

基于 D-S 证据理论的网络表示融合方法

程晓涛^{1,2}, 吉立新¹, 尹 赢¹, 黄瑞阳¹

(1. 战略支援部队信息工程大学, 河南郑州 450002; 2. 中国人民解放军 66061 部队, 北京 100043)

摘 要: 随着网络表示学习技术的发展, 在网络结构信息的基础上, 越来越多的研究者考虑融入额外辅助信息来提升网络表示效果. 针对现有网络表示学习方法中对于多属性特征融合缺乏冲突判别与评价指标的问题, 本文在已有研究基础上, 提出一种基于 D-S 证据理论的网络表示融合方法. 该方法首先通过 SVM 算法给出不同属性信息对融合表示结果的支持度, 然后利用证据组合规则计算网络表示学习中的融合评价指标, 并依据混淆矩阵考虑各类别在节点分类中的局部可信度. 在 3 类数据集上的仿真实验表明: 该方法对于检测网络表示融合中的冲突, 提高表示融合效果具有一定的指导意义.

关键词: 网络表示学习; 特征融合; D-S 证据理论; 冲突判别

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2020)05-0854-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.05.004

Method of Network Representation Fusion Based on D-S Evidence Theory

CHENG Xiao-tao^{1,2}, JI Li-xin¹, YIN Ying¹, HUANG Rui-yang¹

(1. Strategic Support Force Information Engineering University, Zhengzhou, Henan 450002, China;

2. 66061 Troops of the Chinese People's Liberation Army, Beijing 100043, China)

Abstract: With the development of network representation learning technology, researchers are increasingly considering the integration of multi-dimensional attribute information to enhance the performance of network representation. In view of the lack of conflict discrimination and evaluation index for multi-attribute feature fusion in existing network representation learning methods, this paper proposes a network representation learning fusion method based on D-S evidence theory. Firstly, the support degree of different attribute information to the fusion result is given by SVM algorithm. Then, the fusion evaluation index in network representation learning is calculated by using evidence combination rules, and the confidence degree of each attribute is fully considered based on confusion matrix. Simulation results on three types of data sets show that the method can effectively detect conflicts in network representation fusion and improve the performance of fusion representation.

Key words: network representation learning; feature fusion; D-S evidence theory; conflict discrimination

1 引言

表示学习^[1]也称特征学习, 是当前机器学习研究领域的一个重要研究方向. 表示学习的目标是通过对数据表达形式的变化, 使数据所包含的信息更加容易被提取和分析. 例如在文本表示学习中, 通过对文本共现情景的自动学习得到词向量^[2], 这种低维度的表示向量能够有效反映词语之间的语义关系. 而网络表示学习 (Network Representation Learning, NRL) 或者称网络嵌入 (Network Embedding, NE) 就是将网络这种非结构化的数据嵌入到低维空间中, 用低维向量来表征网络中的节点和连边, 或者整个网络, 并以此为输入应用到节点分类、链路预测、社区发现^[3]等网络分析任务中.

当前针对复杂网络的表示学习模型大多根据网络结构特征或者内容特征进行节点的表示学习^[4]. 为了获得更优的网络表示, 研究者们普遍采用在生成网络表示的过程中融合更多外部信息的方法, 如 TADW^[5] (Text Associate Deep Walk) 和 CANE^[6] (Context Aware Network Embedding) 算法是将节点文本信息融入网络表示学习, 作为网络结构信息的补充, 增强网络节点表示的分类效果. Chen 等人提出一种混合群组标签信息的网络表示学习增强算法 GENE^[7] (Group Enhanced Network Embedding), 该算法通过在随机游走序列中添加包含节点标签信息的标签节点, 实现网络结构信息与标签信息的融合; Sandro 等人^[8]则将社区信息融入网络表示学习中, 通过社区嵌入向量与节点嵌入向量的相互影响提高网络表示的准确

性.但上述利用外部信息提高网络表示分类准确性的方法,往往忽视了现实场景中信息来源的差异性与不可靠性,多源信息相互验证可以提升网络表示的性能,但信息冲突也可能会降低融合效果,而当前网络表示学习中针对多属性特征融合的冲突判别与评价指标问题研究还不足.网络表示学习中特征融合的实质是基于不同类别特征信息的多元变量决策问题,D-S 证据理论(Dempster-Shafer evidence theory)作为对概率论的推广,既可以处理数据的不确定性,也能应对数据的多义性,对本文的网络表示融合问题具有一定的借鉴意义.由此,本文在已有网络表示模型^[9]的基础上,提出一种基于 D-S 证据理论的网络表示学习融合方法,将不同网络特征间的融合冲突问题转化为不同类别的后验概率计算问题,利用证据组合规则实现节点特征冲突程度的量化,并在后验概率计算中根据混淆矩阵考虑各类别在分类中的局部可信度,对分类结果进一步优化.

