基于文本引导的注意力图像转发预测排序网络

潘文雯[1] 赵 洲[1] 俞 俊[2] 吴 飞[1]

摘 要 转发预测在社交媒体网站 (Social media sites, SMS) 中是一个很有挑战性的问题. 本文研究了 SMS 中的图像转发预测问题, 预测用户再次转发图像推特的图像共享行为. 与现有的研究不同, 本文首先提出异构图像转发建模网络 (Image retweet modeling, IRM), 所利用的是用户之前转发图像推特中的相关内容、之后在 SMS 中的联系和被转发者的偏好三方面的内容. 在此基础上, 提出文本引导的多模态神经网络, 构建新型多方面注意力排序网络学习框架, 从而学习预测任务中的联合图像推特表征和用户偏好表征. 在 Twitter 的大规模数据集上进行的大量实验表明, 我们的方法较之现有的解决方案而言取得了更好的效果.

关键词 图像转发预测, 多模态学习, 文本引导, 注意力机制

引用格式 潘文雯, 赵洲, 俞俊, 吴飞. 基于文本引导的注意力图像转发预测排序网络. 自动化学报, 2021, **47**(11): 2547−2556

DOI 10.16383/j.aas.c200629

Textually Guided Ranking Network for Attentional Image Retweet Modeling

PAN Wen-Wen[1] ZHAO Zhou[1] YU Jun[2] WU Fei[1]

Abstract Retweet prediction is a challenging problem in social media sites (SMS). In this paper, we study the problem of image retweet prediction in social media, which predicts the image sharing behavior that the user reposts the image tweets from their followees. Unlike previous studies, we learn user preference ranking model from their past retweeted image tweets in SMS. We first propose a heterogeneous image retweet modeling network (IRM) that exploits users past retweeted image tweets with associated contexts, their following relations in SMS and preference of their followees. We then develop a novel attentional multi-faceted ranking network learning framework with textually guided multi-modal neural networks for the proposed heterogenous IRM network to learn the joint image tweet representations and user preference representations for prediction task. The extensive experiments on a large-scale dataset from Twitter site show that our method achieves better performance than other state-of-the-art solutions to the problem.

Key words Image retweet prediction, multi-modal learning, textually guided prediction, attention mechanism

Citation Pan Wen-Wen, Zhao Zhou, Yu Jun, Wu Fei. Textually guided ranking network for attentional image retweet modeling. *Acta Automatica Sinica*, 2021, **47**(11): 2547−2556

如今, 类似 Twitter 的微博服务系统已经成为用户分享媒体内容的重要社交平台, 平台中一项关键机制是转发功能. 在 SMS 中, 跟随其他用户的用户被称为 “跟随者”, 被跟随的用户被称为 “被跟随者”. 转发预测的核心问题是建立用户沿着 “被跟随者−跟随者”这条推特转发链接的推特共享行为模型, 这一点在文献 [1−2] 中引起了广泛的关注.

已有的转发预测方法[1, 3] 能够从用户过去转发的文本推特中学习用户偏好模型进行预测, 但仅限于文字类型的推特. 随着移动设备的普及, 用户生成图像推特的数量急剧增长, 现今 Twitter 中有 17.2 % 的推特内容与图像相关[2]. 因此, 研究社交媒体网站中的图像转发预测问题具有重要意义, 图 1 中给出了一个简单的图像转发预测示例. 但由于图像推特没有区分性的特征表征[2], 且 SMS 的数据稀疏, 现有的转发预测方法对图像转发问题的预测用处不大.

目前, 现有的转发预测方法[1, 3] 大多涉及到媒体信息的选择和表征, 包括推特的图像和标题、用户的社会角色[3] 和情感[4]. 近年来, 利用预先训练好

收稿日期 2020-08-10 录用日期 2020-12-31
Manuscript received August 10, 2020; accepted December 31, 2020
国家重点研发计划 (2018AAA0100603), 浙江省自然基金 (LR19F020006), 国家自然科学基金项目 (61836002, U1611461, 61751209) 资助
Support by National Key R&D Program of China (2018AAA0100603), Zhejiang Natural Science Foundation (LR19F020006), National Natural Science Foundation of China (61836002, U1611461, 61751209)
本文责任编委 桑农
Recommended by Associate Editor SANG Nong
1. 浙江大学计算机科学与技术学院 杭州 310027 2. 杭州电子科技大学计算机科学与技术学院 杭州 310018
1. School of Computer Science and Technology, Zhejiang University, Hangzhou 310027 2. School of Computer and Science, Hangzhou Dianzi University, Hangzhou 310018

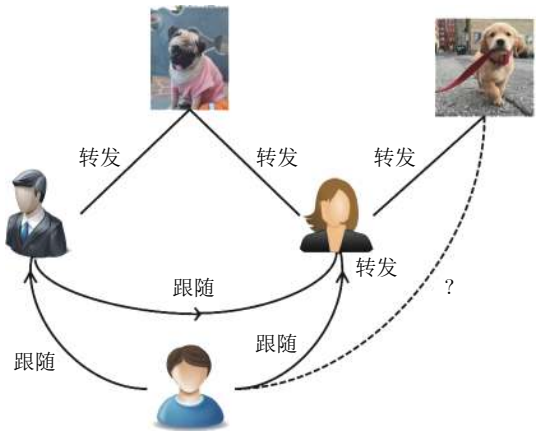


图 1 图像推特行为示例

Fig. 1 An example of image retweet behavior

的卷积神经网络 (Convolutional neural networks, CNN) 获得图像表征的高层视觉特征在各种视觉识别任务中都取得不错的效果[5−6]. 图像推文总是视觉数据, 因此可以使用深度卷积神经网络[7] 来学习图像推文的视觉表示. 另一方面, 图像推特通常与文本上下文信息相关联, 比如用户的评论和标题[2]. 背景图像推文信息通常能传达重要信息, 有助于理解推文, 我们利用深度递归神经网络[8] 学习其语义表征. 目前的研究使用各种模型, 如张量分解[9] 和概率矩阵分解[10] 来模拟用户的转发行为, 但这些模型缺乏探索图像推特与其标题或评论之间关系的能力. 我们使用多模态神经网络[11] 从多模态内容中学习联合图像的推特表征, 从而可提供不同形式的互补信息.

SMS 数据的稀疏性也是图像转发预测的一个挑战性问题. 在 SMS 网站中, 通过用户对图片推文的转发关系来构建图像推文与用户之间的网络. 通常情况下, 每个用户只转发少量的图像推文, 因此 SMS 网络是稀疏的. 受同质性假设[12] 的启发, 我们可以联合考虑用户关注者和用户被转发推文的集合信息来解决图像转发预测的稀疏性问题. 转发行为的社会影响在用户和不同的关注者之间是不同的. 因此我们利用关注机制[13] 自适应地融合用户关注者偏好, 共同预测目标用户的图像转发行为.

