软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2022,33(4):1244−1266 [doi: 10.13328/j.cnki.jos.006481]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

标签推荐方法研究综述*

徐鹏宇[1,2], 刘华锋[1,2], 刘 冰[1,2], 景丽萍[1,2], 于 剑[1,2]

[1](交通数据分析与挖掘北京市重点实验室(北京交通大学), 北京 100044)

[2](北京交通大学 计算机与信息技术学院, 北京 100044)

通信作者: 景丽萍, E-mail: lpjing@bjtu.edu.cn

摘 要: 随着互联网信息的爆炸式增长, 标签(由用户指定用来描述项目的关键词)在互联网信息检索领域中变得越来越重要. 为在线内容赋予合适的标签, 有利于更高效的内容组织和内容消费. 而标签推荐通过辅助用户进行打标签的操作, 极大地提升了标签的质量, 标签推荐也因此受到了研究者们的广泛关注. 总结出标签推荐任务的三大特性, 即项目内容的多样性、标签之间的相关性以及用户偏好的差异性. 根据这些特性, 将标签推荐方法划分为3个类别, 分别是基于内容的方法、基于标签相关性的方法以及基于用户偏好的方法. 之后, 针对这3个类别下的对应方法进行了梳理和剖析. 最后, 提出了当前标签推荐领域面临的主要挑战, 例如标签的长尾问题、用户偏好的动态性以及多模态信息的融合问题等, 并对未来研究方向进行了展望.

关键词: 机器学习; 信息检索; 推荐系统; 标签推荐; 用户偏好

中图法分类号: TP311

中文引用格式: 徐鹏宇, 刘华锋, 刘冰, 景丽萍, 于剑. 标签推荐方法研究综述. 软件学报, 2022, 33(4): 1244–1266. http://www.jos.org.cn/1000-9825/6481.htm

英文引用格式: Xu PY, Liu HF, Liu B, Jing LP, Yu J. Survey of Tag Recommendation Methods. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1244–1266 (in Chinese). http://www.jos.org.cn/1000-9825/6481.htm

Survey of Tag Recommendation Methods

XU Peng-Yu[1,2], LIU Hua-Feng[1,2], LIU Bing[1,2], JING Li-Ping[1,2], YU Jian[1,2]

[1](Beijing Key Lab of Traffic Data Analysis and Mining (Beijing Jiaotong University), Beijing 100044, China)

[2](School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: With the explosive growth of Internet information, tags (keywords specified by users to describe the item) become more and more important in the field of Internet information retrieval. Giving appropriate tags to online content is conducive to more efficient content organization and content consumption. Tag recommendation greatly improves the quality of tags by assisting users to tag. Therefore, tag recommendation has been widely concerned by researchers. This study summarizes the three characteristics of tag recommendation task, that is, the diversity of item content, the correlation between tags, and the difference of user preferences. According to these three characteristics, tag recommendation methods are divided into three categories: content-based method, tag relevance based method, and user preference based method. After that, the corresponding methods are sorted out and analyzed under these three categories. Finally, the main challenges are presented in the field of tag recommendation, such as the long tail problem of tags, the dynamics of user preferences, and the fusion of multimodal information, and the future research is prospected as well.

Key words: machine learning; information retrieval; recommendation system; tag recommendation; user preference

* 基金项目: 国家自然科学基金(61773050)

本文由“面向开放场景的鲁棒机器学习”专刊特约编辑陈恩红教授、李宇峰副教授、邹权教授推荐.

收稿时间: 2021-05-31; 修改时间: 2021-07-16; 采用时间: 2021-08-27; jos 在线出版时间: 2021-10-26

在信息系统中, 标签(tag)指的是分配给一条项目(item)的关键词或术语[1], 有助于描述项目(item), 从而使得项目能够被更好地浏览和检索到[2]. 一个项目的标签通常由项目的创建者赋予[3]. 丰富的标签信息, 将为信息检索和信息归类提供极大的便利[4]. 随着互联网信息的爆炸式增长, 标签在互联网应用中已经变得越来越普遍[5]. 比如: 在知乎、Stack Exchange 和 Stack Overflow 等问答网站上, 每一个问题通常包含标题、描述和标签等信息, 如图 1 所示. 提问者可以借助于标签更准确地吸引回答者[6], 浏览用户可以通过标签更好地检索到感兴趣的问题, 社区运营者也可以通过标签更好地归类信息[7].

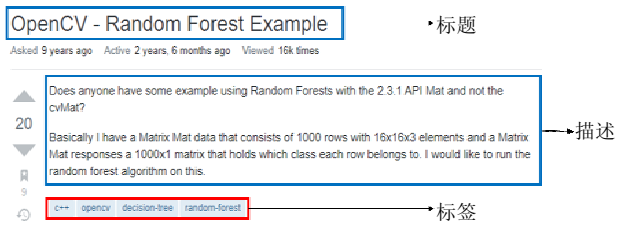


图 1 问答网站 Stack Overflow 上的问题示例

然而, 用户在进行打标签(tagging)操作的过程中会遇到一系列问题. 首先, 由于缺乏专业知识, 用户很难在没有系统辅助的情况下给出完整的标签, 甚至不提供标签[1]; 其次, 如果没有系统指导, 用户可能给出很多近义和冗杂的标签[8]. 比如: 在没有系统辅助的情况下, 用户可能给出“lstm” “lstm-neural-network”或“long-short-term-memory”这 3 种同义的标签. 这样的操作降低了标签信息的数量和质量, 从而会损害到基于标签的后续信息检索任务. 因此, 在用户进行打标签操作的过程中对其进行辅助显得尤为重要. 在这样的背景下, 标签推荐(tag recommendation)应运而生. 标签推荐是指当用户对某个项目(如文本、图片、视频)进行打标签操作时, 平台为用户推荐若干相关标签的过程[9]. 通过标签推荐, 平台不仅提升了用户发布信息时的体验, 还极大地提高了生成标签的数量和质量, 从而有助于建立更高效的信息检索系统[10].

基于上述原因, 我们认为标签推荐是信息检索领域的重要研究方向. 尽管标签推荐相关的论文层出不穷, 据我们所知, 迄今为止只有过一篇综述论文[10]对该领域的方法进行了分类和总结. 由于年代所限, 其只考虑到了 2015 年之前传统的标签推荐方法. 而在近 6 年的时间里, 出现了一大批新的标签推荐方法[5,11−28]. 特别是其中一些基于深度学习的方法在标签推荐领域取得了突破性的结果[7,16,18]. 因此, 我们希望在新的视角下对标签推荐方法进行全面的梳理和总结. 具体来说, 本文根据标签推荐方法要解决的问题, 将现有的方法分为 3 大类, 分别是基于内容(content-based)的方法、基于标签相关性(tag relevance based)的方法以及基于用户偏好(user preference based)的方法. 之后, 我们对这 3 大类下的对应方法进行了梳理和剖析, 分析了模型的优劣. 最后, 我们提出了当前标签推荐领域面临的主要挑战, 例如标签的长尾问题、用户偏好的动态性以及多模态信息的融合问题等, 并且针对上述问题, 对未来的研究方向进行了展望.

本文第 1 节介绍标签推荐的问题描述. 第 2 节介绍标签推荐方法分类. 第 3 节介绍基于内容的标签推荐方法. 第 4 节介绍基于标签相关性的标签推荐方法. 第 5 节介绍基于用户偏好的标签推荐方法. 第 6 节分析现有方法存在的问题, 提出相应的解决方案, 并对未来可能的研究方向和发展趋势加以展望.

为方便读者查阅, 我们将本文中重要词组的英文全称、缩略语以及中文对照在表 1 进行了汇总.

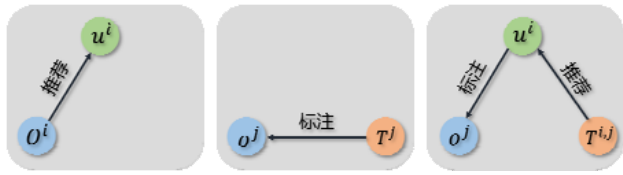
1 问题描述

与标签推荐任务密切相关的两个任务分别是项目推荐(item recommendation)和多标签分类(multi-label classification), 部分标签推荐方法也借鉴了这二者的思路. 一般的项目推荐是指对于当前用户 u^i, 系统根据用户行为等信息为其推荐一个物品集合 $O^j=\{o_1^j,o_2^j,...,o_n^j\}$ 的过程, 如图 2(a)所示. 多标签分类问题一般指的是: 对于当前项目 o^j, 系统根据其内容等信息为其标注一个标签集合 $T^j=\{t_1^j,t_2^j,...,t_n^j\}$ 的过程, 如图 2(b)所示. 而标签推荐任务指的是: 当前用户 u^i 要对目标项目 o^j(往往包含文本、图片或视频等内容信息)标注标签时, 系

统自动为其推荐一个标签集合 $T^{i,j}=\{t_1^{i,j},t_2^{i,j},...,t_n^{i,j}\}$ 的过程. 这一过程也会受到用户信息和项目信息的影响. 用户 u^i、项目 o^j 与标签 $t^{i,j}$ 之间的关系如图 2(c)所示.

表 1 重要词组中英文对照表

英文全称(缩略语)	中文对照
Tag	标签
Item	项目
Tag Recommendation	标签推荐
Content-based	基于内容
Tag Relevance based	基于标签相关性
User Preference based	基于用户偏好
Item Recommendation	项目推荐
Multi-Label Classification	多标签分类
Collaborative Filtering	协同过滤
Tensor Factorization (TF)	张量分解
Neural Graph Collaborative Filtering (NCCF)	图神经网络协同过滤
Graph Convolutional Network (GCN)	图卷积神经网络
Multilayer Perceptron (MLP)	多层感知机
Bag of Words (BoW)	词袋
Term Frequency - Inverse Document Frequency (TF-IDF)	词频-逆文档频率
Recurrent Neural Network (RNN)	循环神经网络
Convolutional Neural Network (CNN)	卷积神经网络
Gated Recurrent Unit (GRU)	门控循环单元
Feature Extractor	特征提取器
Multi-Label Classifier	多标签分类器
Pseudo Probability	伪概率
Cross Entropy	交叉熵
Embedding Matrix	嵌入矩阵
Capsule Networks	胶囊网络
Long Short Term Memory (LSTM)	长短期记忆网络
Autoencoder (AE)	自编码器
Sequence-to-Sequence (Seq2Seq)	序列到序列
Parallel LSTMs (PLSTMs)	并行长短期记忆网络
Pairwise Interaction Tensor Factorization (PITF)	成对交互张量分解



(a) 项目推荐 (b) 多标签分类 (c) 标签推荐

图 2 项目推荐、多标签分类与标签推荐图例

首先, 标签推荐是一个推荐问题, 但与传统的项目推荐任务相比, 其将项目推荐任务下的两个主体(用户和项目)拓展到了 3 个主体(用户、项目和标签), 因此需要考虑更多的影响因素和交互关系. Rendle[29,30]便是将标签推荐看作一个协同过滤(collaborative filtering)任务, 通过张量分解(tensor factorization, TF)建模用户、物品和标签三元组的交互关系, 从而进行标签预测. Wei[17]则是将项目推荐里的图神经网络协同过滤方法(neural graph collaborative filtering, NGCF)[31]拓展到了短视频标签推荐的场景, 利用图卷积神经网络(graph convolutional network, GCN)[32]分别学习用户表示和标签表示, 随后通过多层感知机(multilayer perceptron, MLP)得到融合用户偏好的标签表示和融合用户偏好的内容表示, 并据此得到标签排名. 如果不考虑用户信息, 那么标签推荐可以被简化为一个多标签分类问题. 由于不需要考虑到复杂的用户交互关系, 很大一部分工作[4,7,8,11−14,16,18,21,22,26,28]都是将标签推荐看作一个多标签分类任务, 即先利用特征提取器学习到项目的内容表示, 随后通过多标签分类器进行标签预测.

