

基于多模态融合技术的用户画像方法

张壮¹ 冯小年² 钱铁云^{1,†}

1. 武汉大学计算机学院, 武汉 430072; 2. 中国电力财务有限公司, 北京 100005; † 通信作者, E-mail: qty@whu.edu.cn

摘要 针对当前用户画像工作中各模态信息不能被充分利用的问题, 提出一种跨模态学习思想, 设计一种基于多模态融合的用户画像模型。首先利用Stacking集成方法, 融合多种跨模态学习联合表示网络, 对相应的模型组合进行学习, 然后引入注意力机制, 使得模型能够学习不同模态的表示对预测结果的贡献差异性。改进后的模型具有精心设计的网络结构和目标函数, 能够生成一个由特征级融合和决策级融合组成的联合特征表示, 从而可以合并不同模态的相关特征。在真实数据集上的实验结果表明, 所提模型优于当前最好的基线方法。

关键词 用户画像; 模型组合; stacking; 跨模态学习联合表示; 多层多级模型融合

User Profiling Based on Multimodal Fusion Technology

ZHANG Zhuang¹, FENG Xiaonian², QIAN Tiejun^{1,†}

1. School of Computer Science, Wuhan University, Wuhan 430072; 2. China Power Finance Co., Ltd, Beijing 100005;
† Corresponding author, E-mail: qty@whu.edu.cn

Abstract Existing studies in user profiling are unable to fully utilize the multimodal information. This paper presents a cross-modal joint representation learning network, and develop a multi-modal fusion model. Firstly, a stacking method is adopted to learn the joint representation network which fuse the cross-modal information. Then, attention mechanism is introduced to automatically learn the contribution of different modal to the prediction task. Proposed model has a well defined loss function and network structure, which enables combining the related features in various models by learning the joint representations after feature-level and decision-level fusion. The extensive experiments on real data sets show that proposed model outperforms the baselines.

Key words user profiling; model combination; stacking; cross-modal learning joint representation; multi-layer and multi-level model fusion

近年来, 社交媒体用户画像研究吸引越来越多研究人员的关注。在现有用户画像建模技术中, 对如何融合多个用户数据源或模态, 以便获得更精确的用户画像的研究相当有限, 并且存在如下不足: 一方面, 部分用户画像研究工作仅就单一的模态进行研究, 难以全面地刻画用户的特征; 另一方面, 大多数利用多模态的研究仅在特征级或决策级集成数据源^[1-3], 即使有的研究能够进行两个级别的融合^[4-5], 但仍然缺乏对多个数据源深度融合的探索。

本文提出一种基于多模态融合的用户画像模型UMF-SCF(UMF表示user profiling based on multi-

modal fusion, SCF表示stacking, model combination, and fusion), 用于解决用户画像建模中各模态不能深入交互的问题。

1 UMF-SCF模型

1.1 模型总体架构

本文任务为用户画像, 主要对用户性别和年龄属性进行分类。用户特征集 D 包含两个部分: 有标签用户特征集 D^A (即训练集)和无标签用户特征集 D^U (即测试集), $D=D^A \cup D^U$ 。用户性别和年龄的分类即是给定一系列训练集用户 $D^A \in D$ 进行训练, 继而为

无标签用户 $D^U \in D$ 打标签的过程, 标签集为 C 。用户的信息包括用户的特征词、关注关系抽取的链接、用户生成的文本以及微博数量、发布时间等统计信息。

我们将一个用户的所有微博经分词后合并, 作为该用户的文本信息, 通过计算 tf-idf 值和卡方检验的方法得到用户的特征词嵌入, 通过网络嵌入方法 LINE 获得关系网络中用户节点的嵌入。使用 word2vec 训练得到词向量, 并采用 tf-idf 对用户词向量加权, 得到用户的文本嵌入。最后, 手工提取用户发布微博、评论、转发、关注者、发微博时间和表情符等统计特征, 利用 LIWC 心理学词典提取用户的心理学特征, 将它们作为手工特征嵌入。