本文从多模态注意力排序[14] 网络学习的角度研究图像转发预测问题. 本文首先提出一个异构的图像转发模型 (Image retweet modeling, IRM) 网络, 该模型利用多模态图像推特、用户的转发行为及其跟踪关系三方面进行图像转发预测. 本文引进由文本引导的两个子网络多模态神经网络, 其中递归神经网络学习图像推特上下文信息的语义表征, 卷积神经网络学习视觉表征. 我们在多模态神经网络的基础上使用多面注意力排序方法, 使多面排序的度量值隐含在用户偏好表征中来进行图像转发预测. 本文的主要贡献如下:

1) 与以往研究不同, 本文提出了一种异构的 IRM 网络模型来模拟图像转发的预测, 该模型利用了用户过去转发的图像推特与相关上下文、用户在 SMS 中的跟随关系和用户对后续内容的偏好三方面信息.

2) 提出了基于文本引导的多模态神经网络的注意力多面排序方法来学习基于转发推文的用户偏好表示和图像推文预测的跟踪关系.

3) 使用从 Twitter 收集的数据集来评估方法性能. 大量实验表明, 本文方法优于现有的方法.

1 相关工作

转发预测近年来得到了广泛而深入的研究. 它是当今社会媒体进行信息传播的一种方式. 为了准确地建模用户的转发行为, 我们将当前的研究工作分为三个方面: 用户转发行为的特征选择、转发模型表征和用户转发排序. 在本节中, 我们将简要回顾这三个方面的相关工作.

用户转发行为的特征选择: 如何选择影响用户转发行为的相关因素的问题已被深入研究. Firdaus 等[4] 探讨了用户话题特定情绪对其转发决定的影响. 实验表明, 上下文特征对转发率的贡献很大, 而过去推文的分布并不影响用户的转发率. Yang 等[15] 将社会角色识别和信息扩散集成到一个整体框架中, 对用户社会角色的相互作用进行建模. Chen 等[16] 研究了一些语义特征来学习推文的情感表达. Macskassy 等[17] 解释在不熟悉的领域, 通过评估不同的预测模型和特征, 可以更好地理解用户的转发行为. Xu 等[18] 研究了用户发布行为的因素, 包括突发新闻、用户社交好友发布的帖子和用户的内在兴趣, 同时提出了一个潜在模型来进一步证明这些因素的有效性. Zhang 等[3] 考虑用户 (再) 推文行为, 重点研究自我网络中的朋友如何影响转发行为. 与现有的方法不同, 我们的方法是收集图像推特及其标题或评论. 我们认为不同的文字或评论不仅代表了图片广泛的语义信息, 而且由于用户的社会互动而相互关联.

转发模型表征: 已有大量的研究针对用户转发表征的建模. Zhang 等[1] 使用非参数统计模型结合结构、文本和时间信息来预测转发行为. Luo 等[19] 开发了一个学习排名框架来探索各种转发特性. Bourigault 等[20] 从时间信息扩散的角度考虑任务. 该模型学习了一个扩散核, 其中级联的感染时间用

投影空间中节点的距离表示. Jiang 等[10] 提出了一种基于概率矩阵分解方法的转发预测模型, 将观测到的转发数据、社会影响和消息语义进行整合, 提高了预测的准确性. Hoang 等[9] 将转发行为视为推文、推文作者及其追随者的三维张量, 并通过张量分解同时表示它们. Bi 等[21] 在转发数据上使用了两种新的贝叶斯非参数模型 URM 和 UCM. 两者都能将对推文文本的分析和用户的转发行为整合在同一个概率框架中. Jiang 等[22] 采用矩阵补全方法优化用户转发表示的因子分解. 尽管先前的研究探索学习广泛的表示用户的转发模型, 其中大多数没有特别考虑的共同表示图像转发和他们的标题或评论, 我们提出了文本指导多通道网络数据集和评估其有效性使用推特网站.

用户转发排序: 转发预测的中心问题是对用户转发的推文分享行为进行建模, 用户通过 “被跟随者——跟随者”链接转发推文, 并对社交媒体中出现的所有推文进行排序, 让更多的用户通过短信获得消息, 这一点最近在 Wang 等的工作[23] 中引起了相当大的关注. Liu 等[24]利用模糊理论设计了根基函数神经网络, 对用户热点话题的转发行为进行建模. Firdaus 等[25] 提出了基于用户作者和转发行为的转发预测模型. Zhang 等[1] 提出了非参数模型, 将结构信息、文本信息和时间信息结合在一起来预测转发行为. Wang 等[23] 提出了一个推荐模型来解决 tweet 中提到的问题, 该模型使用了深度神经网络来整合上下文和社会信息. Feng 等[26] 提出特征感知因子分解模型对推文进行重新排序, 该模型统一了线性判别模型和低秩因子分解模型. Peng 等[27] 对转发行为进行建模, 并使用条件随机字段对推文进行排序, 研究了划分社交图和构建转发预测网络关系的方法.

2 基于注意力排序网络学习的图像转发预测

在这一部分, 我们首先会阐述如何使用异构 IRM 网络进行图像转发预测. 然后在此基础上, 提出基于跟随偏好的多面注意力排序方法. 同时提出文本引导的多模态网络, 通过用户的上下文注意力机制来引导图像区域来共同表征相关信息.

2.1 问题定义

首先, 在阐述基本模型之前, 我们需要先介绍模型所涉及的基本概念和专业术语. 我们用 $I=\{\boldsymbol{i}_1, \boldsymbol{i}_2, \cdots, \boldsymbol{i}_n\}$ 表示一组图像推特, 用 $D=\{\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_n\}$ 来表示文本内容, 其中 $\boldsymbol{d}_i=\{\boldsymbol{d}_{i1}, \boldsymbol{d}_{i2}, \cdots, \boldsymbol{d}_{ik}\}$ 表示第 i 个图像推特不同标题和评论的文字嵌入. 用户集表示为 $U=\{u_1, u_2, \cdots, u_m\}$, 用户偏好的排序模型表示为 $R=\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_m\}$, 其中 $\boldsymbol{r}_i$ 是第 i 个用户 u_i 的偏好表征嵌入.

利用上述符号将图像转发预测的问题定义如下: 考虑到用户 U 和输入图像推特 I 及其相关上下文 D, 我们的目标是学习所有用户偏好 R 的多方面排序度量值表征, 然后对目标用户的图像推特进行排序以实现图像转发预测.