2 问题描述及相关定义

2.1 基本定义

D-S 证据理论又称为 D-S 理论,它是一种处理不确定性问题的完整理论^[10],将 D-S 证据理论用于网络表示学习的特征融合问题是一种新的尝试.本文在此给出基本的符号解释和定义.

定义 1 基本概率分配 (Basic Probability Assignment, BPA) 设 Θ 为识别框架, $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 是一个非空的有限子集,其中的元素两两互斥, Θ 中包含了所有可能的命题,映射 $m:2^\Theta \rightarrow [0,1]$ 满足以下条件:

$$m(\phi) = 0$$

$$0 \leq m(A) \leq 1, \forall A \subseteq \Theta$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{1}$$

则 m 称为辨识框架 Θ 上的基本概率赋值函数,也称为 mass 函数.

定义 2 Dempster 组合规则 对于 $\forall A \subseteq \Theta$, 识别框架 Θ 上有限个 mass 函数 m_1, m_2, \dots, m_n 的合成规则为:

$$(m_1 \oplus m_2 \oplus \dots \oplus m_n)(A) = \frac{1}{K} \sum_{A_1 \cap A_2 \cap \dots \cap A_n = A} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_n(A_n) \tag{2}$$

其中,

$$K = \sum_{A_1 \cap A_2 \cap \dots \cap A_n \neq \phi} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_n(A_n) = 1 - \sum_{A_1 \cap A_2 \cap \dots \cap A_n = \phi} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_n(A_n) \tag{3}$$

其中 K 为归一化因子, $1-K$ 反映了证据的冲突程度.

定义 3 信息网络表示学习^[11] 将信息网络记为 $G = (V, E, D, C)$, V 是节点集合, E 是边的集合, D 是各节点的文本属性信息集合, C 是节点类标签集合. 网络的邻接矩阵定义为 $A \in \mathbb{R}^{|V| \times |V|}$, 其中 $A_{ij} = 1$ 如果 $(v_i, v_j) \in E$, 否则 $A_{ij} = 0$. 网络表示学习的目标是学习映射函数 $f: v_i \rightarrow y_i \in \mathbb{R}^d$, 并且 $d \ll |V|$, 使每个节点 $v \in V$ 学习出一个实数向量.

定义 4 网络表示融合 本文中网络表示融合指通过表示学习得到的节点特征向量如何实现 $L3$ 上融合^[12] ($L1$ 数据层, $L2$ 特征层, $L3$ 决策级). 如图 1 所示, 以文献网络为例, 输入的网络数据包含网络结构信息、节点文本信息和节点标签信息, 输出为实现了结构与文本信息融合的网络节点表示^[13].

