layer local information fusion[J]. Journal of Chinese Computer Systems, 2021, 42(4): 690−699 (in Chinese)
(刘德喜，鲍力平，万常选，等. 基于多层局部信息融合的在线论坛用户心理危机识别[J]. 小型微型计算机系统，2021，42(4)：690−699)

[11] De Choudhury M, Gamon M, Counts S, et al. PRedicting depression via social media[C]//Proc of the 7th Int AAAI Conf on Web and Social Media. Palo Alto, CA: AAAI, 2013, 7(1): 128−137

[12] Malmasi S, Zampieri M, Dras M. Predicting post severity in mental health forums[C]//Proc of the 3rd Workshop on Computational Linguistics and Clinical Psychology. Stroudsburg, PA: ACL, 2016: 133−137

[13] Liu Dexi, Xia Xianyi, Wan Changxuan, et al. Automatic evaluation of mental health of online forum users based on multi-feature fusion[J]. Chinese Journal of Computers, 2019, 42(7): 1553−1569 (in Chinese)

(刘德喜, 夏先益, 万常选, 等. 基于多特征融合的在线论坛用户心理健康自动评估[J]. 计算机学报, 2019, 42(7): 1553−1569)

[14] Shrestha A, Spezzano F. Detecting depressed users in online forums[C]//Proc of the 11th IEEE/ACM Int Conf on Advances in Social Networks Analysis and Mining. New York: ACM, 2019: 945−951

[15] Mishra R, Sinha P P, Sawhney R, et al. SNAP-BATNET: Cascading author profiling and social network graphs for suicide ideation detection on social media[C]//Proc of the 17th Conf of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop. Stroudsburg, PA: ACL, 2019: 147−156

[16] Sinha P P, Mishra R, Sawhney R, et al. # Suicidal-A multipronged approach to identify and explore suicidal ideation in Twitter[C]//Proc of the 28th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2019: 941−950

[17] Sawhney R, Joshi H, Shah R, et al. Suicide ideation detection via social and temporal user representations using hyperbolic learning[C]//Proc of the 19th Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2021: 2176−2190

[18] Sawhney R, Agarwal S, Neerkaje A T, et al. Towards suicide ideation detection through online conversational context[C]//Proc of the 45th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2022: 1716−1727

[19] Cao Lei, Zhang Huijun, Feng Ling. Building and using personal knowledge graph to improve suicidal ideation detection on social media[J]. IEEE Transactions on Multimedia, 2020, 24: 87−102

[20] Huang Guanwei, Hong Xiaowen, Zhao Daying. Research on the network structure and communication emotion of virtual community[J]. China Soft Science, 2021(Z1): 108−115 (in Chinese)
(黄冠维, 洪晓文, 赵大营. 虚拟社群网络结构及网络沟通情绪的研究[J]. 中国软科学, 2021(Z1): 108−115)

[21] Wray M, Colen C, Pescosolido B. The sociology of suicide[J]. Annual Review of Sociology, 2011, 37: 505−528

[22] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations[C]//Proc of the 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701−710

[23] Mikolov T, Chen Kai, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint, arXiv: 1301.3781, 2013

[24] Lee D, Kang M, Kim M, et al. Detecting suicidality with a contextual graph neural network[C]//Proc of the 8th Workshop on Computational Linguistics and Clinical Psychology. Stroudsburg, PA: ACL, 2022: 116−125

[25] Sun Yizhou, Han Jiawei, Yan Xifeng, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks[J]. Proceedings of the VLDB Endowment, 2011, 4(11): 992−1003

[26] Sun Yizhou, Norick B, Han Jiawei, et al. Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks[J]. ACM Transactions on Knowledge Discovery from Data, 2013, 7(3): 1−23

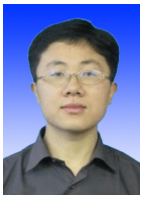
[27] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proc of the 17th Conf of the North American Chapter of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 4171−4186

[28] Hu Ziniu, Dong Yuxiao, Wang Kuansan, et al. Heterogeneous graph transformer[C]//Proc of the 29th Web Conf. New York: ACM, 2020: 2704−2710

[29] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint, arXiv: 1710.10903, 2017

[30] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint, arXiv: 1609.02907, 2016

[31] Guo Jiawen, Danielle L M, Djin L, et al. A corpus analysis of social connections and social isolation in sdolescents suffering from depressive disorders[C]//Proc of the 4th workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality. Vancouver, BC: ACL, 2017: 26−31



Liu Dexi, born in 1975. PhD, professor, PhD supervisor. Member of CCF. His main research interests include social media processing, information retrieval, and natural language processing.

刘德喜, 1975 年生. 博士, 教授, 博士生导师. CCF 会员. 主要研究方向为社会媒体处理、信息检索、自然语言处理.



Zhang Zijing, born in 1994. PhD candidate. His main research interests include social media processing and natural language processing.

张子靖, 1994 年生. 博士研究生. 主要研究方向为社会媒体处理、自然语言处理.



Zou Zhifeng, born in 1998. Master. His main research interest includes natural language processing.

邹志峰, 1998 年生. 硕士. 主要研究方向为自然语言处理.



Wan Qizhi, born in 1988. PhD, lecturer. His main research interests include information extraction, natural language processing, and data mining. (wanqizhi1006@163.com)

万齐智, 1988 年生. 博士, 讲师. 主要研究方向为信息抽取、自然语言处理、数据挖掘.



Liu Xiping, born in 1981. PhD, professor, PhD supervisor. Member of CCF. His main research interests include information extraction, natural language processing, and data mining.

刘喜平, 1981 年生. 博士, 教授, 博士生导师. CCF 会员. 主要研究方向为消息抽取、自然语言处理、数据挖掘.



Liao Guoqiong, born in 1969. PhD, professor, PhD supervisor. His main research interests include social network mining and IoT data management.

廖国琼, 1969 年生. 博士, 教授, 博士生导师. 主要研究方向为社会网络挖掘、物联网数据管理.



Zhu Tingshao, born in 1971. PhD, professor, PhD supervisor. His main research interests include data mining, network behavior analysis, and computational network psychology.

朱廷劭, 1971 年生. 博士, 教授, 博士生导师. 主要研究方向为数据挖掘、网络行为分析、计算网络心理.