2.2 异构 IRM 网络

为探索图像特征和上下文信息的作用, 我们用 $\boldsymbol{z}=\{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n\}$ 来表示联合图像推特表征. 其中, $\boldsymbol{z}_i$ 是由第 i 个图像推特的视觉表征 $\boldsymbol{i}_i$ 和上下文语义表征 $\boldsymbol{d}_i$ 的联合表征组成, 具体的融合方式可见第 3.3 节.

已有的转发预测方法[1, 3] 只从用户过去转发的文本推特中学习用户偏好模型, 进而预测用户的推文共享行为. 与以往的研究不同, 本文提出了多维度注意力排序度量异构 IRM 网络, 该网络利用多模态图像推特、用户过去的转发行为及其后续关系进行图像转发预测. 我们将提出的异构 IRM 网络表示为 $G=(V; E)$, 其中节点集 V 由联合图像表征 Z 和用户偏好表征 R 组成, 边集 E 由用户过去的转发行为 H 及其跟随关系 S 组成, 用矩阵 $H \in \mathbf{R}^{n\times m}$ 表示图像推特与用户之间的转发行为, 其中, 如果第 i 个图像推特被第 j 个用户转发, 则条目 $h_{i,j}$=1, 否则 $h_{i,j}$= 0. 然后通过矩阵 $S \in \mathbf{R}^{m\times m}$ 考虑用户之间的跟随关系, 其中, 如果第 i 个用户跟随第 j 个用户, 则 $s_{ij}=1$. 之后用 N_i 表示第 i 个用户的跟随集合 (如果 $s_{ij}=1$, 则为 $u_j \in N_i$), 用 $N=\{N_1, N_2, \cdots, N_m\}$ 表示用户的跟随集合.

之后, 将从 IRM 网络中导出异质三元组约束作为用户训练多面注意力排序网络的相关偏好. 根据现有的 Twitter 分析工作[28], 我们认为用户可能会对被追随者的未转发图像推特表现出隐性的负面兴趣. 设定第 i 个联合图像推特表征为 z_i, 第 j 个用户为 u_j, 将 u_j 的被跟随者的非转发图像推特 $\boldsymbol{z}_k$ 进行采样. 通过有序元组 (j, i, k, N_j) 对用户的相对偏好进行建模, 表示第 j 个用户更喜欢第 i 个图像推特, 而非第 k 个图像推特. 设 $T=\{(j, i, k, N_j)\}$ 表示从 IRM 网络获得的一组有序元组, 用于一组 n 个图像推特和 m 个用户. 将有序异构元组视为学习用户偏好表示的约束, 即学习用于图像转发预测的多面排序度量函数. 对于任何 $(j, i, k, N_j) \in T$, 以下不等式成立:

$$F_{u_j}(\boldsymbol{z}_i) > F_{u_j}(\boldsymbol{z}_k) \Leftrightarrow$$
$$f_{u_j}(\boldsymbol{z}_i)h_{N_j}(\boldsymbol{z}_i) > f_{u_j}(\boldsymbol{z}_k)h_{N_j}(\boldsymbol{z}_k)$$

其中, $F_{u_j}(\cdot) = f_{u_j}(\cdot)h_{N_j}(\cdot)$ 是用于图像转发预测的第 j 个用户的多面排序模型. $f_{u_j}(\cdot)$ 函数是第 j 个用户的个性化排名模型, $h_{N_j}(\cdot)$ 模型是跟随偏好对第 j 个用户的社会影响. 在假设用户的偏好可能根据同一图像推特的不同部分而变化的情况下, 将用户偏好 U 的维数空间设为联合图像推特表征 Z 的确定权重, 则可用相对排序模块来计算用户对每一个图像推特的相对偏好结果, 其中 $f_{u_j}(\boldsymbol{z}_i) = \boldsymbol{r}_j^{\mathrm{T}}\boldsymbol{z}_i$ 为个性化排序函数, $\boldsymbol{r}_j$ 是第 j 个用户的相对偏好, $\boldsymbol{z}_i$ 是第 i 个图像推特的联合表示. 根据同相假说[12], 社交网络中的个体倾向于按照自己的行为和观点与相似的他人交往和联系. 因此, 我们还将用户对图像推特建模的跟随偏好进行建模, 用 $h_{N_j}(\cdot)$ 来表示.

于是, 可将图像转发预测问题重新表述如下: 给出联合图像推特表征 Z、用户相对偏好 T 的有序元组集和异构 IRM 网络 G, 学习所有用户偏好 R 的表征, 然后由多方面用户偏好函数 $F_u(\cdot)$ 对将要发送给用户 u 的图像推特进行排序.

2.3 文本引导的注意力排序网络学习

在这一部分中, 我们提出基于文本引导的多模层多面注意力排序网络.

我们选择合适的多模态神经网络[29] 来表示 IRM 网络中的图像推特, 它由两个子网络组成: 用于图像数据可视化表示的深度卷积神经网络[30] 和用于文本上下文数据语义表征的深递归神经网络[8]. 这两个子网络在多模融合层中相互作用形成联合表征, 如图 2 所示. 对于一组图像 $I = \{\boldsymbol{i}_1, \boldsymbol{i}_2, \cdots, \boldsymbol{i}_n\}$, 首先通过预训练的 CNN (Convolution neural network) 最后一层卷积层学习图像推特的卷积特征 $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, 其中 $\boldsymbol{x}_i$ 是包含图像视觉信息的三维特征. 通过与 $F = \{\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n\}$ 相同的卷积神经网络最后一层全连接层来学习图像的视觉嵌入. 具体来说, 使用预训练初始网络[30]进行视觉表征, 它在许多视觉表征任务[31−33] 中得到了广泛的应用. 同时训练 LSTM (Long short-term memory) 网络[8] 以获得图像推特的相关上下文, 对于一组文本上下文 $D = \{\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_n\}$, 将 LSTM 最后一个隐藏层的潜在状态作为文本上下文的语义嵌入 $Y = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n\}$, 将 $\boldsymbol{y}_i = \{\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \cdots, \boldsymbol{y}_{ik}\}$ 表示图像推特的不同标题和注释的语义嵌入.