由于有助于信息检索, 大量互联网数据库上的信息都采用了标签进行标注[10]. 因此, 标签所描述的对象多种多样, 标签推荐的任务场景也非常丰富. 最常见的标签描述对象便是文本信息. Xia 等人[33−36]的研究对象是软件信息问答网站上(如 Stack Exchange, Stack Overflow)的问题及其描述. Gong 等人[5,10,37]的研究对象是社交平台上(如 Twitter、微博)帖子的文本内容. Hassan 等人[14]研究的是如何为学术文献进行标签推荐. Gao 等人[38]研究的对象是政务系统中市民对政府的建议. 除了文本信息, 图片信息也是常见的标签描述对象. Zhang 等人[26,27]的研究对象是图片分享平台(如 Instagram)里的图片. Zuin 等人[25]研究的是如何为画作打上合适的标签. 近年来, 随着短视频分享平台(如抖音、快手)的兴起, 也逐渐有学者开始研究面向短视频信息的标签推荐[17,23]. 除了常见的文本、图片和视频信息, 标签推荐场景下标签描述的对象还有商品[29,30]、音乐[39,40]等.

2 标签推荐方法分类

本节我们通过分析和归纳标签推荐任务的 3 大特性, 对标签推荐方法进行分类和概述.

(1) 项目内容的多样性

如同上一节所述, 大量互联网信息条目都采用了标签进行标注, 因此标签所描述的项目内容多种多样. 其中既有较为简单的纯文本信息[35]和纯图片信息[25], 也存在较为复杂的图文信息[21,27](同一个项目既包含有文字又包含有图片), 甚至还有由图片序列、音频序列和文字组成的复杂的视频信息[17,23]. 对于不同类型的信息, 在进行标签推荐时, 应选用不同的策略. 比如: 在进行文本标签推荐时, 往往只需要考虑到提取出更好的文本特征; 而在进行视频标签推荐时, 更需要关注多个模态信息之间的交互. 因此, 标签推荐任务的第一大特性便是项目内容的多样性.

(2) 标签之间的相关性

在标签推荐场景下, 每一条项目所对应的多个标签之间往往有某种程度的相关性. 如图 1 所示: 标签“decision-tree”与“random-forest”之间就具备很强的相关性, 而“decision-tree”与“c++”之间也具备一定程度的相关性. 类似的标签相关关系在每个项目所对应的标签集中普遍存在. 早期的标签推荐方法[9]便是显式地采用条件概率来建模标签相关性. 近期, 一些基于深度学习的标签推荐方法[7,38]也显式地建模了标签相关性, 从而大幅度提升了标签推荐的效果. 因此, 考虑并捕获标签之间的相关性特性, 是进行标签推荐任务的重要方面.

(3) 用户偏好的差异性

如前文所述, 如果不考虑用户信息, 那么标签推荐问题可以看作一个多标签分类任务. 然而, 用户信息在标签推荐中扮演着极为重要的角色[41−43]. 标签推荐任务中的标签由用户直接赋予, 因此与用户的偏好息息相关[44]. 同一位用户所标记的多个标签集之间往往具备一定程度的相关性[16], 而不同用户的打标签习惯往往存在显著的差异. 这是由于不同用户偏好之间存在着明显的差异性. 在标签推荐中, 主要存在两种形式的用户偏好差异: 一是不同用户关注的项目本来就存在着差异, 比如在基于微博文本的标签推荐场景下, 健身博主和美食博主发布的文本内容就存在着明显的差异性; 二是用户使用标签的习惯也存在着差异性, 这是由于每位用户的背景知识不一样, 因而即使对类似的项目进行打标签操作时, 不同用户所标记的标签也很可能不同. 因此, 用户偏好的差异性也是在标签推荐场景下必须要考虑的特性.

以上便是标签推荐任务的 3 大特性: 项目内容的多样性、标签之间的相关性以及用户偏好的差异性. 同时, 它们也是研究者们研究标签推荐任务需要面临的 3 大核心问题. 基于这 3 大核心问题, 我们将现行的标签推荐方法分为 3 个大类, 分别是基于内容的方法、基于标签相关性的方法以及基于用户偏好的方法. 在 3 大类方法下, 又细分为 9 个小类, 分别是基于文本内容的方法、基于图片内容的方法、基于图文内容的方法、基于视频内容的方法、基于标签共现的方法、基于标签结构的方法、基于标签语义的方法、基于交互关系的方法以及基于用户表示的方法. 其分类关系如图 3 所示. 现行的标签推荐方法均可以按此分类标准进行分类. 常见标签推荐方法在本文分类准则下的分类情况见表 2(由上到下按照时间顺序排列). 需要注意的是: 由于我们提出的分类准则是基于标签推荐3大核心问题的, 因此, 如果一个标签推荐方法同时考虑到了多个问题, 则其同时属于多个类. 而在每个大类下的几个小类之间一般存在着互斥关系, 即每个标签推荐方法在每个大类

下仅属于一个小类.

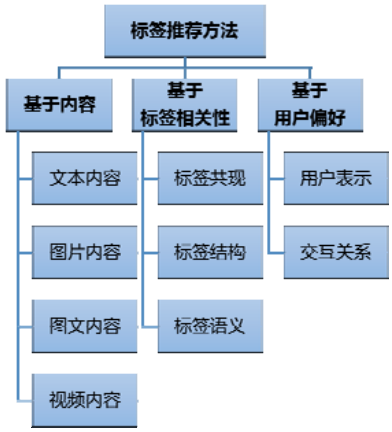


图 3 标签推荐方法分类

表 2 常见标签推荐方法分类

参考文献序号	作者	发表年份	发表刊物(简称)	基于内容				基于标签相关性			基于用户偏好	
				文本内容	图片内容	图文内容	视频内容	标签共现	标签结构	标签语义	交互关系	用户表示
[9]	Sigurbjörnsson B, *et al*.	2008	WWW	√	–	–	–	√	–	–	–	–
[2]	Song Y, *et al*.	2008	SIGIR	√	–	–	–	–	–	–	–	–
[30]	Rendle S, *et al*.	2009	KDD	–	–	–	–	–	–	–	√	–
[5]	Krestel R, *et al*.	2009	RecSys	–	–	–	–	√	–	–	–	–
[42]	Guan Z, *et al*.	2009	SIGIR	√	–	–	–	–	–	–	√	–
[45]	Wang J, *et al*.	2009	PAKDD	√	–	–	–	√	–	–	–	–
[29]	Rendle S, *et al*.	2010	WSDM	–	–	–	–	–	–	–	√	–
[46]	Toderici G, *et al*.	2010	CVPR	–	–	–	√	–	–	–	–	–
[43]	Feng W, *et al*.	2012	KDD	–	–	–	–	–	–	√	–	–
[8]	Xia X, *et al*.	2013	MSR	√	–	–	–	–	–	–	–	–
[41]	Wang H, *et al*.	2013	IJCAI	√	–	–	–	–	–	–	–	–
[34]	Wang S, *et al*.	2014	ICSME	√	–	–	–	–	–	–	–	–
[47]	Wang H, *et al*.	2015	AAAI	√	–	–	–	–	–	–	–	–
[11]	Gong Y, *et al*.	2016	IJCAI	√	–	–	–	–	–	–	–	–
[12]	Li Y, *et al*.	2016	COLING	√	–	–	–	–	–	–	–	–
[19]	Wu Y, *et al*.	2016	CIKM	√	–	–	–	–	–	–	–	–
[22]	Li J, *et al*.	2016	IJCNN	√	–	–	–	–	–	–	–	–
[27]	Rawat YS, *et al*.	2016	MM	–	–	√	–	–	–	–	–	–
[28]	Huang H, *et al*.	2016	COLING	√	–	–	–	–	–	–	–	–
[48]	Yamasaki T, *et al*.	2017	IJCAI	√	–	–	–	–	–	–	√	–
[21]	Zhang Q, *et al*.	2017	IJCAI	–	–	√	–	–	–	–	–	–
[35]	Zhou P, *et al*.	2017	SANER	√	–	–	–	–	–	–	–	–
[49]	Nguyen H, *et al*.	2017	PAKDD	–	√	–	–	–	–	–	–	√
[14]	Hassan HAM, *et al*.	2018	RecSys	√	–	–	–	–	–	–	–	–
[20]	Wu Y, *et al*.	2018	CIKM	√	–	–	–	–	–	–	–	–
[24]	Gong Y, *et al*.	2018	Neurocomputing	–	–	√	–	–	–	–	–	–
[33]	Wang S, *et al*.	2018	ESE	√	–	–	–	–	–	–	–	–
[7]	Tang S, *et al*.	2019	AAAI	√	–	–	–	–	√	–	–	–
[13]	Sun B, *et al*.	2019	TLT	√	–	–	–	–	–	–	–	–
[15]	Shi X, *et al*.	2019	DASFAA	√	–	–	–	–	√	–	–	–
[16]	Zhang S, *et al*.	2019	AAAI	–	–	√	–	–	–	–	–	√
[17]	Wei Y, *et al*.	2019	MM	–	–	–	√	–	–	–	√	–
[38]	Gao J, *et al*.	2019	CIKM	√	–	–	–	–	√	–	–	–
[23]	Li M, *et al*.	2019	CIKM	–	–	–	√	–	–	√	–	√

表 2 常见标签推荐方法分类(续)

参考文献序号	作者	发表年份	发表刊物(简称)	基于内容				基于标签相关性			基于用户偏好	
				文本内容	图片内容	图文内容	视频内容	标签共现	标签结构	标签语义	交互关系	用户表示
[36]	Maity SK, *et al*.	2019	ECIR	√	–	–	–	–	–	–	–	√
[50]	Wang X, *et al*.	2019	MM	√	–	–	–	–	–	–	√	–
[51]	Lima E, *et al*.	2019	TOIT	√	–	–	–	–	√	–	–	–
[52]	Tonge A, *et al*.	2019	TIST	–	√	–	–	–	–	–	–	–
[18]	Quintanilla E, *et al*.	2020	TMM	–	√	–	–	–	–	–	–	√
[25]	Zuin G, *et al*.	2020	IJCNN	–	√	–	–	–	–	–	–	–
[6]	Wang X, *et al*.	2020	WWW	√	–	–	–	–	–	–	√	–
[53]	Chen X, *et al*.	2020	IJCNN	–	–	–	–	–	–	–	√	–
[4]	Lei K, *et al*.	2020	Neurocomputing	√	–	–	–	–	–	–	–	–
[54]	Khezrian N, *et al*.	2020	arXiv	√	–	–	–	–	–		–	–

针对标签推荐任务的第 1 个核心问题——项目内容的多样性, 过去的工作已经探索过: (1) 文本内容; (2) 图片内容; (3) 图文内容; (4) 视频内容. 基于内容的标签推荐方法往往将标签推荐简化为一个多标签分类问题, 即不考虑用户信息, 只考虑项目的内容信息与标签之间的关联. 由于文本内容最为常见, 且更易于提取出有效特征, 因而基于文本内容的标签推荐方法最为常见. 基于内容的标签推荐方法还可以分为基于单模态内容的方法以及基于多模态内容的方法: 基于单模态内容的方法往往以纯文本或图片信息为内容对象, 而基于多模态内容的方法往往利用了文本、图像和音频等多个模态的信息.