将数据集中的所有用户用一个用户特征集 $D=(F, L, T, S)$ 来表示, 特征集的每个用户拥有特征词 F 、文本特征 T 、手工特征 S 和连接用户彼此的链接特征 L 共 4 种模态的特征。为了包含数据源共享与不共享的表示, 本文使用模型组合的策略, 并针对特定的模型组合构建相应的跨模态学习联合表示网络。模型可以分为 (F, L, T, S) 、 (F, L, T) 、 (F, L, S) 、 (F, T, S) 、 (L, T, S) 、 (F, L) 、 (F, T) 、 (F, S) 、 (L, T) 、 (L, S) 和 (T, S) 11 种组合。其中, (F, L, T, S) 对应的跨模态学习联合表示网络称为 FusionFLTS, 其他组合依此类推。

本文提出的多层多级融合模型 UMF-SCF 中的“多层”, 一方面指利用 stacking 产生的双层融合模型, 另一方面, stacking 第一层使用的跨模态学习联合表示网络包含嵌入层、交互层和决策层, 模型级融合发生在交互层; “多级”指融合的模式中同时包含特征级融合和决策级融合。

UMF-SCF 模型的第一层为特征级融合层, 运用跨模态学习联合表示网络对模态组合进行学习, 从而生成各模态联合或共享的特征表示。该层将数据源与非线性函数相结合, 使得不同模态的特征产生深入交互, 有助于各模态融合彼此的互补信息, 并使用注意力机制对提取的特征使用模态级的加权评分, 用以凸显各模态特征对不同任务目标的重要程度不同。UMF-SCF 模型的第二层为决策级融合层, 将第一层产生的预测概率输入神经网络, 从而获得模型最终的预测结果。

本文基于 Farnadi 等^[6]的幂集组合方法, 对给定 k 个数据源的集合 $DS=\{D_1, D_2, \dots, D_k\}$, 去掉空集和单个模态元素的子集(不会产生模态交互), 得到 DS

子集的个数为 2^k-k-1 。为了对各数据源之间的相关性建模, 本文对子集内的所有模态组合建立相应的跨模态学习联合表示网络, 用以进行特征级融合; 将所有跨模态学习联合表示网络的预测结果与 ensemble 方法相结合, 进行决策级融合。因此, 本文的 UMF-SCF 模型是一种混合的数据集成模型。

UMF-SCF 的整体架构如图 1 所示, 首先学习用户的 4 种数据源嵌入, 然后将 4 种数据源嵌入进行模型组合, 产生 11 种组合形式。将所有模型组合输入多模态融合模型中, 每一种模型组合使用相应的跨模态学习联合表示网络来学习, 输出类别的预测概率。将预测概率拼接, 得到 stacking 第 2 层的训练样本和测试样本, 最后输入一个神经网络中做分类。

1.2 FusionFLTS 模型

本文基于跨模态学习联合表示模型 CrossTL^[7], 将原始 CrossTL 模型用于学习文本与链接之间联合表示。CrossTL 采用单层的交互模式, 对模态的融合比较有限。为了达到更深层次的交互, 本文对 CrossTL 模型进行改进, 将其设计为双层的交叉融合结构。因此, 第一交互层的模态表示包含其他模态原始的信息, 第二交互层的模态表示不仅包含其他模态第一交互层的信息, 还包含原始的信息。本文简化了 CrossTL 的模型参数, 并扩展到学习 4 个模态的联合表示 FusionFLTS 中。

1.2.1 FusionFLTS 模型架构

为了使多模态间产生深层交互, 以便学习彼此的互补信息, 本文设计一个 3 层跨模态学习联合表示网络 FusionFLTS, 其结构如图 2 所示。简洁起见, 图中仅展示网络中的特征词 F 、链接特征 L 和文本特征 T , 而略去手工特征 S 。事实上, 这就是 FusionFLT 网络的实际架构, 此处不严格区分。FusionFLT 的 3 层结构为嵌入层、交互层和决策层。为使特征交互更为充分, 我们将交互层设计为双层结构, 由于两层的交互算法一致, 图 2 仅展示单交互层结构。其他模态组合的跨模态学习联合表示网络(如 FusionFL 和 FusionLTS 等)均与 FusionFLT 类似, 通过替换模态组合即可获得。

1) 嵌入层: 对每个输入进行预训练, 得到不同模态的嵌入是有效的方法。本文嵌入层包含用户特征词的嵌入 $F_1(a)$ 、链接源的图形成嵌入 $L_1(a)$ 、文本源的文本嵌入 $T_1(a)$ 和手工特征嵌入 $S_1(a)$, 简写为 F_1, L_1, T_1 和 S_1 。