为了学习不同模式图像推特的联合表征, 建立多模态层, 将递归神经网络部分的文本表示和卷积神经网络部分的视觉表征连接起来, 如图 2 所示. 从而, 可以将图像推特的视觉表征和文本上下文的语义表征映射到同一个多模态特征融合空间中, 并将它们相加以获得多模态融合层的激活, 如

$$\boldsymbol{z}_i = g(\boldsymbol{W}^{(i)}\boldsymbol{f}_i + \boldsymbol{W}^{(d)}\boldsymbol{y}_i)$$

其中矩阵 $\boldsymbol{W}^{(i)}$ 和 $\boldsymbol{W}^{(d)}$ 是权重矩阵. $g(\cdot)$ 是非饱和激活函数 ReLU (Rectified linear unit)[34].

然而, 这种简单的方法并没有利用不同的评论与其匹配的图像推特之间的上下文关系. 为了获得更相关的图像推特和文本评论表征, 本文提出文本引导的多模态融合层, 如图 2 所示, 细节如图 3 所示. 由于每个图像推特都有来自其发布者和订阅者的许多标题和评论, 可假设不同的评论表达了图像

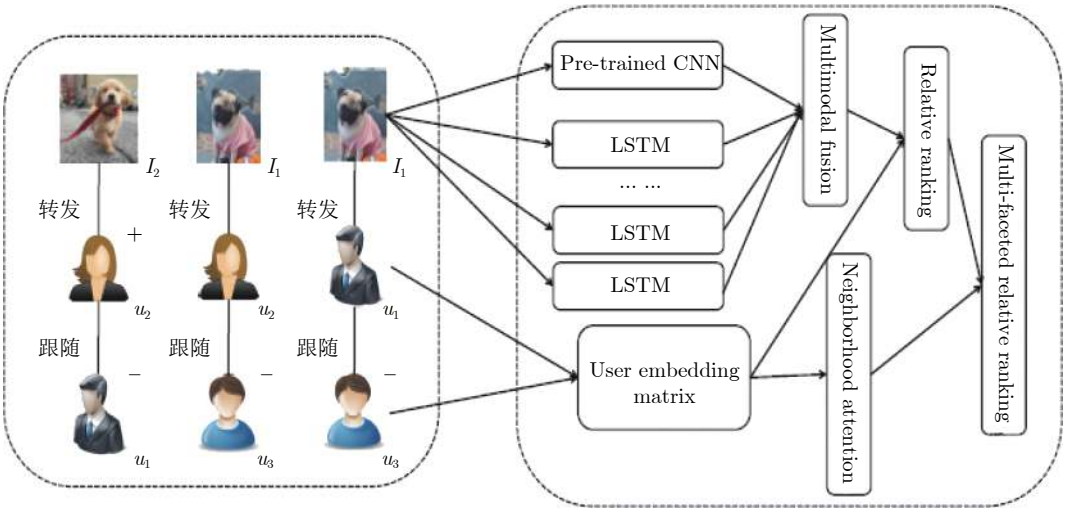


图 2 用于图像转发预测的注意多方面排序网络学习纵览

Fig. 2 The overview of textually guided ranking network for attentional image retweet modeling

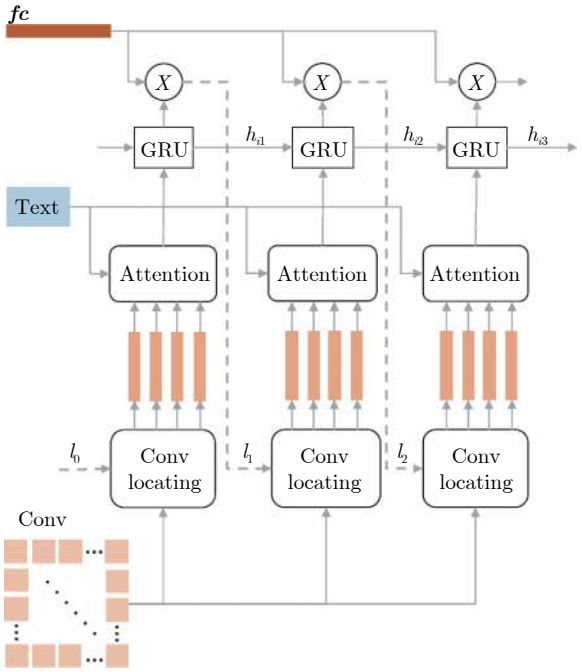


图 3　文本引导的多模融合网络

Fig. 3　Textually guided multi-modal fusion network

的关联和扩展信息. 因此, 本文没有直接使用来自预训练 CNN 最后一层全连接层的视觉特征, 而是使用图像的卷积特征来生成用户对图像推特注意力表征. 在此多模融合网络中, 使用注意力机制来进行一定的约束, 同时以文本信息 $\boldsymbol{y}_i$ 来达到注意力机制的实现, 将文本信息与其图像中的内容进行关联, 如图 3 所示. 注意力模块可以定位图像中适合用户聚焦的区域 $L=\{\boldsymbol{l}_0,\boldsymbol{l}_1,\cdots,\boldsymbol{l}_k\}$, 其中 $\boldsymbol{l}_i=\{\boldsymbol{l}_{xi},\boldsymbol{l}_{yi}\}$ 分别表示图像卷积特征中的 x 轴和 y 轴坐标. 给定卷积特征 $\boldsymbol{x}_i$ 和位置映射向量 $\boldsymbol{l}_j$, 图 3 中的卷积采样自 $\boldsymbol{x}_i$ 中心的 $\boldsymbol{l}_j$ 处的 $3\times3\times S$ 图像特征, 其中 S 是卷积特征的大小. 将采样的 3 维图像特征 $\eta(x_i,l_j)=\{\eta_{i1},\eta_{i2},\cdots,\eta_{i9}\}$ 调整为 $9\times S$. 在选择图像的卷积特征通过位置映射向量之后, 使用注意机制将文本嵌入与提取的卷积特征 $\boldsymbol{fc}$ 融合. 给定第 i 个图像 y_{ij} 的第 j 个评论和多维特征 $\eta(\boldsymbol{x}_i,\boldsymbol{l}_j)$ 的语义特征, 可得到第 j 个评论和第 k 个卷积特征的文本注意力分数为

$$s_{jk}=\boldsymbol{p}\cdot\tanh(\boldsymbol{W}^{(t)}\boldsymbol{y}_{ij}+\boldsymbol{W}^{(u)}\eta_{ik}+\boldsymbol{b})$$

其中 $\boldsymbol{W}^{(t)}$ 和 $\boldsymbol{W}^{(d)}$ 是模型训练更新的参数矩阵. $\boldsymbol{b}$ 是偏向量, $\boldsymbol{p}$ 是计算文本注意力得分的参数向量. 对于 $\eta(\boldsymbol{x}_i,\boldsymbol{l}_j)$ 中的每个被跟随者 η_k, 其得分激活为 $\alpha_k=\exp(s_{jk})/\sum_{n\in\eta(\boldsymbol{x}_i,\boldsymbol{l}_j)}\exp(s_{jk})$, 而文本对第 j 个图像卷积特征的影响为 $g_{ij}=\sum_k\alpha_k\eta_{ik}$.