针对标签推荐任务的第 2 个核心问题——标签之间的相关性, 过去的工作已经探索过: (1) 标签共现; (2) 标签结构; (3) 标签语义. 其共同特点便是通过挖掘标签之间的相关性提升模型的标签推荐性能, 其中, 基于标签共现的方法主要通过条件概率提取出标签之间的共现关系; 基于标签结构的方法主要考虑在模型中显式得构造标签结构, 以捕获标签相关性; 基于标签语义的方法则主要通过引入标签语义向量, 隐式地利用到标签之间的相关关系.

针对标签推荐任务的第 3 个核心问题——用户偏好的差异性, 我们可以将其分为两个方面: (1) 基于交互关系的方法; (2) 基于用户表示的方法. 如果不考虑用户偏好, 那么标签推荐问题往往可以简化为一个多标签分类问题. 然而, 标签推荐的研究对象是用户、项目和标签这 3 种主体及主体之间的相互作用. 不同的用户可能关注于不同领域的项目, 并且存在着不同的标签习惯. 因此, 忽略用户偏好虽然简化了模型, 但同时势必会造成部分关键信息的缺失, 从而影响模型的推荐能力. 虽然以往的大部分标签推荐方法还未关注到用户偏好, 但已有部分工作开始研究基于用户偏好的标签推荐方法, 并且取得了突出的成绩[16,23]. 其中: 基于用户表示的方法从用户的 ID、历史标签、历史项目信息提取出用户表示, 之后与项目表示合并得到用户和项目的联合表示; 而基于交互关系的方法则更进一步, 考虑到了用户、标签和项目三者之间的交互关系.

在本文中, 为了便于公式描述, 我们统一约定粗体的大写字母表示矩阵(如 $\boldsymbol{A}$), 粗体的小写字母表示向量(如 $\boldsymbol{a}$), 非粗体的大写字母或小写字母表示标量(如 a 或 A), $\odot$表示哈达玛积(Hadamard product)符号, $\mathbb{R}$表示实数空间.

3 基于内容的标签推荐方法

基于内容的标签推荐方法往往将标签推荐简化为一个多标签分类问题, 即不考虑用户信息, 只考虑项目的内容信息与标签之间的关系, 通过模型捕获此种关联关系. 在训练阶段, 需要输入项目的内容信息(如文本、图片、视频、音频等)及其相对应的标签集. 模型可以学习到项目内容与标签的隐式关联关系. 在测试阶段, 旨在为每一个新的项目推荐一个或多个标签. 基于内容的标签推荐方法可以分为基于单模态内容的方法以及基于多模态内容的方法: 基于单模态内容的方法往往以纯文本或图片信息为内容对象, 而基于多模态内容的方法往往利用了文本、图像和音频等多个模态的信息.

3.1 单模态内容

3.1.1 文本内容

文本内容是标签推荐任务下最常见的内容信息, 并且由于文本信息更易提取出有效的特征, 而基于文本内容的标签推荐方法最为常见. 因此, 如何在标签推荐场景下学习到一个更好的文本表示, 一直是研究者们关注的核心问题. 而在标签推荐场景下, 为了提取到有效的文本表示, 需要考虑到 3 个层次级别的信息.

(1) 词级别的信息: 考虑到词级别的信息是由于每个词在文档中的重要程度不一样, 同时与标签之间的相关性也存在显著的差别. 如图 1 所示: 一些关键词如标题中的“Random Forest”和标签“decision-tree”, “random-forest”之间具有很强的相关性; 而一些常见词如标题中的“Example”与标签之间显然不具备很强的相关性. 如果能够尽可能准确地提取出文本内容中的关键词, 那么后续任务在学习文本与标签之间关联关系时便会更加具有针对性, 从而提升标签推荐的效果. 因此, 对词级别的信息进行有效的提取是标签推荐场景下重要的一环;

(2) 句子级别的信息: 如果仅仅考虑到词级别的信息, 那么便忽略了蕴含有丰富语义的文本顺序信息[7]. 文本顺序信息是语义信息的重要组成部分, 如果不考虑到文本顺序, 那么语义往往会产生与原义不符的情况. 比如“Beijing belongs to China”和“China belongs to Beijing”, 在不考虑文本顺序的前提下, 前者和后者便具有同样的文本表示. 但显然两者的语义不同. 并且与词级别的信息类似, 每个句子在文档中发挥的作用也不一样. 如图 1 所示, 描述中一共包含有 3 个句子: 第 1 个句子表述了详细的问题, 第 2 个句子在讲具体的操作, 最后一个句子描述了最终的需求. 其中, 每个句子对于文档的重要程度不同. 因此, 如何从句子级别对文本内容进行特征提取, 也是标签推荐场景下的重要研究对象;

(3) 文档级别的信息: 如前文所述, 由于标签推荐场景的多样性, 不同标签推荐任务的文档结构具有很大的差别. 如图 1 所示, 该问答网站上的文档结构就分为标题和描述两部分. 其中: 标题的内容更加精炼, 描述的内容则更加具体. 因此, 对待标题和描述应该采用不同的处理方式, 从而更有效地提取整个文档级别的信息. 标签推荐场景下除了有标题和描述这样的文档结构[7]外, 还包含了问题和答案[13]这样的文档结构以及具有层级性[14]的文档结构等.

早期的传统方法在进行文本特征提取时主要关注于提取词级别的信息, 如 Xia 等人[8]将文本内容看作“词袋(bag of words, BoW)”, 即不考虑文本内容的顺序, 在进行预处理后直接利用单词出现的频次来表示文本内容, 之后将特征输入多标签分类器实现多标签学习的过程. Wu 等人[19,20]同样使用每个单词出现的频次作为文本表示, 提出了一个类似于主题模型[55]的生成模型, 可以挖掘标签词和文档内单词的共现关系. 具体来说, 其将文档的标签看作主题, 文档内的单词依据标签-词分布或标签自身进行生成. Song 等人[2]也利用文档词频作为文档表示, 其将每个文档和其对应的内容表示为一个无向有权的二部图, 权重即为文档中单词出现的次数. 具体来说, 其构建了文档和单词以及文档和标签两个二部图, 采用谱递归嵌入(spectral recursive embedding)[56]方法进行图上的聚类. 之后, 对每个类别内的标签进行重要度的排序. 在测试阶段, 其根据每个新文档的文本特征将其进行分类, 之后根据该类下的标签重要度进行标签推荐.

然而, 以上采用词频来提取词级别信息的方法无法准确地反映出每个词对文档的重要程度. 因为一些常见词往往具有很高的词频(如图 1 中的“example”), 但是却蕴含有较少的信息, 对文档的重要程度低, 因而不应该给予其过高的关注度. 为了缓解此问题, 研究者们[8,57]利用文档的词频-逆文档频率(term frequency-inverse document frequency, TF-IDF)特征作为文本表示. 由于同时考虑到了文档的词频和逆文档频率, 该特征缓解了词频特征对常见词的过度关注, 可以更为准确地提取出每个词对文档的重要程度, 从而获取到更有效的文本表示. 但是以上几种方法在进行文本特征提取时只关注到了词级别的信息, 其将每篇文章看作一个词袋, 从而忽略了文本内容的顺序结构, 进而忽略了文本内容句子级别的信息.

为了同时考虑到文档词级别的信息和句子级别的信息以获得更有效的文本表示, 研究者们开始采用深度学习方法[10,12−14]学习文本内容. 由于考虑到了文本内容的顺序结构, 采用了基于循环神经网络(recurrent

neural network, RNN)[12−14]和卷积神经网络(convolutional neural network, CNN)[10,22]的方法进行文本顺序信息的捕获. 深度学习方法通过对句子级别信息的捕获, 得到了更好的文本表示. 并且相较于传统方法, 深度学习方法具有建模更加灵活的特点, 可以在一定程度上克服标签推荐场景丰富、项目内容多样性的特点. 因而基于文本内容的深度学习标签推荐方法数量众多, 其统一框架如图 4 所示, 主要包含了词表示层(word representation layer)、特征提取器(feature extractor)以及多标签分类器(multi-label classifier)这 3 个部分. 模型输入通常为一个长度为 n 的文本序列(text sequence) $\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_n$, 其中, 每一个 $\boldsymbol{w}_i$ 代表一个单词. $\boldsymbol{w}_i$ 为一个长度为 V_{text} 的 One-Hot 向量, V_{text} 为文本词表的大小. 通过一个词表示层, 将原始的单词 One-Hot 向量转化为其对应的词向量(word vector)表示 $\boldsymbol{e}_1,\boldsymbol{e}_2,\ldots,\boldsymbol{e}_n\in\mathbb{R}^{D_{embedding}}$, 其中, $D_{embedding}$ 是词表示的维度. 随后, 将得到的词表示序列输入到文本特征提取器中, 提取到有效的文本表示. 最后, 将特征提取器得到的文本表示输入到一个多标签分类器中, 得到所有 k 个标签的伪概率(pseudo probability) $t_1,t_2,\ldots,t_k\in[0,1]$, 其中, k 为标签词表的大小. 一般采用交叉熵(cross entropy)函数作为该模型的损失函数.

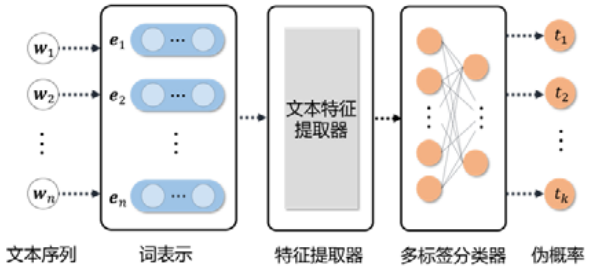


图 4 基于文本内容的深度学习标签推荐方法框架

以下通过该框架的词表示层、特征提取器和多标签分类器对此框架下的不同模型进行阐述.

(1) 词表示层

在词表示层, 模型主要关注词级别信息的特征提取. 文本序列中的每一个单词 $\boldsymbol{w}_i$ 都可以看作是一个 One-Hot 向量, 向量维度为文本词表的大小 V_{text}. 如果直接将该 One-Hot 向量作为词向量表示进行后续任务的话, 会遇到两个问题: ① 模型容易受“维数灾难”的困扰, 计算量会随着向量维度的增加而呈指数倍增长; ② 单词的 One-Hot 向量无法表达单词丰富的语义信息. 由于每个单词的词向量之间都存在正交关系, 因而无法表达单词之间的相关关系, 从而无法使模型显式地捕获到词级别的信息. 因此, 研究者们通常采用词向量技术[57]来表示单词, 获取单词词级别的信息. 具体来说, 通过一个嵌入矩阵(embedding matrix) $\boldsymbol{E}$, 每个单词被投影到一个低维稠密的实值向量空间, 从而获取每个词对应的词表示 $\boldsymbol{e}_i$:

$$\boldsymbol{e}_i=\boldsymbol{w}_i\boldsymbol{E} \tag{1}$$

这里, 嵌入矩阵 $\boldsymbol{E}$ 的维度即为 $V_{text}\times D_{embedding}$; $D_{embedding}$ 即为低维嵌入空间的大小, 同时也是词表示 $\boldsymbol{e}_i$ 的长度. 通过这样的转换, 在深度网络学习的阶段, 便可以学习到每个单词对应的词向量表示, 从而学习到文本内容的词级别信息. 在词表示层, 不同方法的区别主要在于嵌入矩阵的设置不同. 可以选择采取随机初始化的嵌入矩阵或预训练好的嵌入矩阵.