2) 交互层: 首先使用隐藏层将各模态表示进行

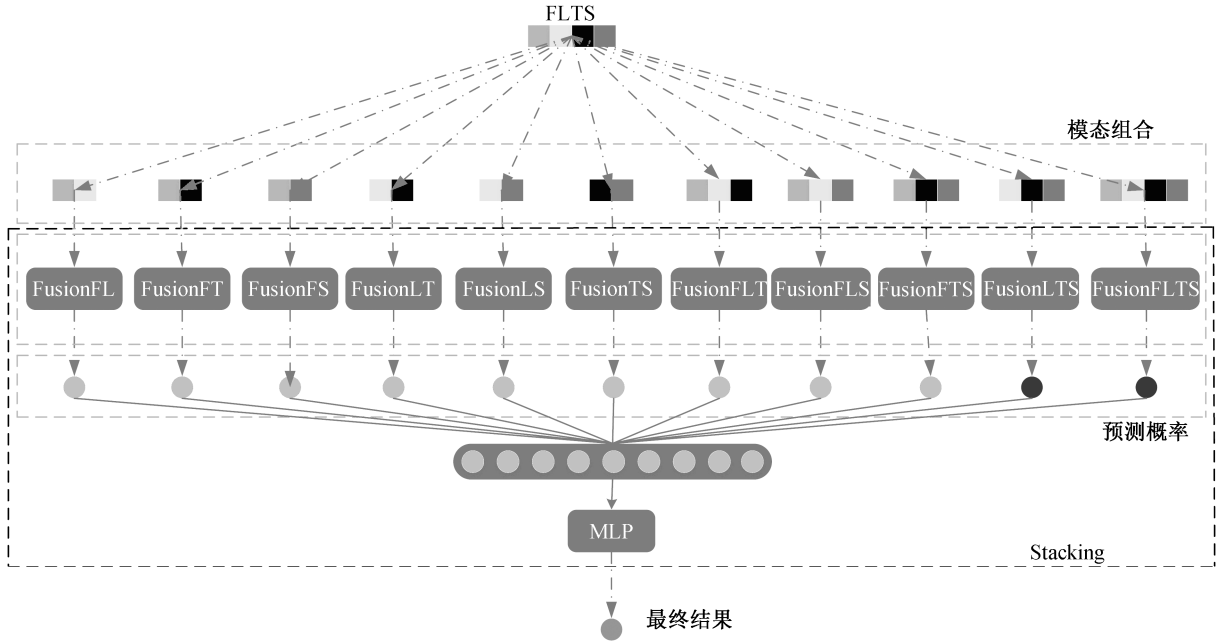


图1 UMF-SCF 整体架构

Fig. 1 Overall architecture diagram of UMF-SCF

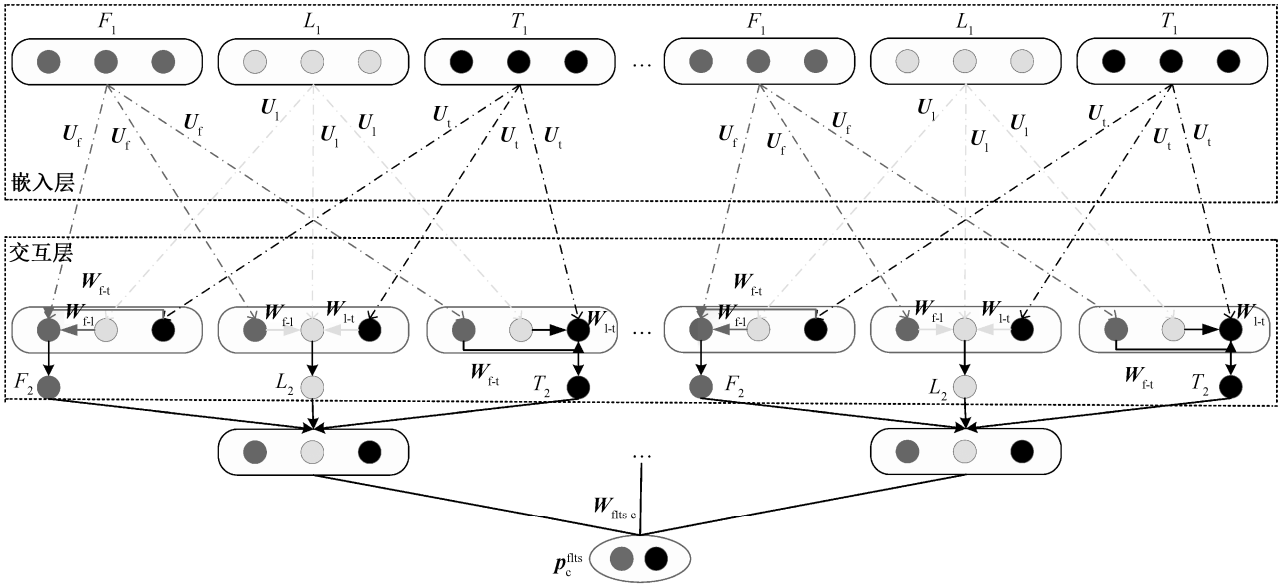


图2 FusionFLT 网络架构

Fig. 2 Network architecture diagram of FusionFLT

转换, 然后针对每个已经转换过的模态表示, 补充与其他转换过的模态嵌入的关联表示, 由此, 得到各模态经过交互层之后的特征表示, 其中包含与其他模态对应的交互信息:

$$\begin{aligned} M_b^1 = & \tanh(U_{m^1} M_a^1 + W_{m^1 m^2} U_{m^2} M_a^2 + \\ & W_{m^1 m^3} U_{m^3} M_a^3 + W_{m^1 m^4} U_{m^4} M_a^4), \end{aligned} \quad (1)$$

其中, M^* 分别表示 4 种模态嵌入 F, L, T 和 S ; a 为交互层的层数, 若 $a=1$, 则 $b=2$, 若 $a=2$, 则 $b=3$; \tanh 是神经网络激活函数; U_* 是第 a 层交互层隐藏层中相应模态嵌入对应的转换矩阵; W_{**} 是第 a 层交互层中相应两模态嵌入的关联权重矩阵。

3) 决策层: 经过交互层后, 每个用户都可以得到 5 个高层表示: F_3, L_3, T_3, S_3 和 $F_3 \oplus L_3 \oplus T_3 \oplus$