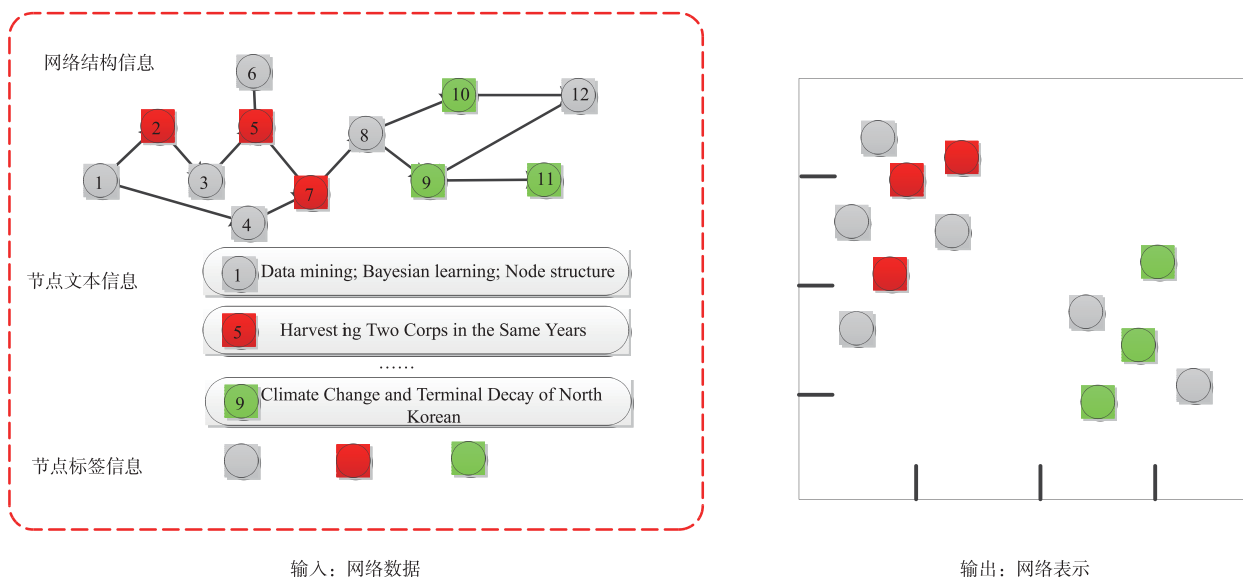


图1 文献网络中表示融合示意图

2.2 网络用户特征表示的构建

本节利用向量表示网络用户的不同类别特征,不同特征在用户分类结果上的差异程度反映特征间的冲突大小,由此可以将不同网络特征间的融合冲突问题转化为不同类别的后验概率计算问题,然后通过 D-S 证据理论度量不同特征间的冲突程度,对于冲突较大的特征,可以根据用户需求制定合理的融合策略.在不

同类别特征向量表示的构建过程中,本节以用户属性特征和网络结构特征为例进行阐述.对于网络结构特征,可以通过常用的网络表示学习方法,如 DeepWalk^[14]、LINE^[15]或 Node2vec^[16]算法得出;对于用户属性特征,可通过词向量方法进行构建;用户节点的类别集合为 $C = \{c_1, c_2, \dots, c_n\}$,其中 n 为节点从属的类别个数,网络用户特征表示的构建过程如图 2 所示.

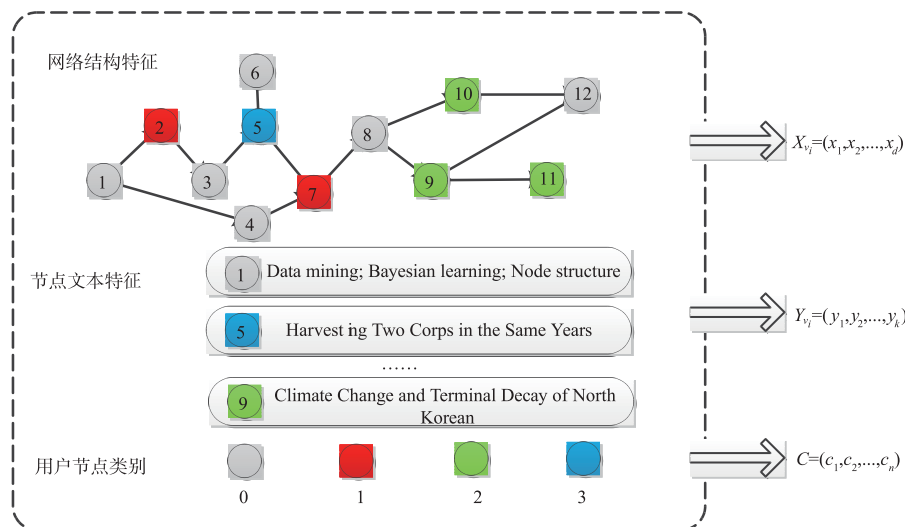


图2 文献网络中特征向量构建示意图

3 基于证据理论的网络表示融合方法

3.1 模型框架

本文提出的基于证据理论的网络表示融合方法的模型框架如图 3 所示,将用户 v_i 网络结构特征表示为 $X_{v_i} = (x_1, x_2, \dots, x_d)$,用户 v_i 属性特征表示为 $Y_{v_i} = (y_1, y_2, \dots, y_k)$;用户节点的类别集合为 $C = \{c_1, c_2, \dots, c_n\}$,可视为证据理论中的识别框架 $\Theta = \{c_1, c_2, \dots, c_n\}$,其中 n 为网络节点的类别数.本方法首先根据各特征的输入数据构造基本概率赋值函数,然后根据用户的网络结构特征,计算得到用户类别的后验概率分布为 $P(X_{v_i}) = \{m_x(c_1), m_x(c_2), \dots, m_x(c_n)\}$,其中 $m_x(c_i)$ 为根据特征 X 计算得到的用户 v_i 属于类别 c_i 的概率;根据用户属性特征表示计算得到的用户类别后验概率分布为 $P(Y_{v_i}) = \{m_y(c_1), m_y(c_2), \dots, m_y(c_n)\}$,其中 $m_y(c_i)$ 为根据特征 Y 计算得到的用户 v_i 属于类别 c_i 的概率.最后根据上述各类别的概率分配函数,定义融合指数 F 如下:

$$F = \sum_{i=1}^n m_x(c_i) \cdot m_y(c_i) \quad (4)$$

其中, $C_1 \cap C_2 \cdots \cap C_n = \phi$, $1 - F$ 反映了单个用户分类中两类特征 X, Y 的冲突程度;融合指数值越高,表明两类特征对用户归属同一类别的支持度越高,反之则表明两类特征对用户归属同一类别的支持度低.用户特征