接着, 使用另一个递归神经网络来推断下一个图像区域的位置. 以 g_{ij} 作为第 j 时间步的输入, 用 $\boldsymbol{h}_{ij}$ 和 $\boldsymbol{c}_{ij}$ 表示 RNN (Recurrent neural networks) 的隐藏状态和输出, 这里的视觉特征来自预先训练的 CNN 的最后一层全连接层. 给定图像的视觉嵌入 $\boldsymbol{f}_i$ 和 RNN 的第 j 步的输出 $\boldsymbol{c}_{ij}$, 下一个位置映射向量为

$$\boldsymbol{l}_{j+1}=g(\boldsymbol{W}^{(j)}\boldsymbol{f}_i+\boldsymbol{W}^{(c)}\boldsymbol{c}_{ij})$$

$+$ 为不同形式的元素相加. $\boldsymbol{W}^{(j)}$ 和 $\boldsymbol{W}^{(c)}$ 是权重矩阵. $g(\cdot)$ 是按单元缩放的双曲正切函数. 定义 $TG(\boldsymbol{l}_j,\boldsymbol{x}_i,\boldsymbol{y}_{ij},\boldsymbol{f}_i)$ 为文本指导过程, 通过将模型与递归神经网络叠加, 可以得到下一个位置映射向量和 RNN 的隐藏状态.

在此基础上提出多面注意力排序函数学习方法用于图像转发预测. 受注意机制[14, 35] 的启发, 在图 2 中设计了邻域注意模块, 对社会影响函数 $h_{N_j}(\cdot)$ 进行计算. 基于有序元组约束 $T=\{(j,i,k,N_j)\}$ 计算 $h_{N_j}(\cdot)$. 给定用户偏好表征 $R=\{\boldsymbol{r}_1,\boldsymbol{r}_2,\cdots,\boldsymbol{r}_n\}$, 用户与其跟随者的社会影响注意力分数表示为

$$s_{pq}=\boldsymbol{p}\cdot\tanh(\boldsymbol{W}^{(s)}\boldsymbol{r}_p+\boldsymbol{W}^{(n)}\boldsymbol{r}_q+\boldsymbol{b})$$

其中 $\boldsymbol{W}^{(s)}$ 和 $\boldsymbol{W}^{(n)}$ 是模型训练更新的参数矩阵. $\boldsymbol{b}$ 是偏向量, $\boldsymbol{p}$ 是计算社会影响注意力得分的参数向量.

最后, 我们可以定义注意多面排名损失函数如下:

$$L_{(j,i,k,N_j)}=\max(0,c+F^{-}_{u_j}(\boldsymbol{z}_k)-F^{+}_{u_j}(\boldsymbol{z}_i))$$

其中排名函数 $F_{u_j}(\boldsymbol{z}_i)=f_{u_j}(\boldsymbol{z}_i)h_{N_j}(\boldsymbol{z}_i)$, 上标 $F^{+}_{u_j}(\cdot)$ 表示正偏好, $F^{-}_{u_j}(\cdot)$ 表示负偏好, 超参数 c $(0<c<1)$ 控制损失函数的边界.

3　基于注意力排序网络学习的图像转发预测

3.1　数据集准备

1) 数据集信息: 我们从 Twitter 收集数据, Twitter 是一种流行的微博服务, 供网络用户分享他们的媒体内容[36]. 用户通常通过在社交媒体网站上转发图片来显示他们对图像推特的积极偏好. 我们抓取用户的个人资料, 包括他们过去转发的图像推特和他们的跟随关系. 已知, 信息扩散有随距离和用户兴趣指数衰减的趋势. 图像转发作为信息级联的一种形式, 在用户关系网络中呈现出重尾分布. 为了尽可能避免这种现象, 数据集在 “推特–用户–推特” 循环中被爬取. 也就是说, 我们首先选择一条转发次数超过 5 次的微博, 找到它的发布者. 然后, 我们收集出版商最近一个月的图像推特, 并找到转发这

些推特的用户. 只有不到 8 个关注者的用户被过滤, 然后我们再次使用其余的用户找到他们的图片推特. 我们将 “推特用户”循环 5 次, 抽取所有用户、其追随者和关注者的信息, 构建图像转发建模网络. 我们总共收集了 15 500 个用户, 74 927 条图片推文和 274 851 条跟随关系.

2) 图像特征提取: 我们对收集到的图像 tweets 进行如下预处理. 我们从预训练的 Inception-V4 网络的最后一个全连接层中提取全局特征用于图像的特征嵌入, 共 1 536 维向量. 为了满足文本引导多模网络的需求, 我们还从同一个预训练网络的最后一个卷积层提取图像特征, 从而得到每个图像的 8×8×1 536 特征向量.

3) 文本特征提取: 我们首先过滤所有的表情符号和感叹词获得所有的标题和注释. 然后对句子中的每一个词, 我们采用预训练 Glove 模型[37] 来提取语义表征. 字向量的维数是 300. 具体来说, 为每个图片推特设置了四个句子, 每个句子的长度为 12. 对于那些少于 4 个标题或评论的图像推特, 我们复制最后一条评论作为填充. 对于我们的数据集, 词汇表的大小设置为 12 500. 因此, 我们使用单词标记 <unk> 和 <eos> 来标记标题或注释的结尾.

3.2 评价标准

在大多数在线媒体服务中, 转发预测任务旨在向用户提供前 K 个图片推送. 为了评估我们的方法在排名前 K 位的图像推特方面的有效性, 我们采用了两个基于排名的评估标准 Precision@K[2] 和 ROC (Receiver operating characteristic curve) 曲线下与坐标轴围成的面积 (Area under curve, AUC)[38−40] 来评估图像转发预测的性能. 给定用户 U^t 和图像推特 i^t 的测试集, 用 R^{u_i} 表示测试集中某个用户 u_i 的前 K 个图像推特的预测排名, 其中排名列表 $|R^{u_i}|$ 的大小为 K.

3.3 效果比较

评估本文的方法 AMNL (仅使用线性融合方法) 和 AMNL+ (使用文本引导的多模态网络) 的性能, 以及其他五个最新的图像转发预测问题的解决方案.