S_3 。这些表示都包含来自特征词、链接、文本和手工特征的融合信息,但每种表示仍然有独特的聚焦点。例如, T_3 主要是用户级文本表示,同时包含附加的特征词、链接和手工特征信息。其他表示同理。为了将上述 5 种跨模态学习到的高层表示应用于目标任务中,本文使用一个线性层,将 5 个表示映射到其标签类别空间,分别是 C^f , C^l , C^t , C^s 和 C^{flts} ,然后使用一个 softmax 层来获得类别 c 的概率分布。

线性映射定义如式(2)所示:

$$C^* = W_{*-c}A_3 + b_{*-c}, \quad (2)$$

其中, $A \in \{F_3, L_3, T_3, S_3, F_3 \oplus L_3 \oplus T_3 \oplus S_3\}$, C^* 表示 A 对应的标签类别空间, W_{*-c} 表示 A 对应的线性层权重矩阵, b_{*-c} 是相应的偏置值。

softmax 层定义如式(3)所示:

$$p_c^* = \frac{\exp(C_c^*)}{\sum_{m=1}^M \exp(C_m^*)}, \quad (3)$$

其中, p_c^* 表示 F_3, L_3, T_3, S_3 和 $F_3 \oplus L_3 \oplus T_3 \oplus S_3$ 预测类别 c 的概率, M 是类别个数, C_c^* 表示标签类别空间的第 c 个类别。

1.2.2 FusionFLTS 目标函数

为了充分利用 FusionFLTS 学到的表示,我们设计了由多种损失组成的目标函数:各模态表示的损失 J_{mod} 、联合表示的损失 J_{flts} 和模态表示间的一致性损失 J_{dif} 。其中, J_{mod} 和 J_{flts} 用来衡量不同模态组合各自特征的优良程度, J_{flts} 则促使不同模态组合的预测结果尽量接近。