融合后归属类别 c_i 的概率为:

$$\begin{aligned} (m_x \oplus m_y)(\{c_i\}) &= \frac{1}{F} \sum_{B \cap C = \{c_i\}} m_x(B) \cdot m_y(C) \\ &= \frac{1}{F} \cdot m_x(c_i) \cdot m_y(c_i) \end{aligned} \quad (5)$$

当融合的特征类别由 2 类推广到 Z 类 ($Z \geq 2$) 时,融合指数的计算方法及融合后类别归属概率如下,其中 m_j 代表第 j 种类型的用户特征:

$$\begin{aligned} F &= \sum_{i=1}^n \prod_{j=1}^Z m_j(c_i) (m_1 \oplus m_2 \oplus \cdots \oplus m_Z)(\{c_i\}) \\ &= \frac{1}{F} \sum_{A_1 \cap A_2 \cap \cdots \cap A_Z = \{c_i\}} m_1(A_1) \cdot m_2(A_2) \cdots m_Z(A_Z) \end{aligned} \quad (6)$$

由此,根据融合指数大小设置一定阈值,即可找出网络多源信息融合中的高冲突节点,判定引入额外信息对用户分类效果是提升还是降低.本文结合支持向量机(Support Vector Machine, SVM)方法与 D-S 证据理论^[17]进行决策融合实验,计算不同属性特征融合后的用户分类结果.标准 SVM 算法输出的是测试样本的类别标签,而本文基于证据理论进行网络表示融合中,需要计算用户归属不同类别的后验概率值得到基本可信度函数(mass 函数),下节将介绍如何将 SVM 输出转换为后验概率值的输出.

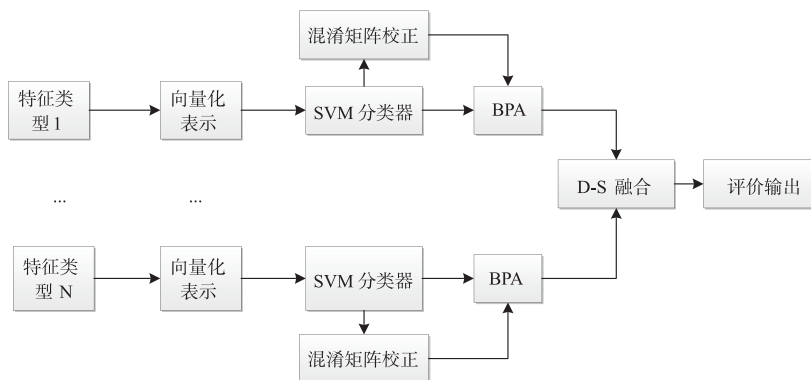


图3 模型框架示意图

3.2 非线性 SVM 的后验概率计算

将标准 SVM 输出的类别标签映射为后验概率输出,Platt 等人^[18]在二分类 SVM 中给出了相关计算方法,SVM 的标准输出如下式所示:

$$y = \text{sign}(f(x)) = \text{sign}(\alpha_i^* y_i (x \cdot x_i) + b^*) \quad (7)$$

针对二分类的情况,应用 Sigmoid 函数将 SVM 的输出值映射到 $[0, 1]$ 之间:

$$\text{Pr}(y = 1 | x) \approx P_{A,B}(f(x)) = \frac{1}{1 + \exp(Af(x) + B)} \quad (8)$$

其中 $f(x)$ 为标准的 SVM 输出结果, $\text{Pr}(y = 1 | x)$ 是在输出值为 $f(x)$ 的条件下分类正确的概率. A 和 B 是参数值,可通过计算参数集的最小负对数似然值来求解.

$$\min_{z=(A,B)} F(z) = - \sum_{i=1}^l (t_i \log(p_i) + (1 - t_i) \log(1 - p_i)) \quad (9)$$

其中 p_i, t_i 值如下:

$$p_i = P_{A,B}(f_i) = p(y_i = 1 | x_i) \quad (10)$$

$$t_i = \begin{cases} \frac{N_+ + 1}{N_+ + 2}, & \text{if } y_i = +1 \\ \frac{1}{N_- + 2}, & \text{if } y_i = -1 \end{cases}, i = 1, \dots, l \quad (11)$$