其中, 由于预训练好的嵌入矩阵包含了预训练语料丰富的语义信息, 因而在当前标签推荐任务的语料与预训练语料的语义接近时, 选择预训练好的嵌入矩阵会有助于词级别信息的提取. 比如, Kai 等人[4]采用 Word2Vec[57]预训练好的词向量初始化词表示层, 并在训练过程中不断进行更新, 以期获得更符合当前标签推荐场景的词表示. 由于在预训练阶段 Word2Vec 可以根据词的上下文信息学到词与词之间的相似性以及一些内在联系, 因而可以赋予模型较好的初始化词表示. 由于标签推荐场景的多样性, 其词表中的部分单词可能未出现在预训练词表中,在此情况下, 一般对于未出现在预训练词表中的词使用随机初始化的方式[10]. 而另一种预训练模型 GloVe[58]不仅考虑到了词的上下文信息, 还考虑到了词在整个输入语料的全局信息, 因而还可

以捕获到词的全局共现关系, 从而提取到更有效的词级别信息. Hassan 等人[14]便利用了 GloVe 预训练好的词向量初始化词表示层, 并在训练过程中不断进行更新, 以获取更有效的词表示. 在当前语料较少, 并且与预训练语料语义接近的情况下, Sun 等人[13]利用 GloVe 预训练好的词向量初始化词表示层, 并且不进行词表示层的更新, 避免由于训练资料的不足破坏原有的语义.

随机初始化的嵌入矩阵则更加适合专业性和领域性较强的标签推荐任务, 因为在此种场景下, 预训练语料和当前语料的语义不一致, 甚至存在较大的语义鸿沟, 从而导致预训练好的嵌入矩阵无法为模型提供较为匹配的指导. 比如在专业性较强的软件信息问答网站场景下进行标签推荐时, Huang 等人[28]便直接使用了随机初始化的嵌入矩阵, 并在训练过程中不断进行更新, 在端到端的深度学习训练框架下, 模型会自动学习到每个词所对应的词表示, 从而捕获到匹配当前标签推荐场景的词级别信息.

(2) 特征提取器

在特征提取器部分, 模型主要进行句子级别信息和文档级别信息的提取. 为了捕获到标签推荐场景下句子级别的信息, 研究者们普遍采用基于卷积神经网络、循环神经网络或胶囊网络(capsule networks)[59,60]的方法进行特征提取. Gong 等人[11]采用卷积神经网络学习文本表示, 并且引入了局部注意力机制[61]捕获句子中的重要部分. 具体来说, 其方法分为全局和局部两个通道: 全局通道直接利用卷积神经网络作为特征提取器提取文本内容全局范围内的特征; 而局部通道利用一个长度固定的时间窗获取每个时间窗内各个单词的注意力权重, 选取权重大于一定阈值的单词作为触发词, 最后利用一个卷积层将全局和局部的信息进行融合. 该方法中的全局部分利用卷积捕获到了文本的顺序信息, 其局部部分通过对触发词的关注, 捕获到了词级别的信息.

然而, 基于卷积神经网络的方法却有着无法考虑到文本内容相对位置信息的缺点. 胶囊网络在综合了卷积神经网络局部特征提取优点的同时, 考虑到了其缺失的相对位置等其他信息. Kai 等人[4]便将胶囊网络作为其模型特征提取器的一部分, 捕获到了更加丰富的文本结构信息. 具体来说, 文本内容首先通过词表示层和注意力层, 注意力层的目的在于依据注意力分数, 提取文本中的重要的单词. 之后, 文本内容的隐表示再次通过一个卷积层, 用来学习文本内容的局部关联关系. 卷积层的输出结果被输入到一个胶囊网络中, 模型从而可以学习到文本内容相对位置的关联关系.

基于卷积神经网络或胶囊网络的方法只能提取到文本的局部而非全局的顺序信息. 为了进一步捕获文本内容的全局顺序信息, 研究者们采用基于循环神经网络的方法进行特征提取. Li 等人[12]提出了一种基于注意力机制的长短期记忆网络(long short term memory, LSTM)模型, 通过引入文本内容的主题分布, 指导特征提取器学习到更符合当前文档主题的文本表示. 具体来说, 该方法通过主题模型[62]得到当前文本的主题分布 $\boldsymbol{\theta} \in \mathbb{R}^{V_{topic}}$, 其中, V_{topic} 是主题数. 而 $\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_n \in \mathbb{R}^{D_{hidden}}$ 是 LSTM 提取到的 n 个隐状态, 其中, D_{hidden} 为 LSTM 隐状态的维度. 则当前文本的主题注意力分数 a_j 可以通过下面的公式得到:

$$g_j = \boldsymbol{v}^{\mathrm{T}} \tanh(\boldsymbol{W}\boldsymbol{\theta} + \boldsymbol{U}\boldsymbol{h}_j) \tag{2}$$

$$a_j = \frac{\exp(g_j)}{\sum_{j'=1}^{n} \exp(g_{j'})} \tag{3}$$

其中, $\boldsymbol{W} \in \mathbb{R}^{D_{hidden} \times V_{topic}}$, $\boldsymbol{U} \in \mathbb{R}^{D_{hidden} \times D_{hidden}}$, $\boldsymbol{v} \in \mathbb{R}^{D_{hidden}}$ 是可训练的参数矩阵. 其在计算注意力分数时, 将主题分布的信息压缩到隐状态的空间中, 从而得到主题注意力分数 a_j. 循环神经网络适合用于序列建模但无法以并行的方式提取特征. 另一方面, 卷积神经网络对局部响应的学习较好, 但缺乏学习长期关联的能力. 因此, 一些研究者[63,64]结合了循环神经网络和卷积神经网络进行特征提取, 旨在同时获取文档的局部特征和全局特征. 比如: Li 等人[22]先用卷积神经网络对文本内容做局部的特征提取, 在得到不同时间窗下的多个特征图后, 使用 LSTM 提取其上下文依赖关系, 代替了普通卷积神经网络的池化操作, 从而得到原文档最终的文档表示; 而 Lai 等人[34]则先用一个双向的循环神经网络获取文本的顺序和逆序信息后, 将其输入到一个最大池化层, 得到最终的文档表示.

为了进一步捕获到标签推荐场景下文档级别的信息, 研究者们普遍针对不同的标签推荐场景设计不同的

网络架构进行特征提取. 如前文所述, 由于标签推荐场景的多样性, 不同标签推荐任务的文档结构具有很大的差别. 如文献[13]的研究对象是英语试题中的选择题, 其文档结构包含了选择题的题干和选项两部分, 对待这两部分应该采用不同的处理方式和重视程度, 这样才能更好地提取到有效的文本特征. 为了捕获到文本内容文档级别的信息, Sun 等人[13]采用了双向的 LSTM 结合位置注意力机制作为特征提取器, 从而得到了更符合标签推荐场景的文本表示. 具体来说, 在注意力机制部分, 由于题干和选项有着不同的重要程度, 因此作者采用了一个重要度向量 $\boldsymbol{p}$ 指示一段文本中的每个词的重要程度:

$$\boldsymbol{p}=(p_1,p_2,\ldots,p_n)\in\{P_A,P_N\}^n \tag{4}$$

其中, P_A 和 P_N 是两个超参数, 分别表示选项词的重要度和题干词的重要度. 作者认为, 选项词蕴含的判别式信息比题干词更加丰富, 因此假设 $P_A>P_N$. 接下来, 使用该重要度向量 $\boldsymbol{p}$ 指导注意力权重 a_j 的生成:

$$g_j=\boldsymbol{p}^T\boldsymbol{w} \tag{5}$$

$$a_j=\frac{\exp(g_j)}{\sum_{j'=1}^{n}\exp(g_{j'})} \tag{6}$$

其中, $\boldsymbol{w}\in\mathbb{R}^n$ 为注意力机制中可训练的参数向量. 与文献[13]的情形类似, 文献[14]的研究对象是学术论文, 由于学术论文摘要部分的规范性, 其往往由几个具有逻辑关系的句子组成. Hassan 等人[14]便考虑到了这样具有层次性的文档结构, 采用了双向的门控循环单元(gated recurrent unit, GRU)结合层级注意力机制[65]作为特征提取器. 在层级注意力部分, 模型分别在“单词-句子”层面捕获每个单词对句子的重要程度, 以及在“句子-文档”层面捕获每个句子对文档的重要程度, 从而使得模型更加关注文档中的重要句子以及句子中的重要单词, 进而学习到更好的文本表示. 具体来说, 当前文档的第 j 个句子对句子内第 t 个单词的注意力分数 $a_{j,t}$ 为

$$g_{j,t}=\boldsymbol{v}_w^{\mathrm{T}}\tanh(\boldsymbol{W}_w\boldsymbol{h}_{j,t}+\boldsymbol{b}_w) \tag{7}$$

$$a_{j,t}=\frac{\exp(g_{j,t})}{\sum_{t'=1}^{n}\exp(g_{j,t'})} \tag{8}$$

其中, $\boldsymbol{W}_w$, $\boldsymbol{b}_w$ 和 $\boldsymbol{v}_w$ 为“单词-句子”级别注意力机制的可学参数, $\boldsymbol{h}_{j,t}$ 是第 j 个句子的第 t 个单词通过双向 GRU 网络进行特征提取后得到的隐表示. 同样, 在“句子-文档”级别, 当前整个文档对第 j 个句子的注意力分数 $\boldsymbol{a}_j$ 为

$$g_j=\boldsymbol{v}_s^{\mathrm{T}}\tanh(\boldsymbol{W}_s\boldsymbol{h}_j+\boldsymbol{b}_s) \tag{9}$$

$$a_j=\frac{\exp(g_j)}{\sum_{j'=1}^{n}\exp(g_{j'})} \tag{10}$$

其中, $\boldsymbol{W}_s$, $\boldsymbol{b}_s$ 和 $\boldsymbol{v}_s$ 为“句子-文档”级别注意力机制的可学参数. 其中, $\boldsymbol{h}_j$ 是第 j 个句子通过双向 GRU 网络以及上一层注意力机制后得到的隐表示. 通过这样的方式, 模型通过对文档级别信息的提取, 得到了更有效的文本表示.

(3) 多标签分类器

如图 4 所示: 在利用特征提取器获取到有效的文本表示后, 大部分方法都采用了 MLP 作为多标签分类器. 顶层的神经元个数即为标签推荐任务中标签词表的大小 k, 预测结果即为每个标签的伪概率 $\boldsymbol{t}=(t_1,t_2,\ldots,t_k)$, 其中, $t_i\in[0,1]$. 在标签推荐领域, 主要有两类多标签分类器, 其区别主要在于顶层激活函数以及模型损失函数的不同. 第 1 类多标签分类器[27]在顶层构建了由 Sigmoid 函数激活的全连接层, $\boldsymbol{m}_i$ 为第 i 个文档经过特征提取器之后的文本表示, $\boldsymbol{W}$ 为第 1 类多标签分类器顶层全连接层的参数矩阵. 其预测结果如下:

$$\boldsymbol{t}_i=Sigmoid(\boldsymbol{W}\boldsymbol{m}_i) \tag{11}$$

其中, $\boldsymbol{t}_i=(t_{i,1},\ldots,t_{i,j},\ldots,t_{i,k})$. $t_{i,j}\in[0,1]$是模型预测得到的第 i 个文档是否含有第 j 个标签的伪概率. 第 1 类多标签分类器使用的损失函数是 k 个二分类交叉熵损失函数之和:

$$L=-\sum_{i=1}^{N}\sum_{j=1}^{k}(y_{i,j}\log(t_{i,j}))+(1-y_{i,j})\log(1-t_{i,j}) \tag{12}$$

其中, N 是训练集的文档数量, k 是标签的类别个数, $y_{i,j}\in\{0,1\}$表示实际上第 i 个文档是否含有第 j 个标签.