为了方便公式的表达我们使用 $P_{\text{mod}} = \{p_c^f, p_c^l, p_c^t, p_c^s\}$ 表示计算 J_{mod} 的预测概率的集合,使用 $P_{\text{dif}} = \{(p_c^f, p_c^l), (p_c^f, p_c^t), (p_c^f, p_c^s), (p_c^l, p_c^t), (p_c^l, p_c^s), (p_c^t, p_c^s)\}$ 表示计算 J_{dif} 的预测概率的集合。使用交叉熵作为损失,则 3 个部分的损失函数可以用式(4)~(6)定义:

$$J_{\text{mod}} = - \sum_{p \in P_{\text{mod}}} \sum_{a \in D^4} \sum_{c=1}^C p_c^g(a) \cdot \log(p(a)), \quad (4)$$

$$J_{\text{flts}} = - \sum_{a \in D^4} \sum_{c=1}^C p_c^g(a) \cdot \log(p_c^{\text{flts}}(F_3(a) \oplus L_3(a) \oplus T_3(a) \oplus S_3(a))), \quad (5)$$

$$J_{\text{dif}} = - \sum_{(p_1, p_2) \in P_{\text{dif}}} \sum_{a \in D^4} \sum_{c=1}^C p_1(a) \cdot \log(p_2(a)), \quad (6)$$

其中, a 是有标签数据 D^4 中的一个用户, p_c^g 是类别 c 的真实标签。

如果简单地将上述损失相加来构成最终的损失,则不能揭示不同模态的实际重要性,原因在于相同的模态对不同的用户属性分类任务可能有不同的贡献。例如,通过对用户微博文本数据进行分析,发现手工提取的各种统计类特征对于预测用户的年龄更有利(例如,不同年龄段的用户发微博的时间、活跃程度以及使用的表情符等都具有分辨性),而文本内容对用户性别的分类更重要。即使是同一任务,不同的模态也可能对预测的结果产生不同的影响。因此,本文引入注意力机制,即引入权重系数列表 w_s , 总体损失是以上损失函数的线性加权合并,如式(7)所示:

$$J_{\text{FLTS}} = \sum_{i=0}^3 w_s[i] \cdot \text{List}_J^{\text{mod}}[i] + w_s[4] \cdot J_{\text{flts}} + \sum_{j=5}^{10} w_s[j] \cdot \text{List}_J^{\text{dif}}[j-5] + w_s[11] \cdot \text{reg_loss}, \quad (7)$$

其中, reg_loss 为 L2 正则化损失, $\text{List}_J^{\text{mod}}$ 表示 J_{mod} 各项损失的列表, $\text{List}_J^{\text{dif}}$ 表示 J_{dif} 各项损失的列表。用权重系数列表 w_s 来平衡各模态表示、它们的联合表示以及一致性。如果将 w_s 中的权重均设为 1, 则代表 3 个损失函数的简单相加。应用于 3 个和两个模态的跨模态学习联合表示模型(FusionFLT 和 FusionFL)的算法同理。

2 实验及分析

2.1 实验设置

为了评测本文提出的 UMF-SCF 模型的性能,我们在 SMP CUP 2016 微博用户画像数据集上对模型进行训练和测试。由于数据集微博文本数据包含很多对推测用户属性无用的信息,我们对原始数据进行预处理,包括分词、去停用词、缺失值补零以及去除用户重复微博等。

本文从 4 种数据源中得到嵌入表示:用户特征词表示、关系网络中的用户节点表示、用户级文本表示和手工特征表示。

在预测用户性别和年龄的任务中,为了与 SMP CUP 2016 竞赛评估指标一致,本研究选取准确率(Accuracy)作为评测指标。还选用分类任务中常用的精确率(Precision)、召回率(Recall)和 F1 值作为评测指标。对于三分类的年龄预测任务,采用宏平均

的F1值作为评价指标。我们选取如下基线方法作为对比。

特征词 考虑到卡方检验的低频词缺陷, 本文计算得到用户微博文本和平台文本中所有词的tf-idf值, 根据该值计算每个词语与类别的卡方值, 选取微博文本中卡方值最高的5000个词和平台文本中卡方值最高的300个词作为用户文本的特征词, 使用SVD降维的方法将特征词嵌入降维至300维, 即为用户特征词表示。

LINE^[8] 用于学习用户关注关系网络的用户节点表示。

Word2Vec^[9]+TF-IDF 使用word2vec预训练的词向量作为模型中的词向量矩阵, 通过计算用户文本所有词(包含微博文本与平台文本)的tf-idf值为词向量赋予权重, 得到用户的300维文本表示。