其中 N_+ 是 $y_i = +1$ 的样本数量, N_- 是 $y_i = -1$ 的样本数量. 对于本文的多分类问题,利用多个二分类器组合投票的方式可以近似计算该分类的后验概率输出.

$$P(i | x) = \frac{\sum_{j=1, j \neq i}^n p_{ij}(i | j; x)}{\sum_{k=1, k \neq i}^n \sum_{j=1, j \neq k}^n p_{kj}(k | j; x)} = \frac{2}{n(n-1)} \sum_{j=1, j \neq i}^n p_{ij}(i | j; x) \quad (12)$$

其中, $p_{ij}(i | j; x)$ 表示由第 i 类和第 j 类组合的二分类计算得到的样本属于第 i 类的后验概率.

3.3 基于混淆矩阵的后验概率校正

由于用户在不同社交网络中填写个人属性信息或

建立通联关系中存在差异^[19],并且不同特征对于不同类别目标的表征能力也不同,因此通过不同特征获得的类别预测需要综合考虑错判为其他类的概率,同时估计到分类器对各个目标类别的识别可靠性,本文采用混淆矩阵来估计各特征的局部识别可信度,利用混淆矩阵对分类器的局部可信度进行描述,在构造分类器的基本概率分配时进行加权校正,理论上可以使得网络表示融合效果更优. 假设用户类别数为 k ,待识别数据集中共有 N 个样本,每个类别分别含有 N_i ($i = 1, 2, \dots, k$) 个样本,分类后的混淆矩阵 C 一般表示为:

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1k} \\ c_{21} & c_{22} & \dots & c_{2k} \\ \vdots & \vdots & c_{ij} & \vdots \\ c_{k1} & c_{k2} & \dots & c_{kk} \end{bmatrix} \quad (13)$$

其中, c_{ij} 表示 C_i 类被分类器判断成 C_j 类的样本数量. 混淆矩阵中元素的行下标对应目标的真实类别,列下标对应分类器输出的类别,对角元素表示各类别能够被分类器正确识别的样本数量,非对角元素表示发生错误判断的样本数量. 混淆矩阵的行向量代表了类别 C_i 的对象在进行分类时对各类别的倾向性. 针对当前网络表示的分类问题,根据混淆矩阵可知,当分类器输出类别 C_j 时,当前样本的真实类别 C_i 的概率为:

$$PC(C_i | C_j) = \frac{c_{ij}}{\sum_{j=1}^k c_{ij}} \quad (14)$$

将 $PC(C_i | C_j)$ 记作 $PC(C_i)$,则 $PC(C_i)$ 可视为当前目标 x 属于 C_i 类的支持度. 由此,将 $PC(C_i)$ 定义为分类器关于类别 i 的局部可信度. 当对样本 x 输出一个真实类别的判决结果 C_i 时,我们对这个结果的可信度为 $PC(C_i)$,则分类器给出的 x 属于 C_i 类的基本概率赋值 $m(C_i)$ 为:

$$m(C_i) = P_i \times PC(C_i) \quad (15)$$

其中 P_i 为 SVM 输出的属于 C_i 类的后验概率, $PC(C_i)$ 为由混淆矩阵提供的局部可信度信息. 对于每个分类器经上式加权融合校正后得到的基本概率赋值, 再利用证据理论的组合规则融合得到最终的识别结果.

4 实验与结果

4.1 数据集及对比算法

本文利用文献[13]中给出的 CiteSeer-M10、Cora、DBLP 数据集测试本文提出的方法, CiteSeer-M10 数据集包含 10,310 篇论文, 77,218 条连边, 10 个类别标签. Cora 数据集包含 2,708 篇论文, 5,429 条连边, 7 个类别标签. DBLP 数据集包含 60,744 篇论文, 52,890 条边, 4 个类别标签. 对比算法包括 Doc2vec^[2] 算法: 使用神经网络模型将文本内容信息转化为分布式向量. DeepWalk^[14] 算法: 使用神经网络模型将网络结构信息转化为网络表示向量. D2V + DW 算法: 本文采用将 Doc2vec 和 DeepWalk 算法生成的表示向量进行拼接的方法.