然而, 第 1 类多标签分类器使用 *Sigmoid* 函数作为顶层激活函数的方法将多标签分类问题看作了 k 个二分类问题, 没有显式地考虑到标签之间的关联关系. 因此, 第 2 类多标签分类器使用 *Softmax* 函数作为顶层激活函数. 由于 *Softmax* 函数可以为预测结果增加和为 1 的约束, 因而在一定程度上考虑到了顶层标签之间的相关性, 从而更加适合标签推荐的应用场景. 因此, 大部分基于深度学习的标签推荐方法[11,12,14,16,22]都使用了第 2 类多标签分类器, 即: 在得到有效的文本表示 $\boldsymbol{m}_i$ 之后, 在顶层构建了由 *Softmax* 函数激活的全连接层, 并且使用交叉熵损失函数:

$$\boldsymbol{t}_i=Softmax(\boldsymbol{W}\boldsymbol{m}_i) \tag{13}$$

$$L=-\sum_{i=1}^{N}\sum_{j=1}^{k}(y_{i,j}\log(t_{i,j})) \tag{14}$$

3.1.2 图片内容

随着图片分享应用(如 Instagram)的普及, 越来越多的信息以图片的方式进行传递, 在图片分享场景下的标签推荐也逐渐变成一个热点问题. 由于卷积神经网络在图像特征提取方面的迅猛发展, 基于图片内容的标签推荐方法大都采用了基于卷积神经网络的方法进行特征提取. 并且与上一节类似, 大部分基于图片内容的方法同样将标签推荐看作一个多标签分类问题[16−18].

此类方法的统一框架如图 5 所示, 主要包含了视觉特征提取器和多标签分类器两部分. 在多标签分类器部分, 此类方法与上一节一致. 而在视觉特征提取器部分, Nguyen 等人[49]最早研究该问题, 其采用了一个简单的卷积神经网络进行图片内容的特征提取. Zuin 等人[25]使用了基于卷积神经网络的自编码器(autoencoder, AE)学习画作的视觉特征, 并且利用学到的隐特征进行后续的标签推荐. Quintanilla 等人[18]则是采用了 ResNet50[66]提取视觉特征, 并且在特征提取的过程中采用了对抗学习的思想, 利用真实标签指导特征提取器学习到更符合标签推荐的有效视觉特征, 最后将对抗学习损失和多标签分类器损失进行加权得到最终的损失函数.

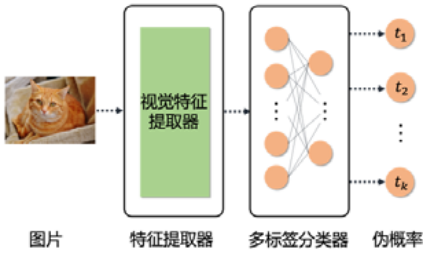


图 5 基于图片内容的深度学习标签推荐方法框架

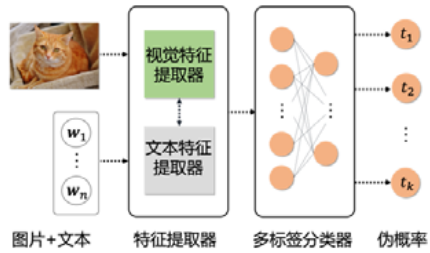


图 6 基于图文内容的深度学习标签推荐方法框架

3.2 多模态内容

3.2.1 图文内容

在图片分享应用(如 Instagram, Twitter)中, 用户分享的内容信息往往同时包含了图片和文本信息, 而这些信息都会影响其对应标签的生成. 因此, 在图文场景下进行标签推荐时, 需要同时考虑到文本特征和视觉特征. 在该场景下, 研究者们主要关注于如何联合提取内容的视觉和文本特征[16,21,27], 特别是关注两个模态信息之间的交互[67]. 同时, 其大都采用了基于深度学习的方法, 其统一框架如图 6 所示, 主要包含了联合特征提取器和多标签分类器两部分.

Rawat 等人[27]在进行标签推荐时, 同时利用到图片的视觉信息和上下文信息. 其分别利用 AlexNet[68]提取图片的视觉信息得到隐特征 $\boldsymbol{u}$, 利用两个全连接层提取上下文信息得到隐特征 $\boldsymbol{h}$. 在得到两部分的隐特征后,

直接将其拼接在一起, 作为最终的内容表示, 随后进行标签推荐任务. 然而, 对于图文信息这样的多模态信息, 每个模态之间存在着相互影响, 直接将隐特征进行拼接的方法并不能很好地实现特征融合[22,69]. 因此, Zhang 等人[21]使用了基于注意力机制的深度学习方法, 在进行特征融合时考虑到两个模态之间的交互. 具体来说, 在通过 VGG 网络[70]提取到视觉特征矩阵 $\boldsymbol{U}\in\mathbb{R}^{d\times m}$, 通过 LSTM 和最大池化操作提取到文本特征向量 $\boldsymbol{h}\in\mathbb{R}^{d}$ 后, 其采用了基于文本特征 $\boldsymbol{h}$ 的视觉注意力机制获取与文本特征 $\boldsymbol{h}$ 交互过的视觉特征 $\tilde{\boldsymbol{u}}$. 此处, d 为视觉(或文本)特征隐空间的大小, m 为每张图片中区域的个数:

$$\boldsymbol{U}_h=\tanh(concat(\boldsymbol{W}_u\boldsymbol{U},\boldsymbol{W}_h\boldsymbol{h})) \tag{15}$$

$$\boldsymbol{a}_u=Softmax(\boldsymbol{w}_{a_u}\boldsymbol{U}_h+\boldsymbol{b}_{a_u}) \tag{16}$$

$$\tilde{\boldsymbol{u}}=\sum_{i=1}^{m}a_u^i\boldsymbol{u}^i \tag{17}$$

其中, $\boldsymbol{W}_u\in\mathbb{R}^{r\times d}$ 和 $\boldsymbol{W}_h\in\mathbb{R}^{r\times d}$ 为参数矩阵, $\boldsymbol{U}_h\in\mathbb{R}^{2r\times m}$; $concat(\boldsymbol{Q},\boldsymbol{q})$符号表示将矩阵 $\boldsymbol{Q}$ 和向量 $\boldsymbol{q}$ 进行拼接, 具体而言, 将矩阵 $\boldsymbol{Q}$ 的每一列和向量 $\boldsymbol{q}$ 进行拼接; $\boldsymbol{w}_{a_u}\in\mathbb{R}^{1\times 2r}$ 和 $\boldsymbol{b}_{a_u}\in\mathbb{R}^{1\times 2r}$ 为参数向量; $\boldsymbol{a}_u\in\mathbb{R}^{m}$ 为对应的注意力权重向量, 代表着通过与文本内容交互之后, 模型对图片每个区域的关注程度不同. 最终, 通过对每个区域的特征按照注意力分数进行加权, 得到与文本特征交互后的视觉特征 $\tilde{\boldsymbol{u}}\in\mathbb{R}^{d}$.

接着, 模型采用了基于视觉特征 $\tilde{\boldsymbol{u}}$ 的文本注意力机制, 获取与视觉特征 $\tilde{\boldsymbol{u}}$ 交互过的文本特征 $\tilde{\boldsymbol{h}}$. 具体而言, 其首先将文本内容通过 LSTM 得到的文本特征矩阵 $\boldsymbol{H}\in\mathbb{R}^{d\times n}$, 此处, n 为文本序列长度; 随后, 通过视觉特征 $\tilde{\boldsymbol{u}}$ 的引导得到交互矩阵 $\boldsymbol{H}_u$. 根据该交互矩阵得到文本特征的注意力权重:

$$\boldsymbol{H}_u=\tanh(concat(\boldsymbol{W}_{\tilde{u}}\tilde{\boldsymbol{u}},\boldsymbol{W}_H\boldsymbol{H})) \tag{18}$$

$$\boldsymbol{a}_h=Softmax(\boldsymbol{w}_{a_h}\boldsymbol{H}_u+\boldsymbol{b}_{a_h}) \tag{19}$$

$$\tilde{\boldsymbol{h}}=\sum_{i=1}^{n}a_h^i\boldsymbol{h}^i \tag{20}$$

同样地: $\boldsymbol{W}_{\tilde{u}}\in\mathbb{R}^{r\times d}$ 和 $\boldsymbol{W}_H\in\mathbb{R}^{r\times d}$ 为参数矩阵, $\boldsymbol{H}_u\in\mathbb{R}^{2r\times n}$; $\boldsymbol{w}_{a_h}\in\mathbb{R}^{1\times 2r}$ 和 $\boldsymbol{b}_{a_h}\in\mathbb{R}^{1\times 2r}$ 为参数向量; $\boldsymbol{a}_h\in\mathbb{R}^{n}$ 为对应的注意力权重向量, 代表着通过与视觉内容交互之后, 模型对文本序列每个单词的关注程度不同. 最终, 通过对每个单词的特征按照注意力分数进行加权, 得到与视觉特征交互后的文本特征 $\tilde{\boldsymbol{h}}\in\mathbb{R}^{d}$. 输入多标签分类器的特征 $\boldsymbol{f}$ 为两部分交互特征 $\tilde{\boldsymbol{u}}$, $\tilde{\boldsymbol{h}}$ 之和:

$$\boldsymbol{f}=\tilde{\boldsymbol{u}}+\tilde{\boldsymbol{h}} \tag{21}$$

虽然以上方法考虑到了文本特征与视觉特征的交互, 然而其假设了在特征交互时文本特征先影响视觉特征, 之后, 视觉特征再影响文本特征. 而在直观上, 文本特征和视觉特征不存在特征交互的先后关系. 因此, Zhang 等人[16]采取了共行注意力机制[71]进行文本和视觉信息的同步提取, 从而得到更有效的混合特征. 具体来说, 在分别利用 LSTM 以及 VGG 网络提取到的视觉特征矩阵 $\boldsymbol{U}\in\mathbb{R}^{d\times m}$ 和文本特征矩阵 $\boldsymbol{H}\in\mathbb{R}^{d\times n}$ 后, 其建立了一个亲和度矩阵 $\boldsymbol{C}\in\mathbb{R}^{m\times n}$:

$$\boldsymbol{C}=\tanh(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{W}_b\boldsymbol{H}) \tag{22}$$

其中, $\boldsymbol{W}_b\in\mathbb{R}^{d\times d}$ 为可学参数. 之后, 模型利用亲和度矩阵 $\boldsymbol{C}$ 与视觉特征 $\boldsymbol{U}$ 进行交互, 得到交互后的文本特征矩阵 $\boldsymbol{H}_u$; 同理, 可以得到交互后的视觉特征矩阵 $\boldsymbol{U}_h$:

$$\boldsymbol{H}_u=\tanh(\boldsymbol{W}_h\boldsymbol{H}+(\boldsymbol{W}_u\boldsymbol{U})\boldsymbol{C}) \tag{23}$$

$$\boldsymbol{U}_h=\tanh(\boldsymbol{W}_u\boldsymbol{U}+(\boldsymbol{W}_h\boldsymbol{H})\boldsymbol{C}^T) \tag{24}$$