手工特征 提取用户关注者数目、评论数目、发帖总数等统计信息以及表情符特征、用户心理学特征等, 使用卡方检验的方法提取与用户性别、年龄预测最相关的300个手工特征。

针对以上4种基线方法, 本文使用支持向量机SVM对特征进行分类。我们也尝试使用多层感知机(multilayer perceptron, MLP)分类, 效果与SVM无显著差异。

Concat+SVM 通过数据源嵌入得到各模态嵌入拼接, 然后输入支持向量机分类器进行分类。

Concat+MLP 通过数据源嵌入得到各模态嵌入拼接, 然后输入多层感知机MLP进行分类。

UHAN^[10] 是一种基于用户引导的分层注意力网络(UHAN)模型, 研究涉及视觉和文本模态的非结构化特性, 并进一步考虑从流行度预测的多模态数据中学习有效的多模态表示。

UDMF^[6] 该混合用户画像框架利用模态之间的共享表示, 在特征层集成3个数据源, 并结合在决策层来操作每个数据源组合的独立网络。

SMP CUP 2016 竞赛第一名(SMP CUP 1st): 该模型也利用用户特征词特征、关系网络特征和文本特征3个模态, 但做法与本文略有不同。学习用户特征词表示时, 为了降低噪声(广告词、分享和新闻等)的影响, 每个维度用0/1表示, 而本文用tf-idf表示, 可以更精确地表征用户。得到3个模态后, 该模型对3种表示stacking进行集成, 并输出最终预测结果。

在以上的基线方法中, 对于没有使用本文数据

集的方法(如UDMF和UHAN等), 我们在本文数据集上实现, 并保持原模型中的参数设置不变。

2.2 结果分析

2.2.1 对比实验

在SMP CUP 2016微博用户画像数据集上各项基线方法准确率的实验结果见表1和2。

由表1和2可以看出, 本文的UMF-SCF模型在用户性别、年龄预测的任务上均取得领先的效果。

特征词、LINE、W2V+TF-IDF和手工特征的方法是单个模态嵌入的方法, 与Concat+MLP, UHAN, UDMF, SMP CUP 1st以及本文的模型UMF-SCF等多模态融合的方法相比, 在用户画像问题上, 多模态融合的方法明显优于单模态嵌入的方法。

LINE方法效果差的原因在于, 在SMP CUP 2016微博用户画像数据集中, 训练集和测试集中部分用户的链接表示缺失。实验中, 对于没有在关注

表1 微博用户画像数据集上各基线方法的年龄预测结果
Table 1 Age prediction results of each baseline method on the Weibo user profiling data set

方法	准确率	精确率	召回率	F1
特征词	0.622	0.569	0.449	0.457
LINE	0.574	0.191	0.333	0.243
W2V+TF-IDF	0.613	0.530	0.478	0.489
手工特征	0.620	0.536	0.426	0.417
Concat+SVM	0.606	0.530	0.514	0.522
Concat+MLP	0.645	0.597	0.526	0.559
UHAN	0.643	0.581	0.471	0.520
UDMF	0.659	0.589	0.491	0.536
SMP CUP 1st	0.648	0.609	0.483	0.539
UMF-SCF (本文)	0.681	0.676	0.521	0.589

表2 微博用户画像数据集上各基线方法的性别预测结果
Table 2 Gender prediction results of each baseline method on the Weibo user profiling data set

方法	准确率	精确率	召回率	F1
特征词	0.855	0.879	0.946	0.912
LINE	0.822	0.841	0.948	0.891
W2V+TF-IDF	0.869	0.911	0.922	0.916
手工特征	0.802	0.826	0.946	0.882
Concat+SVM	0.811	0.897	0.860	0.878
Concat+MLP	0.866	0.879	0.963	0.919
UHAN	0.856	0.903	0.916	0.909
UDMF	0.870	0.888	0.955	0.920
SMP CUP 1st	0.883	0.901	0.960	0.930
UMF-SCF (本文)	0.887	0.907	0.955	0.930