表 1、表 2 和表 3 分别显示了所有方法对排名标准 Precision@1, Precision@3 和 AUC 的评价结果. 本文以 60 %、70 % 和 80 % 的数据为训练集进行评价, 使用三个等级评估标准比较所有方法的结果值. 然后, 比较模型在不同模式下的性能, 其中用户偏好表示的维度设置为 400, 80 % 的数据用于培训. 所有其他参数和超参数也被选择以保证验证集的最佳性能. 我们在六种方法上评估所有三个标准的平均值. 这些实验结果揭示了一些有趣的观点:

表 1 不同方法的 Precision@1 结果
Table 1 Experimental results on precision@1 of different approaches

方法	Precision@1		
	60 %	70 %	80 %
RRFM	0.6253	0.6474	0.6583
VBPR	0.6399	0.6525	0.6793
D-RNN	0.7001	0.7191	0.7385
IRBLRUS	0.7193	0.7295	0.7516
ADABPR	0.6394	0.6488	0.6692
CITING	0.7463	0.7608	0.7773
AMNL	0.8691	0.8975	0.9008
AMNL+	**0.9341**	**0.9444**	**0.9585**

表 2 不同方法的 Precision@3 结果
Table 2 Experimental results on precision@3 of different approaches

方法	Precision@3		
	60 %	70 %	80 %
RRFM	0.5973	0.6284	0.6400
VBPR	0.6082	0.6304	0.6432
D-RNN	0.6468	0.6702	0.6879
IRBLRUS	0.6593	0.6684	0.6813
ADABPR	0.5980	0.6198	0.6301
CITING	0.7304	0.7467	0.7677
AMNL	0.7519	0.7791	0.7959
AMNL+	**0.8680**	**0.8796**	**0.8823**

表 3 不同方法的 AUC 结果
Table 3 Experimental results on AUC of different approaches

方法	AUC		
	60 %	70 %	80 %
RRFM	0.5032	0.5195	0.5282
VBPR	0.5491	0.5799	0.5814
D-RNN	0.6834	0.6973	0.6999
IRBLRUS	0.7145	0.7342	0.7440
ADABPR	0.5393	0.5601	0.5782
CITING	0.5802	0.5982	0.6425
AMNL	0.7703	0.7998	0.8486
AMNL+	**0.8792**	**0.8986**	**0.9126**

1) 以内容特征作为学习排序度量的辅助信息的方法 CITING、D-RNN 和 VBPR 的性能优于低秩因子化排序度量方法 ADABPR 和 RRFM, 这说

明同时包含图像推送和相关上下文信息的深层神经网络对图像转发预测问题至关重要.

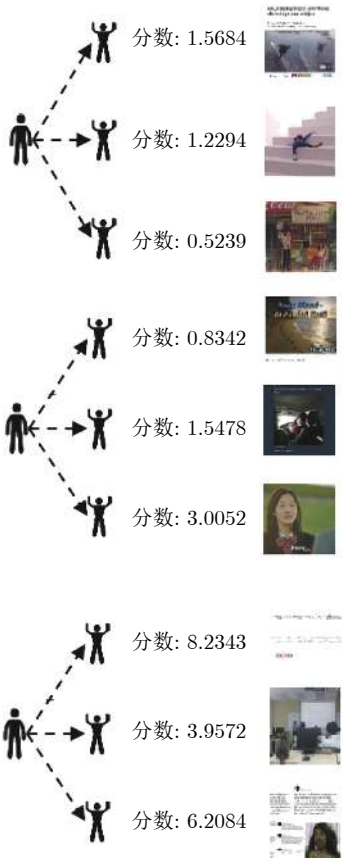
2) 与其他含边信息的排序方法相比, AMNL_{i} 的性能优于 VBPR, AMNL_{d} 的性能优于 CITING. 这表明多方面的排名指标很重要.

3) 与 AMNL 方法相比, AMNL+ 方法具有更好的性能. 这表明通过文本引导的多模态融合方法, 可以更好地将图像推特与包含相关语义信息的不同标题或评论联合表征, 从而在图像转发预测中获得更好的性能.

4) 在所有情况下, AMNL+ 方法都能达到最佳性能. 这表明利用多模态图像推特的联合图像推特表征及其关联上下文和多维度排序度量的注意多维度排序网络学习框架可以进一步提高图像转发预测的性能.

图 4 (a) 和 4 (b) 展示了 AMNL+ 对一些用户的图像转发预测的实验结果实例. 在文本选取方面, 我们主要选取了与图像推特相关联程度和语法语义复杂程度不同的句子组合, 比如有不同的对象等, 从而来测试注意力机制的实际效用. 而且在实际现象中, 有推荐意义的推特往往具有比较精彩和相对复杂的语义内容. 而在图 4 列举的例子中, 由于篇幅有限, 以 3 个句子为例. 图 4 由两个部分组成, 其中图 4 (a) 显示用户和用户的跟随者发布的图像和模型所预测的排名分数, 图 4 (b) 显示的是预测的推特图像及其注释对于注意力模块的反馈效果. 图 4 (a) 列举出了对不同推特的预测喜爱程度, 其低排名分数表明越不值得推荐. 我们可以发现, 其中关注者发布的未被转发的图像推特获得了比较差的排名分数. 这说明这些推特更容易被用户看到, 但不受用户欢迎, 与事实的情况相吻合. 图 4 (b) 显示的是几个图像推特中注释得分较高的例子, 不同的注释评论中的关键词在图像中得到了很好的注意力显现. 这表明, 用我们的方法预测的图像对于图 4 (a) 中的用户是可取的. 值得一提的是, 一些特定的词语与图像中颜色相同的物体相匹配, 体现注释和字幕对预测有一定引导效果.

本文所提出的注意力多面排名网络学习方法的训练更新规则本质上是迭代的. 接着, 本文将继续研究 AMNL 方法收敛的方式. 图 5 (a) 和图 5 (b)



(a)

评论1: Bird in love
评论2: Unlock the car Sharon, i love you. we can make this work.
评论3: Gotta go home, it will rain soon. goodbye honey.
分数: 1.8492