表 1 DBLP 数据集上的节点分类准确率

λ	Doc2vec	Deepwalk	D2V + DW	D-S fusion	D-S fusion +	Improvement	Conflict
0.2	0.7276	0.5338	0.7575	0.7582	0.7712	1.8%	0.5242
0.4	0.7343	0.5398	0.7656	0.7660	0.7803	1.9%	0.5303
0.6	0.7354	0.5380	0.7691	0.7662	0.7872	2.4%	0.5289
0.8	0.7425	0.5348	0.7768	0.7729	0.7954	2.4%	0.5291

表 2 M10 数据集上的节点分类准确率

λ	Doc2vec	Deepwalk	D2V + DW	D-S fusion	D-S fusion +	Improvement	Conflict
0.2	0.5687	0.4879	0.6336	0.6721	0.6784	7.1%	0.3791
0.4	0.5944	0.5284	0.6783	0.7084	0.7121	5.0%	0.4066
0.6	0.6028	0.5371	0.7017	0.7277	0.7297	4.0%	0.4095
0.8	0.6149	0.5712	0.7347	0.7546	0.7602	3.5%	0.4408

表 3 Cora 数据集上的节点分类准确率

λ	Doc2vec	Deepwalk	D2V + DW	D-S fusion	D-S fusion +	Improvement	Conflict
0.2	0.6602	0.6269	0.7444	0.7682	0.7702	3.5%	0.6003
0.4	0.7565	0.7001	0.8085	0.8116	0.8134	1.0%	0.6307
0.6	0.7615	0.7242	0.8203	0.8338	0.8421	2.7%	0.6440
0.8	0.7832	0.7223	0.8307	0.8374	0.8456	1.8%	0.6523

实验 1 研究不同训练比例 λ 下各算法的分类性能. 在 DBLP 数据集上进行多标签分类实验的仿真结果如表 1 所示, 由表可知, 多特征信息融合方法 (D2V + DW \ D-S fusion \ D-S fusion +) 的准确度优于仅利用任意单一特征 (Doc2vec \ DeepWalk) 的分类准确度, 证明了特征融合的优势与必要性; 在算法复杂度上, D2V + DW 采用先将特征向量进行拼接然后分类器分类的方式, 而 D-S fusion 和 D-S fusion + 方法先利用各个特征向量计算类别归属, 然后对分类结果进行融合判别, 三种方法的算法复杂度接近 (在同一量级), 实际运算时间上

4.2 实验与结果

对于网络表示融合中的结果评价问题, 本实验中采用的评价指标为准确率、召回率和 F_1 值, 并设计了 3 类实验验证提出的网络表示融合方法的可行性和有效性. 网络表示融合实验的参数设置如下: 采用的表示向量维度为 100, 实验结果在 10 次实验中取均值; DeepWalk 算法中随机游走长度为 8, 游走次数为 20 次; Doc2vec 算法中参与训练的最小词频为 3, 滑动窗口大小为 5. 实验结果如表 1、2、3 所示, 表中分别展示了 Doc2vec 算法、Deepwalk 算法、D2V + DW 算法的分类准确率, 其中 D-S fusion 表示不添加基于混淆矩阵修正的决策融合方法准确率, D-S fusion + 表示添加了基于混淆矩阵修正的决策融合方法准确率; Conflict 值表示应用 Doc2vec 算法与 DeepWalk 算法在节点分类结果上不一致的比例; Improvement 值表示 D-S fusion + 算法相比于 D2V + DW 算法的准确率提升比率.

D-S fusion + 方法略高, 但 D-S fusion 和 D-S fusion + 方法对数据噪声的鲁棒性更强, 分类准确率更高. 同时从实验结果看出: (1) 属性特征较网络结构特征在用户分类中的效果更显著, 分类精度更高; (2) 通过添加混淆矩阵的校正, 可以使得 D-S 融合方法在 DBLP 数据集上的分类准确率有进一步的提升. 通过进一步分析实验中混淆矩阵可知, 该数据集下错分类别较为集中, 因此根据混淆矩阵设置不同的局部可信度用于纠偏可以取得一定的效果. 在 M10、Cora 数据集上进行多标签分类实验的结果如表 2、表 3 所示. 从结果可知: (1) 该数据

集上属性特征与网络结构特征在用户分类中的效果相差不大,原因是 M10 和 Cora 网络中连边数量更多,结构更加紧密,网络结构特征的效果更突出;(2)基于 D-S 证据理论融合的方法优于将表示向量进行拼接的方法,同时拼接方法的计算复杂度更高.此外,添加混淆矩阵校正的方法,在此两类数据集上效果并不明显,因为该混淆矩阵中错误类别分布较为均匀,添加修正对最终结果的影响不大,该方法对于错误类别中有明显错分某类偏好的数据集会有更好的效果.