关系网络中的用户,我们采用特征补零作为用户表示的策略,在一定程度上会影响链接表示的效果。

Concat+SVM 虽然利用了多模态的信息,但是简单拼接并不能使得多个模态间产生交互,支持向量机分类的效果反而比特征词、W2V+TF-IDF 等单模态嵌入还要差,说明如果不能合理地使用多模态融合,反而会削弱单模态的效果。

Concat+MLP 虽然也是简单拼接多模态嵌入,但效果远超 Concat+SVM。由此可见,神经网络能使多种模态的数据之间产生交互,在一定程度上发挥多模态融合的优势。

UHAN 模型在本实验中效果一般。由于本文实验数据集的关系网络中约 45% 的用户存在节点缺失,所以对该模型使用引入用户向量的方法并不能奏效。

与本文的 UMF-SCF 模型相比,UDMF 模型由于只采用简单的神经网络来融合各个数据源,在多模态表示的深度交互方面有所欠缺。同时,UDMF 模型采用的多任务学习方式在本实验中并没有效果的提升。经分析,我们认为是由性别和年龄两个任务相关性较小导致。

2.2.2 模型简化学习

本文对 UMF-SCF 模型进行模型简化学习,即在去掉或替换模型的关键组件情况下,分析简化模型的效果,用以说明模型各组件的真实作用。本文考虑 UMF-SCF 模型的 4 个主要模块:跨模态学习联合表示网络、stacking 决策级集成方法、模型组合以及注意力机制,检测它们对最终预测的贡献。我们设计以下 4 个简化模型实验。

Concat+Stacking+DNN 与 UMF-SCF 模型全部架构相比,将模型中使用的跨模态学习联合表示网络替换为普通的神经网络。

Fusions+Attention 与 UMF-SCF 模型全部架构相比,去除 stacking 集成的模块,缺失第二层决策级融合。

FusionFLTS+Stacking+Attention 与 UMF-SCF 模型全部架构相比,去除 FusionFLTS 以外的其他跨模态学习联合表示网络。

除去注意力机制 将 UMF-SCF 模型的目标函数的各项损失系数全置为 1,即将各模态和各一致性损失的重要程度一视同仁。

表 3 为在 SMP CUP 2016 微博用户画像数据集上各项简化模型准确率的实验结果。可以看出,

表 3 微博用户画像数据集上各项简化模型的准确率
Table 3 Accuracy of simplified models in weibo user profiling data set

方法	年龄	性别
Concat+Stacking+DNN	0.656	0.872
Fusions+Attention	0.658	0.874
FusionFLTS+Stacking+Attention	0.665	0.878
除去注意力机制	0.666	0.881
全部架构	0.681	0.887

UMF-SCF 模型全部架构与 Concat+Stacking+DNN 对比表明,本文的跨模态学习联合表示网络比普通神经网络优越。Fusions+Attention 的效果表明,使用 stacking 集成的作用同样明显,可见第二层决策级别的数据源集成对模型效果的提高至关重要,该层的集成也是本文设计多层多级模型融合的优势所在。Fusion+Stacking+Attention 和除去注意力机制的影响相对小一些,在性别预测的任务上与 UMF-SCF 全部架构非常接近,但在年龄预测的任务上依然存在差距。可见,相对于只使用 stacking 集成一种组合方式进行分类的方法,集成多种模型组合的方法更有利于模型的性能,注意力机制亦然。UMF-SCF 模型各个组件都发挥着不可或缺的作用。

2.2.3 参数分析

正则化参数和学习率对 UMF-SCF 模型的效果有重要影响。

由于 UMF-SCF 模型采用多层多级模型融合的方法,网络参数较多,模型较为复杂,为防止模型产生过拟合现象,采用合理的正则化方法十分必要。图 3(a)给出不同正则化参数 λ 下 UMF-SCF 模型的准确率。其中,0.001/0.005 表示用于融合两个模态的跨模态学习联合表示网络使用的正则化参数是 0.001,用于融合 3 个和 4 个模态的跨模态学习联合表示网络使用的正则化参数是 0.005,其他同理。由图 3(a)可知,正则化参数 λ 取 0.01/0.05 时,UMF-SCF 模型可以达到最佳效果,因此本文采取此参数组合。

学习率是神经网络训练中的一个重要的超参数,控制着神经网络基于损失梯度调整网络权重的速度。我们选取 0.0001, 0.001, 0.01 和 0.1 共 4 个学习率进行实验,结果如图 3(b)所示。可以看到,当学习率为 0.01 时,模型的效果最佳,所以本文所有神经网络都采用该学习率。