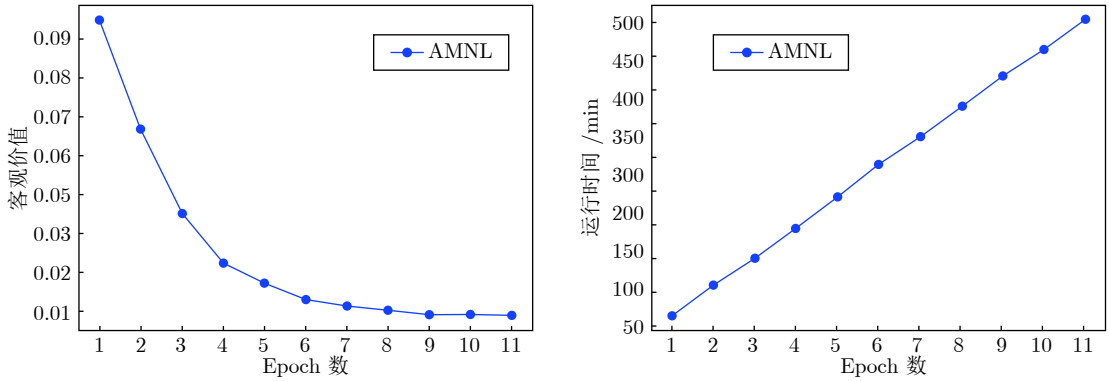
评论1: A pet isn't a toy, or a item.
评论2: A pet is member of a family, not a thing you can throw away.
评论3: Poor dog in rainy day.
分数: 3.9844

评论1: Lunch with my son at Nobu!
评论2: This woman has no clue who's behind her.
评论3: Who's the celebrity behind her?
分数: 10.4782

(b)

图 4　AMNL+ 在图像转发预测任务中的实验结果

Fig. 4　Experimental results of AMNL+ on the image retweet prediction task



(a) 客观价值
(a) Objective value

(b) 运行时间
(b) Running time

图 5 随着 Epoch 客观价值和运行时间的变化

Fig. 5 Objective value and running time versus the number of epochs

分别为 AMNL 方法的收敛曲线和运行时间曲线. 图中 x 轴表示迭代次数. 图 5 (a) 中的 y 轴为目标值, 图 5 (b) 中的 y 轴为本文所提方法的运行时间. 每个 Epoch 包含 231 539 次迭代更新. 我们将用户偏好表示的维数设置为 500, 并使用 80 % 的数据进行训练. 结果表明, 该方法在 10 轮后完成收敛, 计算时间约为 500 分钟. 这项研究验证了 AMNL 的有效性.

3.4 消融实验

我们评估了我们的模型模块部分的贡献, 主要评估了文本引导的多模式融合网络和社会影响功能. 同时, 我们还评估了图像推特的视觉表征、关联上下文的语义表征和联合图像推文表示对模型的影响.

为了理解组件的贡献和不同介质对我们的模型的影响, 我们提出消融研究并在表模型: AMNL_{i} 方法意味着我们只使用图像推特的视觉表征. AMNL_{d} 方法意味着只对相关上下文进行语义表征. $\text{AMNL}_{+\text{i}}$ 模型是在文本引导的多模态融合网络中, 将图像推特卷积特征的平均池化直接输入到递归神经网络中, 而不是在文本表示中使用注意机制. $\text{AMNL}_{\text{hfunc}}$ 和 $\text{AMNL}_{+\text{hfunc}}$ 模型是指我们直接对两个模型的排名函数进行计算, 而不使用社会影响函数. 如表 4 所示, 我们还发现了一些有趣的结果:

1) 与 AMNL_{i} 和 AMNL_{d} 方法相比, AMNL 方法取得了更好的性能. 这表明相比于只利用视觉特征或者文本特征, 注意力多方面的排名网络学习框架利用多模式图像的联合图像推特表征及其相关的上下文可以获得更好的性能.

2) 与 $\text{AMNL}_{+\text{hfunc}}$ 的结果相比, AMNL_{+} 在三项标准中得分较高. 这表明社会影响函数可以帮助提高我们的方法的性能. 而 $\text{AMNL}_{+\text{hfunc}}$ 和 AMNL 的实验结果进一步证明了我们的上述结果在不同分量之间是一致的.

表 4 用 80 % 的数据进行训练, 消融实验的实验结果

Table 4 Experimental results with different modalities and components using 80 % of the data for training

方法	Precision@1	Precision@3	AUC
$\text{AMNL}_{+\text{i}}$	0.8427	0.7673	0.8204
AMNL_{d}	0.7892	0.7719	0.7962
$\text{AMNL}_{\text{hfunc}}$	0.8598	0.7900	0.8095
AMNL	0.9008	0.7959	0.8486
$\text{AMNL}_{+\text{i}}$	0.9227	0.8276	0.8724
$\text{AMNL}_{+\text{hfunc}}$	0.9199	0.8195	0.8689
AMNL+	**0.9585**	**0.8823**	**0.9126**

4 结论

本文提出了一种基于异构 IRM 网络的图像转发预测模型. 具体来说, 我们的 IRM 网络利用用户过去转发的图像推文、关联的文本上下文和用户的后续关系来采样用户转发行为的适当表示. 在此基础上, 我们提出了一种基于文本引导的多模态神经网络的注意多方面排序方法, 以学习联合图像推特表征和用户偏好表征, 从而将多方面的排序度量嵌入到表示中进行预测. 我们使用来自 Twitter 的数据集来评估我们的方法的性能. 大量的实验表明, 我们的方法比几种最新的解决方案能获得更好的性能.

References

1 Zhang Q, Gong Y Y, Guo Y, Huang X J. Retweet behavior prediction using hierarchical dirichlet process. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, Texas, USA: AAAI Press, 2015. 403−409

2 Chen T, He X N, Kan M Y. Context-aware image tweet modelling and recommendation. In: Proceedings of the 24th ACM In-

ternational Conference on Multimedia. Amsterdam, The Netherlands: ACM, 2016. 1018−1027

3 Zhang J, Tang J, Li J Z, Liu Y, Xing C X. Who influenced you? Predicting retweet via social influence locality. *ACM Transactions on Knowledge Discovery from Data*, 2015, **9**(3): Article No. 25

4 Firdaus S N, Ding C, Sadeghian A. Topic specific emotion detection for retweet prediction. *International Journal of Machine Learning and Cybernetics*, 2019, **10**(8): 2071−2083

5 Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, United States: Curran Associates Inc., 2013. 2553−2561

6 Zhao Z, Yang Q F, Lu H Q, Weninger T, Cai D, He X F, et al. Social-aware movie recommendation via multimodal network learning. *IEEE Transactions on Multimedia*, 2018, **20**(2): 430−440

7 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA, 2015.