实验 2 中分析了特征冲突与分类准确率的关系曲线.如图 4 所示,随着训练样本数量的增加,冲突比例(Conflict 值)有一定提升.从各算法的准确率曲线可知,由于数据中噪声干扰的存在,识别错误的样例存在交叉,即每种特征都有自己擅长区分的样例,关键在于选择最适合的特征进行分类,特征融合后的分类准确率有一定提升.本文方法根据混淆矩阵对后验概率进行校正,确定容易错分的类别,从而设置不同的局部可信度权值用于纠偏,可以起到一定的提升效果.实验 3 分析节点融合指数 F 与融合分类结果的关系,即统计数据集中各节点融合指数 F 值对应下的分类结果(正确或错误)准确率.从图 5 可以看出,在一定范围内,随着融合指数 F 的增大,分类正确的比例逐渐上升,融合指数与融合结果中分类正确的比例呈现一

定的正相关关系.同时,在分类实验中通过去除部分高冲突节点后,也能获得分类准确率的一定提升.在实际应用中,可以对此类数据噪声干扰下的冲突难分节点进行人工判别,为后续应用任务做好预处理工作.

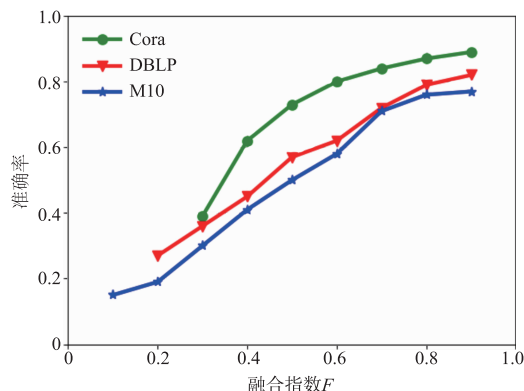


图5 融合指数 F 与融合分类结果关系

5 结束语

在引入外部信息增强网络表示学习算法效果的研究基础上,本文重点分析了多源网络表示信息融合中的特征冲突问题,提出一种基于证据理论的网络表示融合方法,有效提升了节点分类任务中的准确率.多源信息融合方法作为有效提升网络表示学习性能的重要途径,在处理新实体的稀疏表示问题上具有一定优势,但目前仍处于起步阶段,后续仍具有广阔的研究前景.

参考文献

- [1] 齐金山,梁循,李志宇.大规模复杂信息网络表示学习:概念、方法与挑战[J].计算机学报,2018,41(10):222-248.
Qi Jin-shan,Liang Xun, Li Zhi-yu. Representation learning of large-scale complex information network: concepts, methods and challenges [J]. Chinese Journal of Computers, 2018, 41(10): 222-248. (in Chinese)
- [2] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[A]. Proceedings of the Advances in Neural Information Processing Systems[C]. Nevada; NIPS, 2013. 3111-3119.
- [3] Shi C, et al. Heterogeneous information network embedding for recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(2): 357-370.
- [4] 温雯,等.一种融合节点先验信息的图表示学习方法[J].软件学报,2018,29(3):786-798.
Wen Wen, et al. Graph embedding by incorporating prior knowledge on vertex information[J]. Journal of Software, 2018, 29(3): 786-798. (in Chinese)
- [5] Yang C, Liu Z, Zhao D, et al. Network representation learn-

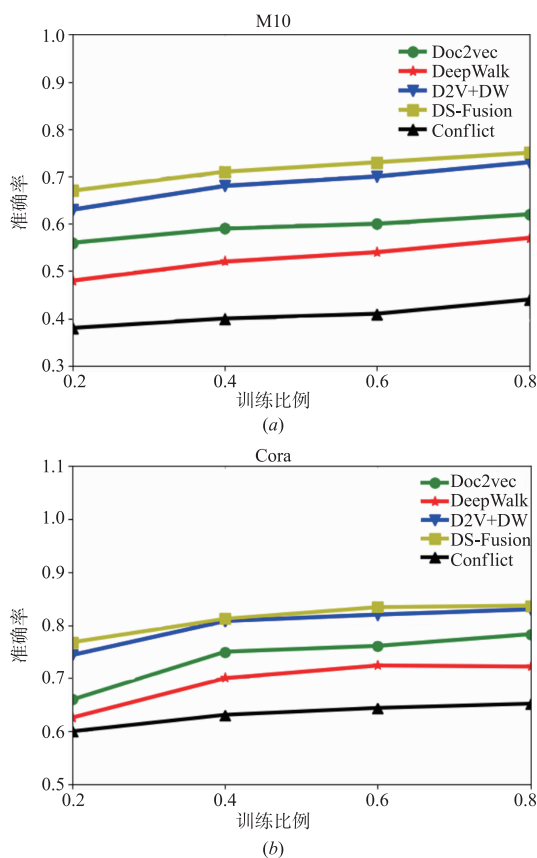
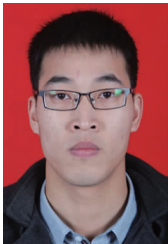