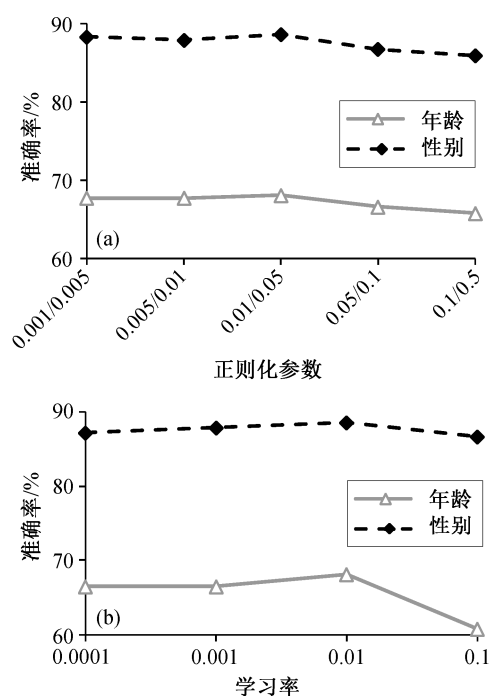


图3 正则化参数和学习率对分类结果的影响
Fig. 3 Influence of regularization parameters and learning rate on classification results

3 结语

本文提出一种跨模态学习的模型UMF-SCF,通过学习多模态联合表示来融合多种数据源。该模型是多层多级的融合模型,包含3个简单而有效的方法,即模型组合策略、跨模态学习联合表示网络和stacking集成方法。通过将不同数据源的用户数据整合到联合表示来完成多模态融合,从而更好地实现用户属性预测。UMF-SCF模型在特征级和决策级融合不同的模态,根据用户生成的内容和社会关系内容,准确地预测社交媒体用户的多个属性。

本文在来自新浪微博的超过4000个用户的用户建模任务上,对UMF-SCF体系结构进行评估,构建4个数据源。通过对新浪微博用户的年龄和性别进行预测,得出用户画像。结果表明,本文的UMF-SCF模型提高了数据源组合的学习能力,取得领先于其他基线方法的结果,其中年龄预测任务准确率得分超过0.68,性别预测任务准确率得分超过0.88。在未来的工作中,我们将进一步将跨模态学习联合表示网络拓展到图学习领域,并采用GCN等方法实现多模态数据源的更深层融合。

参考文献

- [1] Xiao J, Hao Y, He X, et al. Attentional factorization machines: learning the weight of feature interactions via attention networks // Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, 2017: 2119–3125
- [2] Wei H, Zhang F, Yuan N J, et al. Beyond the words: predicting user personality from heterogeneous information // Tenth ACM International Conference on Web Search and Data Mining. Cambridge, 2017: 305–314
- [3] Wollmer M, Weninger F, Knaup T, et al. YouTube movie reviews: sentiment analysis in an audio-visual context. IEEE Intelligent Systems, 2013, 28(3): 46–53
- [4] Gu Y, Yang K, Fu S, et al. Hybrid attention based multimodal network for spoken language classification // International Conference on Computational Linguistics. Santa Fe, 2018: 2379–2390
- [5] Gu Y, Chen S, Marsic I. Deep Multimodal learning for emotion recognition in spoken language // 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, 2018: 5079–5083
- [6] Farnadi G, Jie T, Cock M D, et al. User profiling through deep multimodal fusion // Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey, 2018: 171–179
- [7] You Z, Qian T. Learning a joint representation for classification of networked documents // International Conference on Neural Information Processing. Siem Reap, 2018: 199–209
- [8] Tang J, Qu M, Wang M, et al. LINE: large-scale information network embedding // 24th International Conference on World Wide Web. Florence, 2015: 1067–1077
- [9] Mikolov T, Chen K, Corrado G S, et al. Efficient estimation of word representations in vector space // 1st International Conference on Learning Representations. Scottsdale, 2013: 1–12
- [10] Zhang W, Wang W, Wang J, et al. User-guided hierarchical attention network for multi-modal social image popularity prediction // Proceedings of the 2018 World Wide Web. Lyon, 2018: 1277–1286