8 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, **9**(8): 1735−1780

9 Hoang T A, Lim E P. Retweeting: An act of viral users, susceptible users, or viral topics? In: Proceedings of the 2013 SIAM International Conference on Data Mining (SDM). Austin, Texas, USA: SIAM, 2013. 569−577

10 Jiang B, Lu Z G, Li N, Wu J J, Jiang Z W. Retweet prediction using social-aware probabilistic matrix factorization. In: Proceedings of the 18th International Conference on Computational Science. Wuxi, China: Springer, 2018. 316−327

11 Atrey P K, Hossain M A, El Saddik A, Kankanhalli M S. Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, 2010, **16**(6): 345−379

12 Yuan Z Q, Sang J T, Xu C S, Liu Y. A unified framework of latent feature learning in social media. *IEEE Transactions on Multimedia*, 2014, **16**(6): 1624−1635

13 Luong T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: The Association for Computational Linguistics, 2015. 1412−1421

14 Nie L Q, Yan S C, Wang M, Hong R C, Chua T S. Harvesting visual concepts for image search with complex queries. In: Proceedings of the 20th ACM International Conference on Multimedia. Nara, Japan: ACM, 2012. 59−68

15 Yang Y, Tang J, Leung C W K, Sun Y Z, Chen Q C, Li J Z, et al. RAIN: Social role-aware information diffusion. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence. Texas, USA: AAAI Press, 2015. 367−373

16 Chen J D, Li H, Wu Z J, Hossain M S. Sentiment analysis of the correlation between regular tweets and retweets. In: Proceedings of the 16th International Symposium on Network Computing and Applications (NCA). Cambridge, MA, USA: IEEE, 2017. 1−5

17 Macskassy S A, Michelson M. Why do people retweet? Anti-homophily wins the day! In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. Barcelona, Catalonia, Spain: AAAI Press, 2011. 209−216

18 Xu Z H, Zhang Y, Wu Y, Yang Q. Modeling user posting behavior on social media. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. Portland, Oregon, USA: ACM, 2012. 545−554

19 Luo Z C, Osborne M, Tang J T, Wang T. Who will retweet me? Finding retweeters in twitter. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland: ACM, 2013. 869−872

20 Bourigault S, Lagnier C, Lamprier S, Denoyer L, Gallinari P. Learning social network embeddings for predicting information diffusion. In: Proceedings of the 7th ACM International Conference on Web Search and Data Mining. New York, USA: ACM, 2014. 393−402

21 Bi B, Cho J. Modeling a retweet network via an adaptive Bayesian approach. In: Proceedings of the 25th International Conference on World Wide Web. Montréal, Québec, Canada: ACM, 2016. 459−469

22 Jiang B, Liang J G, Sha Y, Wang L H. Message clustering based matrix factorization model for retweeting behavior prediction. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne, Australia: ACM, 2015. 1843−1846

23 Wang B D, Wang C, Bu J J, Chen C, Zhang W V, Cai D, et al. Whom to mention: Expand the diffusion of tweets by @ recommendation on micro-blogging systems. In: Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro, Brazil: ACM, 2013. 1331−1340

24 Liu Y B, Zhao J Z, Xiao Y P. C-RBFNN: A user retweet behavior prediction method for hotspot topics based on improved RBF neural network. *Neurocomputing*, 2018, **275**: 733−746

25 Firdaus S N, Ding C, Sadeghian A. Retweet prediction considering user′s difference as an author and retweeter. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). San Francisco, CA, USA: IEEE, 2016. 852−859

26 Feng W, Wang J Y. Retweet or not? Personalized tweet reranking. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining. Rome, Italy: ACM, 2013. 577−586

27 Peng H K, Zhu J, Piao D Z, Yan R, Zhang Y. Retweet modeling using conditional random fields. In: Proceedings of the 11th International Conference on Data Mining Workshops. Vancouver, BC, Canada: IEEE, 2011. 336−343

28 Chen K L, Chen T Q, Zheng G Q, Jin O, Yao E P, Yu Y. Collaborative personalized tweet recommendation. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. Portland, Oregon, USA: ACM, 2012. 661−670

29 Nie L Q, Song X M, Chua T S. Learning from multiple social networks. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 2016, **8**(2): 1−118

30 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. 1−9

31 Zhang H W, Kyaw Z, Chang S F, Chua T S. Visual translation embedding network for visual relation detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 3107−3115

32 Zhao W Q, Guan Z Y, Luo H Z, Peng J Y, Fan J P. Deep multiple instance hashing for object-based image retrieval. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: AAAI Press, 2017. 3504−3510

33 Zhao Z, Lin J H, Jiang X H, Cai D, He X F, Zhuang Y T. Video question answering via hierarchical dual-level attention network learning. In: Proceedings of the 25th ACM International Conference on Multimedia. Mountain View, CA, USA: ACM, 2017. 1050−1058

34 Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. Haifa, Israel: Omnipress, 2010. 807−814

35 Zhao Z, Yang Q F, Cai D, He X F, Zhuang Y T. Video ques-

tion answering via hierarchical spatio-temporal attention networks. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: AAAI Press, 2017. 3518−3524

36 Java A, Song X D, Finin T, Tseng B. Why we twitter: Understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis. San Jose, CA, USA: ACM, 2007. 56−65

37 Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: ACL, 2014. 1532−1543

38 He R N, McAuley J. VBPR: Visual Bayesian personalized ranking from implicit feedback. In: Proceedings of the 13th AAAI Conference on Artificial Intelligence. Phoenix, Arizona, USA: AAAI Press, 2016. 144−150

39 Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L. BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, QC, Canada: AUAI Press, 2009. 452−461

40 Li H Y, Hong R C, Lian D F, Wu Z A, Wang M, Ge Y. A relaxed ranking-based factor model for recommender system from implicit feedback. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA: AAAI Press, 2016. 1683−1689

潘文雯 浙江大学计算机科学博士研究生. 主要研究方向为机器学习和自然语言处理.
E-mail: wenwenpan@zju.edu.cn
(**PAN Wen-Wen** Ph. D. candidate in computer science at Zhejiang University. Her research interest covers machine learning and natural language processing.)

赵 洲 浙江大学计算机科学与技术学院副教授. 主要研究方向为机器学习和数据挖掘. 本文通信作者.
E-mail: zhaozhou@zju.edu.cn
(**ZHAO Zhou** Associate professor at the College of Computer Science and Technology, Zhejiang University. His research interest covers machine learning and data mining. Corresponding author of this paper.)

俞 俊 杭州电子科技大学计算机科学与技术学院教授. 主要研究领域为计算机动画图像处理, 机器学习.
E-mail: yujun@hdu.edu.cn
(**YU Jun** Professor at the School of Computer Science and Technology, Hangzhou Dianzi University. His research interest covers computer animation image processing, machine learning.)

吴 飞 浙江大学求是特聘教授. 主要研究方向为人工智能, 多媒体分析与检索和统计学习理论.
E-mail: wufei@cs.zju.edu.cn
(**WU Fei** Qiushi distinguished professor at Zhejiang University. His research interest covers artificial intelligence, multimedia analysis and retrieval, and statistical learning theory.)