图4 不同算法的分类准确率和冲突比较

- ing with rich text information [A]. Proceedings of the International Joint Conference on Artificial Intelligence [C]. Argentina; AAAI, 2015. 2111 – 2117.
- [6] Tu C, Liu H, Liu Z, et al. CANE: Context-aware network embedding for relation modeling [A]. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics [C]. Vancouver; ACL, 2017. 1722 – 1731.
- [7] Chen J, et al. Incorporate group information to enhance network embedding [A]. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management [C]. Indianapolis; ACM, 2016. 1901 – 1904.
- [8] Cavallari S, et al. Learning community embedding with community detection and node embedding on graphs [A]. Proceedings of the Conference on Information and Knowledge Management [C]. Singapore; ACM, 2017. 377 – 386.
- [9] 林海伦, 等. 面向网络大数据的知识融合方法综述 [J]. 计算机学报, 2017, 40(1): 1 – 27.
Lin Hai-lun, et al. Network big data oriented knowledge fusion methods: a survey [J]. Chinese Journal of Computers, 2017, 40(1): 1 – 27. (in Chinese)
- [10] Wang X, Song Y. Uncertainty measure in evidence theory with its applications [J]. Applied Intelligence, 2018, 48(7): 1672 – 1688.
- [11] 涂存超, 杨成, 刘知远. 网络表示学习综述 [J]. 中国科学: 信息科学, 2017, 47(8): 980 – 996.
Tu Cun-chao, Yang Cheng, Liu Zhi-yuan. Network representation learning: an overview [J]. Scientia Sinica Informationis, 2017, 47(8): 980 – 996. (in Chinese)
- [12] 李艳秋, 任福继, 胡敏. 动态模糊密度的多分类器融合算法 [J]. 电子学报, 2018, 46(5): 1246 – 1252.
LI Yan-qi, REN Fu-ji, HU Min. Dynamic fuzzy density for multi-classifier fusion algorithm [J]. Acta Electronica Sinica, 2018, 46(5): 1246 – 1252. (in Chinese)
- [13] Pan S, Wu J, Zhu X. Tri-party deep network representation [A]. Proceedings of the International Joint Conference on Artificial Intelligence [C]. New York; AAAI, 2016. 1895 – 1901.
- [14] Perozzi B, et al. Deepwalk: Online learning of social representations [A]. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. New York; ACM, 2014. 701 – 710.
- [15] Tang J, Qu M, Wang M, et al. Line: large-scale information network embedding [A]. Proceedings of the International Conference on World Wide Web [C]. Florence, Italy; WWW, 2015. 1067 – 1077.
- [16] Grover A, Leskovec J. node2vec: Scalable feature learning for networks [A]. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. San Francisco, USA; ACM, 2016. 855 – 864.
- [17] 雷蕾, 等. 结合 SVM 与 DS 证据理论的信息融合分类方法 [J]. 计算机工程与应用, 2013, 49(11): 114 – 117.
Lei Lei, et al. Approach of information fusion and classification by SVM and DS evidence theory [J]. Computer Engineering and Applications, 2013, 49(11): 114 – 117. (in Chinese)
- [18] Platt J C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods [J]. Advances in Large Margin Classifiers, 1999, 10(4): 61 – 74.
- [19] 陈鸿昶, 等. 一种基于用于轨迹的跨社交网络用户身份识别 [J]. 电子与信息学报, 2018, 40(11): 2758 – 2764.
Chen Hong-chang, et al. User identification across social networks based on user trajectory [J]. Journal of Electronics and Information Technology, 2018, 40(11): 2758 – 2764. (in Chinese)

作者简介



程晓涛 (通信作者) 男, 1990 年出生, 河北邢台人, 现为战略支援部队信息工程大学博士研究生, 研究方向为网络大数据处理与分析、社交媒体挖掘。
E-mail: chengxt90@gmail.com



吉立新 男, 1970 年出生, 江苏淮安人, 现为战略支援部队信息工程大学研究员, 博士生导师, 研究方向为电信网信息关防。



尹赢 女, 1994 年出生, 四川绵竹人, 现为战略支援部队信息工程大学硕士研究生, 研究方向为网络大数据处理与分析、复杂网络, 网络表示学习。



黄瑞阳 男, 1986 年出生, 福建漳州人, 现为战略支援部队信息工程大学副研究员, 研究方向为网络大数据处理与分析。