其中, $\boldsymbol{W}_u$ 和 $\boldsymbol{W}_h$ 为可学参数, 维度与参数 $\boldsymbol{W}_b$ 一致. 在该方法中, 交互后的视觉特征矩阵 $\boldsymbol{U}_h$ 和文本特征矩阵 $\boldsymbol{H}_u$ 的捕获同步进行, 摒弃了文献[21]关于特征交互先后顺序的不合理假设, 因而可以提取到更有效的交互特征. 接下来, 通过交互矩阵 $\boldsymbol{U}_h$ 和 $\boldsymbol{H}_u$ 可以得视觉特征注意力权重 $\boldsymbol{a}_u$ 和文本特征注意力权重 $\boldsymbol{a}_h$:

$$\boldsymbol{a}_u=Softmax(\boldsymbol{W}_{a_u}^{\mathrm{T}}\boldsymbol{U}_h+\boldsymbol{b}_{a_u}) \tag{25}$$

$$\boldsymbol{a}_h=Softmax(\boldsymbol{W}_{a_h}^{\mathrm{T}}\boldsymbol{H}_u+\boldsymbol{b}_{a_h}) \tag{26}$$

其中，$\boldsymbol{W}_{a_u}^{\mathrm{T}},\boldsymbol{W}_{a_h}^{\mathrm{T}}\in\mathbb{R}^d$. 之后，通过原始的特征矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{H}$ 与对应注意力向量 $\boldsymbol{a}_u$ 和 $\boldsymbol{a}_h$ 的加权求和，可以得到与视觉特征交互后的文本特征 $\tilde{\boldsymbol{h}}$ 以及与文本特征交互后的视觉特征 $\tilde{\boldsymbol{u}}$. 最终，与 Zhang 等人[21]的方法类似，将两部分交互特征 $\tilde{\boldsymbol{u}}$, $\tilde{\boldsymbol{v}}$ 相加，即可得到最终的内容表示 $\boldsymbol{f}$:

$$\tilde{\boldsymbol{u}}=\sum_{i=1}^{m}\boldsymbol{a}_u^i\boldsymbol{u}^i \tag{27}$$

$$\tilde{\boldsymbol{h}}=\sum_{i=1}^{n}a_h^i\boldsymbol{h}^i \tag{28}$$

$$\boldsymbol{f}=\tilde{\boldsymbol{u}}+\tilde{\boldsymbol{h}} \tag{29}$$

3.2.2 视频内容

随着短视频分享平台(如 TikTok、快手)的普及，越来越多的信息以视频的形式进行传递[46,48,50,72]. 视频信息往往由图片序列、音频以及文字组成，可以看作一种包含了视觉、听觉以及文本内容的多模态信息[73]. 现阶段，基于视频内容的标签推荐方法较少，并且主要以深度学习方法为主[17,23]. 此类方法的统一框架如图 7 所示，主要包含了联合特征提取器和多标签分类器两部分. 其中，多标签分类器部分与上一节相同. 模型的主要关注点依旧在如何进行有效的特征提取. 而在视频标签推荐场景下，进行有效的特征提取需要考虑到视频信息的序列性和多模态性.

(1) 序列性: 序列性指的是视频信息先后片段之间往往具备明显的序列关系，如图 7 所示，其视频的图片序列片段均是关于猫的内容，只是每个片段之间有一些略微的区别；并且对于标签推荐场景下视频信息的每个模态——视觉、音频和文本信息都具有明显的序列性. 因此，通过考虑视频的序列性可以更好地进行视频特征提取；

(2) 多模态性: 多模态性指的是由于该场景下的视频信息可以看作多个模态信息的结合，各个模态之间具有某种程度的交互关系，比如一段关于宠物猫的短视频信息，就可能同时包含有视觉信息——猫的图片、听觉信息——猫叫声以及文字信息——关于猫的文字描述. 因此，如果能够考虑到这多个模态之间的相关性，则更有助于特征提取，从而更好地服务于后续标签推荐任务.

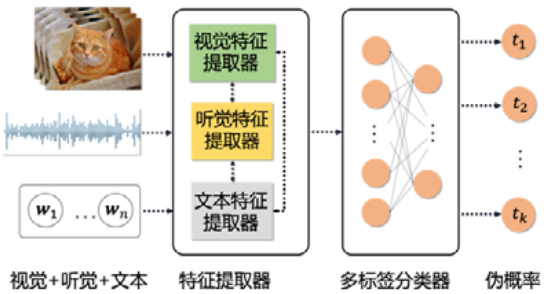


图 7 基于视频内容的深度学习标签推荐方法框架

Wei 等人[17]在进行短视频标签推荐时，在对原始数据进行特征提取时选用了 FFmpeg (http://ffmpeg.org/)进行短视频关键帧的提取和听觉片段的切分，之后使用预训练好的 ResNet50[58]提取视觉特征，使用 VGGish[74]学习听觉内容的深度特征. 除此以外，其使用 Sentence2Vector[75]提取视频对应的文本描述的文本特征. 在提取完每一个片段的视觉和听觉特征后，其采取了直接拼接的方法将其合成为最终的视觉和听觉特征. 由于视频信息具有序列性，其先后片段之间存在着关联关系. 如果进行先后片段特征的直接拼接，则忽略了视频信息的序列性，因此难以提取到更有效的视频特征.

而 Li 等人[23]则充分考虑到了视频信息提取时的序列性. 与上述方法类似，在进行原始数据的特征提取时，其分别采用了 FFmpeg 进行短视频关键帧的提取和听觉片段的切分，之后使用预训练好的 ResNet[58]提取视觉特征，使用 Librosa (https://github.com/librosa)提取每个听觉片段的特征. 除此以外，其使用 Word2Vec[48]

提取视频对应的文本单词的向量表示. 而在序列性建模部分, 其在视觉、听觉和文本的初级特征的基础上进一步采用了并行长短期记忆网络(parallel LSTMs, PLSTMs)对 3 部分初级特征进行特征提取, 以捕获视频信息的序列关系. 其 PLSTMs 结构如下所示, 其中, i 代表当前是第 i 个样本, m 代表当前信息的模态, n 代表当前的时间步, $\boldsymbol{s}_{i,n}^{m}$ 是 PLSTMs 对应的隐状态. PLSTMs 的输入 $\boldsymbol{x}_{i,n}^{m}$ 是由初级特征提取器提取后的初级特征:

$$\begin{cases} \boldsymbol{in}_{i,n}^{m}=\sigma(\boldsymbol{W}_{i}^{m}\boldsymbol{x}_{i,n}^{m}+\boldsymbol{U}_{i}^{m}\boldsymbol{s}_{i,n-1}^{m}+\boldsymbol{b}_{i}^{m}) \\ \boldsymbol{f}_{i,n}^{m}=\sigma(\boldsymbol{W}_{f}^{m}\boldsymbol{x}_{i,n}^{m}+\boldsymbol{U}_{f}^{m}\boldsymbol{s}_{i,n-1}^{m}+\boldsymbol{b}_{f}^{m}) \\ \boldsymbol{o}_{i,n}^{m}=\sigma(\boldsymbol{W}_{o}^{m}\boldsymbol{x}_{i,n}^{m}+\boldsymbol{U}_{o}^{m}\boldsymbol{s}_{i,n-1}^{m}+\boldsymbol{b}_{o}^{m}) \\ \tilde{\boldsymbol{C}}_{i,n}^{m}=\tanh(\boldsymbol{W}_{C}^{m}\boldsymbol{x}_{i,n}^{m}+\boldsymbol{U}_{C}^{m}\boldsymbol{s}_{i,n-1}^{m}+\boldsymbol{b}_{C}^{m}) \\ \boldsymbol{C}_{i,n}^{m}=\boldsymbol{f}_{i,n}^{m}\odot\boldsymbol{C}_{i,n-1}^{m}+\boldsymbol{in}_{i,n}^{m}\odot\boldsymbol{C}_{i,n}^{m} \\ \boldsymbol{s}_{i,n}^{m}=\boldsymbol{o}_{i,n}^{m}\odot\tanh(\boldsymbol{C}_{i,n}^{m}) \end{cases} \tag{30}$$

其中, $\boldsymbol{in}_{i,n}^{m}$, $\boldsymbol{f}_{i,n}^{m}$, $\boldsymbol{o}_{i,n}^{m}$ 分别代表输入门(input gate)、遗忘门(forget gate)和输出门(output gate); $\boldsymbol{C}_{i,n-1}^{m}$ 是记忆单元向量(memory cell vector); $\boldsymbol{W}_{l}^{m}$, $\boldsymbol{U}_{l}^{m}$, $\boldsymbol{b}_{l}^{m}$ 是平行 LSTM 的可学参数, 其中, $l\in\{in,f,o,C\}$; $\sigma(\cdot)$表示 *Sigmoid* 函数; $\tanh(\cdot)$表示双曲正切函数. 除此以外, 由于视频中不同片段内容的重要程度不同, 作者在平行 LSTM 层后增加了注意力机制[76], 从而可以学习到每个模态下不同片段的重要程度:

$$\boldsymbol{\theta}(i,m,n,j)=ReLU(\boldsymbol{W}_{att}^{m}\boldsymbol{s}_{i,n}^{m}+\boldsymbol{U}_{att}^{m}\boldsymbol{x}_{j}+\boldsymbol{b}_{att}^{m}) \tag{31}$$

$$\boldsymbol{\alpha}(i,m,n,j)=Softmax(\boldsymbol{\theta}(i,m,n,j)) \tag{32}$$

$$\boldsymbol{s}_{i}^{m}=\sum_{n=1}^{N}\boldsymbol{\alpha}(i,m,n,j)\boldsymbol{s}_{i,n}^{m} \tag{33}$$

其中, $\boldsymbol{W}_{att}^{m}$ 和 $\boldsymbol{U}_{att}^{m}$ 是注意力网络中的权重矩阵, $\boldsymbol{b}_{att}^{m}$ 是偏差向量(bias vector), 均为可学参数. 最终得到的 $\boldsymbol{s}_{i}^{m}$ 即为每个模态下的内容表示, 其充分考虑到了视频信息的序列性特性, 可以更好地服务于后续任务. 如上所述, 现阶段基于视频内容的标签推荐方法仅仅考虑到视频信息的序列性, 还没有考虑到视频多模态信息之间的交互性、互补性和同步性, 因而未来还有进一步挖掘的空间.

4 基于标签相关性的标签推荐方法

基于标签相关性的标签推荐方法主要分为基于标签共现的方法、基于标签结构的方法以及基于标签语义的方法. 其共同特点便是通过挖掘标签之间的相关性提升模型的标签推荐性能. 基于标签共现的方法主要通过条件概率提取出标签之间的共现关系; 基于标签结构的方法主要考虑在模型中显式得构造标签结构, 以捕获标签相关性; 基于标签语义的方法主要通过引入标签语义向量, 隐式地利用到标签之间的相关关系.

4.1 基于标签共现的方法

最早的标签推荐方法[9]便是基于标签共现的方法, 其主要利用到了每两个标签之间出现的条件概率, 因此也可称之为基于条件概率的标签推荐方法. 在训练阶段, 该方法通过统计标签之间的共现关系得到任意两个标签之间共现的条件概率. 在实际应用阶段, 其标签推荐流程如图 8 所示: 首先需要用户预先为当前项目给出少量标签; 之后根据用户给出的每一个标签, 计算以该标签为条件、共现概率最大的标签集合, 从而得到候选标签; 最后, 选取候选标签中条件概率最大的一些标签进行推荐.

与基于条件概率的标签推荐方法类似, 早期基于主题模型的方法[5]也需要用户预先为项目提供 1–5 个标签; 随后, 其将每个项目的少量标签词看作一个文档, 利用主题模型拟合该标签生成过程, 并且为每个项目确定一个主题; 最后, 对每一个项目选取其所属主题下隶属度最高的标签词进行标签推荐. 以上两种方法都需要用户预先给出少许标签才能进行推荐. 而现有的标签推荐场景往往是冷启动的, 即不需要用户预先给定标签. 因而这两种方法不符合当前标签推荐常见的应用场景.

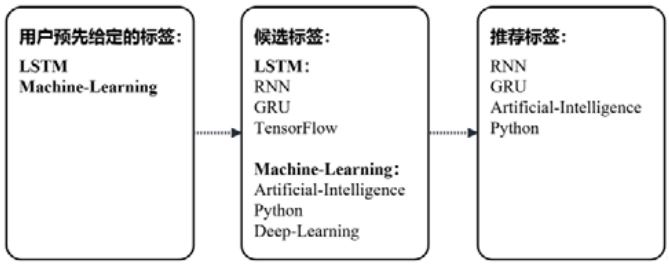


图 8 基于条件概率的标签推荐方法

4.2 基于标签结构的方法

基于标签结构的方法更多考虑的是在模型中引入显式的标签结构以捕获标签相关性.

Tang 等人[7]使用了一个序列到序列(sequence-to-sequence, Seq2Seq)[77]模型来建模标签推荐问题. 如图 9 所示, 由于采用了 Seq2Seq 结构, 其在预测每一个新标签时都会考虑到所有已经预测出的标签信息, 因此可以看作其显式地建立了标签之间的依赖关系.

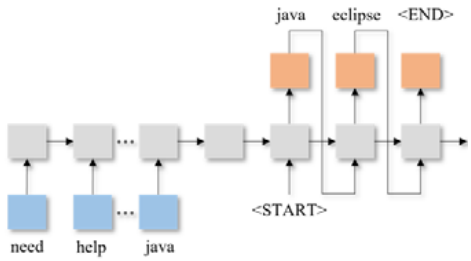


图 9 基于 Seq2Seq 模型的标签推荐方法

而 Gao 等人[38]则考虑到了标签之间的层级结构. 如图 10 所示: 其首先构建了一个多层的标签树, 标签所描述的概念从根节点到叶子节点逐渐由粗粒度到细粒度. 模型在编码器(encoder)部分进行特征提取, 在解码器(decoder)部分采用链式神经网络(chained neural networks)[78]进行标签的逐层预测, 其中, 链式神经网络中的每一层与标签树中的每一层一一对应. 此种方法在原始标签集合具有显式的层级结构时具有非常好的效果.

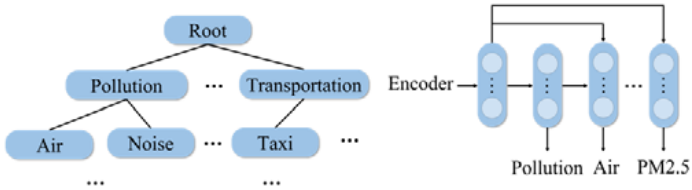


图 10 基于链式神经网络的标签推荐方法

4.3 基于标签语义的方法

与基于标签结构的方法相比, 基于标签语义的方法则主要通过引入标签语义向量, 利用到外部知识, 为低频标签赋予丰富的语义信息, 从而丰富了低频标签和高频标签之间的相关性. Li 等人[23]通过构建具有外部知识的标签图来探索标签相关性. 标签图的构建包含了 4 种关系: 包含关系(cp)、上下级关系(ss)、正向相关关系(po)以及共现关系(co). 其中, 包含关系指的是一个由多个单词组成的标签包含了其他标签; 上下级关系指的是两个标签之间存在词义上的上下义关系, 直接从 WordNet[79]中获得; 正向相关关系指的是两个标签的 WUP 相似度[80]超过了一定阈值; 共现关系指的是两个标签在历史数据中共同出现过, 通过统计历史数据即可得到.

如图 11 所示: 根据外部知识, 可以得到具有 4 种不同边的异构标签图 G; 之后, 初始化每个标签 h_j 的隐

表示 $\boldsymbol{e}(h_j)\in\mathbb{R}^{d_D}$, 其中, d_D 表示标签嵌入空间(embedding space)的维度; 随后, 通过 GCN[81]进行节点的表示学习, 从而得到每个标签的最终表示 $\boldsymbol{e}(h'_j)$. 最终表示即隐含了标签的相关性信息, 从而使得频繁出现的标签可以与其相关的长尾标签共享知识.

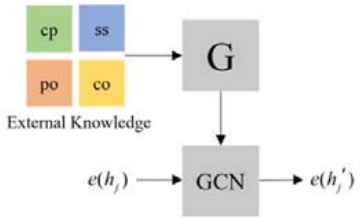


图 11 基于 GCN 的标签表示方法

标签语义信息的引入, 可以很好地利用到标签相关性. 除了使用基于标签图的 GCN 引入标签语义信息之外, Xiao 等人[82]直接使用 GloVe 预训练好的词向量作为标签表示, 为模型引入丰富的标签语义信息. 其通过标签语义的引入, 引导文本内容学习到更有利于分类的内容表示. Chen 等人[83]在双曲空间下基于标签的树状结构提取出标签的语义信息, 同样为后续任务引入了标签相关性信息.

5 基于用户偏好的标签推荐方法

如前文所述, 如果不考虑用户信息, 那么标签推荐问题可以看作是一个多标签分类任务. 然而, 用户信息在标签推荐中扮演着极为重要的角色[41−43]. 标签推荐任务中的标签由用户直接赋予, 因此与用户偏好息息相关[44]. 在标签推荐中, 主要存在两种形式的用户偏好差异: 一是不同用户关注的项目本来就存在着差异, 比如在基于微博文本的标签推荐场景下, 健身博主和美食博主发布的文本内容就存在着明显的差异性; 二是用户使用标签的习惯也存在着差异性, 这是由于每位用户的背景知识不一样, 因而即使对类似的项目进行打标签操作时, 不同用户所标记的标签也很可能不同. 因此, 忽略用户偏好虽然简化了模型, 但同时势必会造成部分关键信息的缺失, 从而影响模型的推荐能力. 虽然以往的大部分标签推荐方法还未关注到用户偏好, 但已有部分工作开始研究基于用户偏好的标签推荐方法, 并且取得了突出的成绩[16,23]. 其中, 基于用户表示的方法从用户的 ID、历史标签、历史项目信息提取出用户表示, 之后与项目表示合并, 得到用户和项目的联合表示. 而基于交互关系的方法则更进一步, 考虑到了用户、标签和项目三者之间的交互关系.

5.1 基于用户表示的方法

由于用户与项目和标签之间的交互关系较难捕获, 因此多数基于用户偏好的方法在获取到用户表示后, 直接将用户表示与项目表示进行结合, 以供后续任务使用. 此类方法我们称之为基于用户表示的方法. 不同方法的区别主要在于用户信息的来源不同, 其主要来源于 3 个方面.

(1) 用户 ID 信息. 用户 ID 信息是最简单的用户信息. Nguyen 等人[57]直接利用了用户的 ID 信息, 将其和项目特征进行拼接后输入到 MLP 中进行标签推荐. 通过引入用户 ID 信息, 模型可以学习到每名用户在打标签时的不同偏好;

(2) 用户的历史标签信息. 然而, 由于用户 ID 信息的稀疏性, 仅仅利用用户 ID 信息, 模型能够学习到的用户偏好信息有限. 因此, Maity 等人[36]则进一步利用到了标签信息学习用户表示, 首先将所有用户和标签构成一个异构图, 之后使用 node2vec[84]方法学习到每个用户节点的节点表示, 将其作为每位用户的用户表示. 与此类似, Quintanilla 等人[18]也使用了标签信息学习用户表示, 其将每名用户的历史标签信息输入到一个自编码器网络中, 将中间的隐含层作为用户表示. 这样的方法缓解了仅仅使用用户 ID 信息的稀疏性, 为模型引入了更加丰富的标签信息来表示用户偏好, 因而此类方法可以捕获到更有效的用户偏好;

(3) 用户的历史项目信息. 除了使用用户的 ID 信息和历史标签信息外, Zhang 等人[16]则是同时关注到了用户的 ID 信息、历史标签信息和历史项目信息, 通过一个注意力网络学习到用户偏好表示. Li 等人[23]也通过用户的历史标签信息和项目信息获取到用户偏好表示, 其分别用预先训练好的 Word2Vec 和 CNN 网络来提取用户的历史标签信息和项目信息, 最后经过一个 MLP 得到用户偏好表示. 由于引入了更加丰富的用户相关信息, 模型可以捕获到更加具体的用户偏好, 从而有助于后续标签推荐任务.

5.2 基于交互关系的方法

大部分基于用户表示的方法并没有直接关注到用户、项目和标签之间的交互关系. 然而, 由于标签推荐的研究对象是用户、项目和标签这 3 种主体及主体之间的相互作用, 因而考虑到三元组之间交互关系有利于个性化的标签推荐[85,86]. 基于交互关系的方法最早可以追溯至 Rendle 等人提出的基于张量分解的标签推荐方法[29,30]. 首先, 其将原始的用户、项目和标签的三元组交互信息看作一个三阶张量, 张量的每个维度分别代表用户、项目和标签. 张量的每个元素 $y_{u,i,t}\in\{0,1\}$, 含义是用户 u 对项目 i 是否打过标签 t. 之后, 如公式(34)所示, 其采用成对交互张量分解(pairwise interaction tensor factorization, PITF)的方法, 分别考虑到了用户 $\hat{\boldsymbol{u}}^T_{u,f}$ 和标签 $\hat{\boldsymbol{t}}^U_{t,f}$ 的交互关系、项目 $\hat{\boldsymbol{i}}^T_{i,f}$ 和标签 $\hat{\boldsymbol{t}}^I_{t,f}$ 的交互关系以及用户 $\hat{\boldsymbol{u}}^I_{u,f}$ 和项目 $\hat{\boldsymbol{i}}^U_{i,f}$ 的交互关系.

$$\hat{y}_{u,i,t}=\sum_f \hat{\boldsymbol{u}}^T_{u,f}\cdot\hat{\boldsymbol{t}}^U_{t,f}+\sum_f \hat{\boldsymbol{i}}^T_{i,f}\cdot\hat{\boldsymbol{t}}^I_{t,f}+\sum_f \hat{\boldsymbol{u}}^I_{u,f}\cdot\hat{\boldsymbol{i}}^U_{i,f} \tag{34}$$

然而, 由于该方法的用户数量、项目数量以及标签数量恒定, 因此其只能进行已有项目的标签推荐, 无法处理冷启动项目, 从而无法满足当前的应用场景. 除了使用基于张量分解的方法捕获交互关系之外, Wei 等人[17]采用了基于图神经网络协同过滤的方法[31]进行交互关系的学习. 如图 12 所示, 其首先构建了由用户、项目和标签组成的异构图 G, 之后采用 GCN[32]学习用户和标签的节点表示, 再根据项目表示和用户表示学习到用户交互的项目表示, 根据用户表示和标签表示学习到用户交互的标签表示, 最终得到预测得分. 通过对用户交互信息的关注, 模型可以更好地捕获到用户偏好, 从而更好地进行个性化的标签推荐.

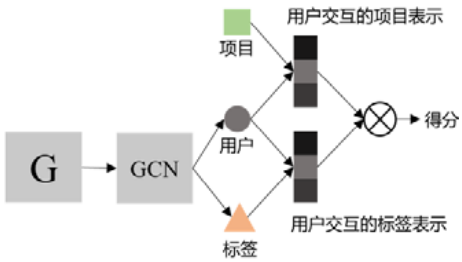


图 12 基于图神经网络协同过滤的标签推荐方法

6 难点与展望

通过对标签推荐现有方法的总结, 我们分析得到现阶段标签推荐方法一些亟待解决的难点以及可能的研究思路.

(1) 长尾问题

在标签推荐场景下, 标签词表的大小普遍较大, 而不同标签出现的次数又极度的不均衡. 因而标签的分布往往呈现出幂律分布的特性, 即大部分标签出现次数很少, 只有少部分标签具有较高的出现频次. 其中, 出现次数较少的标签称为长尾标签[87]. 这样的现象, 称为标签推荐的长尾问题. 由于大部分标签的训练样本过少, 因而模型很难学习到此类样本的分类边界, 从而影响模型的分类性能. 现阶段标签推荐方法主要从标签相关性的角度缓解了这一问题, 即通过挖掘标签的结构信息和层级信息使得长尾标签与高频标签之间产生某种联系, 进而辅助长尾标签的学习, 或通过外部知识赋予长尾标签更加丰富的语义信息. 未来的研究可以进

一步探索标签相关性, 如采用迁移学习的方法将高频标签的部分知识迁移到长尾标签上来. 另外, 可以从数据的不均衡性出发, 采用重采样、重平衡或代价敏感学习[88−91]的方法缓解数据不均衡问题所带来的影响, 使得模型可以学习到更好的分类边界.

(2) 用户偏好的动态性

如前文所述, 标签推荐本质上是面向用户进行的推荐任务. 不同的用户可能有不同的标签习惯, 类似的帖子会因此产生不同的标签. 因而, 对用户偏好的挖掘在标签推荐场景下尤为重要. 现阶段的方法主要从用户表示和交互关系的角度为模型引入用户偏好信息, 然而现有的基于用户偏好的方法都是静态的, 其假设了用户偏好在时间维度上的一致性, 即用户偏好不会随着时间发生变化. 然而在实际场景下, 用户的偏好却是动态变化的[92]. 忽略了用户偏好的动态性, 便无法实时得为标签推荐提供可靠的用户资料. 因此, 如何有效刻画标签推荐场景下用户偏好的动态变化特点, 从而提出基于动态用户偏好的标签推荐模型, 成为标签推荐的难点之一.

(3) 多模态信息的融合问题

如第 3.2 节所述, 随着多媒体社交应用(如 Twitter、TikTok、快手)的普及, 用户分享的信息也越来越多地从文字信息转化为富媒体信息(如图文信息、视频信息等). 富媒体信息往往包含了多个模态的信息, 并且其各个模态之间具有某种程度的相关关系. 因此, 在进行多模态信息融合的过程中, 应该考虑到每个模态之间的相关关系, 从而有利于更有效的特征提取. 在基于图文内容的标签推荐方法下, 研究者们采用了基于注意力机制的方法, 考虑到了图片信息和文本信息的交互, 从而提取到了更有效的内容信息. 然而, 在现阶段基于视频内容的标签推荐方法下[17−23], 还没有研究者考虑到视频多模态信息之间的交互关系. 这一点也可以作为后续研究者们进行研究的方向.

(4) 标签噪声

大部分标签推荐方法的前提假设是现有的标签是准确无误的. 然而, 由于原始标签是由用户赋予的, 从而无法避免人为因素的干扰, 因此往往会有不合适的标签、错误的标签、甚至无意义的标签[1](比如用户错误操作将一些无意义的字符作为标签)出现. 在以往的标签推荐工作中, 大部分工作并没有考虑到标签噪声的存在, 或仅仅是在预处理阶段将低频标签删去. 这样的操作一方面丢失了部分原始信息, 另一方面简化了原始的标签推荐问题, 从而使得真实的标签推荐问题没有得以解决. 因此, 如何在考虑标签噪声的前提下提升标签推荐的性能, 也是当前标签推荐问题的难点之一.

(5) 可解释标签推荐

已有的标签推荐方法主要关注于提升推荐的准确度. 然而, 推荐准确度可能并不是标签推荐的唯一目标或终极目标[10]. 本质上, 标签推荐是面向用户进行的推荐任务, 因此其任务的最终目的是为用户服务, 方便用户进行打标签操作, 同时提升用户的舒适度. 所以, 如果用户不知系统为何为其推荐某种标签的话, 其很可能不会选择系统推荐的标签. 这就要求研究者们为用户解释为何为其推荐某个标签, 从而提升用户对标签推荐结果的接受度, 进而提升用户体验.

(6) 缺少标准数据集

如前文所述, 标签推荐任务的应用场景多种多样. 然而, 无论是在文本场景、图片场景、图文场景还是视频场景下的标签推荐方法, 都没有相对应的标准数据集. 大部分研究者都基于自己的研究目的, 独自爬取和处理了对应的数据集. 由于不同的数据集具备各自的特性, 并且不同的处理方法也会为数据集带来难以预测的影响. 因此, 研究者们无法直观地进行多种算法间的比较. 所以, 为标签推荐领域制作开源的标准数据集, 对该领域具有一定的推动作用.

(7) 预训练模型在标签推荐上的应用

如第 4.3 节所述, 标签语义信息的引入, 无疑可以为标签推荐带来性能提升. 现有方法已经探索过 Word2Vec, GloVe 等预训练模型在标签推荐领域的应用, 然而, 标签推荐场景下的文本语料往往具有专业性和领域性. 如问答网站 Ask Ubuntu 上的内容均围绕着 Ubuntu 系统, 并且可能包含代码块等富文本信息. 因而在

该语料下的语义往往和通用语料之间具有一定的差距, 并且该语料下的词也有许多并不存在于预训练词表中, 从而导致无法直接利用到预训练模型. 随着 ELMo, BERT 等预训练模型[54,93−95]在自然语言处理方面的巨大成功, 如何将强大的预训练模型应用到标签推荐领域, 成为了亟待解决的问题.

(8) 标签推荐模型的快速求解

由于标签推荐任务的实时性, 高效、快速地求解标签推荐模型一直是研究人员关注的热点问题. 大数据时代, 互联网平台的海量数据量为标签推荐模型求解带来了极大的挑战[96]. 尤其是在图文、视频等多模态场景下, 标签推荐模型的复杂度急剧上升[23]. 因此, 如何设计面向多模态场景下的标签推荐算法, 使之能够高效而快速地为用户提供高质量标签推荐结果, 成为标签推荐的难点之一.

7 结 论

作为信息检索领域的重要研究方向, 标签推荐通过辅助用户进行打标签的操作, 极大地提升了标签的质量, 从而有助于更高效的信息检索. 因而, 标签推荐在近年来也获得了研究者们广泛的关注. 本文根据标签推荐方法要解决的问题, 将近年来常见的标签推荐方法划分为 3 个类别, 其分别是基于内容的方法、基于标签相关性的方法以及基于用户偏好的方法. 之后, 本文对这 3 个类别下的对应方法进行了梳理和剖析. 最后, 本文提出了当前标签推荐领域面临的主要挑战, 例如标签的长尾问题、用户偏好的动态性、多模态信息的融合问题以及标签推荐的领域性问题等, 并提出了可能的解决思路. 希望本文对标签推荐方法的综述能够为相关学者提供一定程度的帮助, 同时更好地促进标签推荐领域的研究.

References:

[1] Smith G. Tagging: People-powered Metadata for the Social Web. Berkeley: Peachpit New Riders, 2007.

[2] Song Y, Zhuang Z, Li H, *et al*. Real-time automatic tag recommendation. In: Proc. of the Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2008. 515−522.

[3] Berendt B, Christoph H. Tags are not metadata, but "just more content"—To some people. In: Proc. of the Int'l Conf. on Weblogs and Social Media. 2007.

[4] Lei K, Fu Q, Yang M, *et al*. Tag recommendation by text classification with attention-based capsule network. Neurocomputing, 2020, 391: 65−73.

[5] Krestel R, Fankhauser P, Nejdl W. Latent dirichlet allocation for tag recommendation. In: Proc. of the ACM Conf. on Recommender Systems. 2009. 61−68.

[6] Wang X, Zhang Y, Yamasaki T. Earn more social attention: user popularity based tag recommendation system. In: Companion Proc. of the Web Conf. 2020. 212−216.

[7] Tang S, Yao Y, Zhang S, *et al*. An integral tag recommendation model for textual content. In: Proc. of the AAAI Conf. on Artificial Intelligence, Vol.33. 2019. 5109−5116.

[8] Xia X, Lo D, Wang X, *et al*. Tag recommendation in software information sites. In: Proc. of the Working Conf. on Mining Software Repositories. 2013. 287−296.

[9] Sigurbjörnsson B, Zwol R. Flickr tag recommendation based on collective knowledge. In: Proc. of the Int'l Conf. on World Wide Web. 2008. 327−336.

[10] Belém FM, Almeida JM, Gonçalves MA. A survey on tag recommendation methods. Journal of the Association for Information Science and Technology, 2017, 68(4): 830−844.

[11] Gong Y, Zhang Q. Hashtag recommendation using attention-based convolutional neural network. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence. 2016. 2782−2788.

[12] Li Y, Liu T, Jiang J, *et al*. Hashtag recommendation with topical attention-based LSTM. In: Proc. of the Int'l Conf. on Computational Linguistics. 2016. 3019−3029.

[13] Sun B, Zhu Y, Xiao Y, *et al*. Automatic question tagging with deep neural networks. IEEE Trans. on Learning Technologies, 2018, 12(1): 29−43.

[14] Hassan HAM, Sansonetti G, Gasparetti F, *et al*. Semantic-based tag recommendation in scientific bookmarking systems. In: Proc. of the ACM Conf. on Recommender Systems. 2018. 465−469.

[15] Shi X, Huang H, Zhao S, *et al*. Tag recommendation by word-level tag sequence modeling. In: Proc. of the Int'l Conf. on Database Systems for Advanced Applications. 2019. 420−424.

[16] Zhang S, Yao Y, Xu F, *et al*. Hashtag recommendation for photo sharing services. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2019. 5805−5812.

[17] Wei Y, Cheng Z, Yu X, *et al*. Personalized hashtag recommendation for micro-videos. In: Proc. of the ACM Int'l Conf. on Multimedia. 2019. 1446−1454.

[18] Quintanilla E, Rawat Y, Sakryukin A, *et al*. Adversarial learning for personalized tag recommendation. IEEE Trans. on Multimedia, 2020, 23: 1083−1094.

[19] Wu Y, Yao Y, Xu F, *et al*. Tag2Word: Using tags to generate words for content based tag recommendation. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management. 2016. 2287−2292.

[20] Wu Y, Xi S, Yao Y, *et al*. Guiding supervised topic modeling for content based tag recommendation. Neurocomputing, 2018, 314: 479−489.

[21] Zhang Q, Wang J, Huang H, *et al*. Hashtag recommendation for multimodal microblog using co-attention network. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence. 2017. 3420−3426.

[22] Li J, Xu H, He X, *et al*. Tweet modeling with LSTM recurrent neural networks for hashtag recommendation. In: Proc. of the Int'l Joint Conf. on Neural Networks. 2016. 1570−1577.

[23] Li M, Gan T, Liu M, *et al*. Long-tail hashtag recommendation for micro-videos with graph convolutional network. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management. 2019. 509−518.

[24] Gong Y, Zhang Q, Huang X. Hashtag recommendation for multimodal microblog posts. Neurocomputing, 2018, 272: 170−177.

[25] Zuin G, Veloso A, Portinari JC, *et al*. Automatic tag recommendation for painting artworks using diachronic descriptions. In: Proc. of the Int'l Joint Conf. on Neural Networks. 2020. 1−8.

[26] Wang Y, Wang S, Tang J, *et al*. CLARE: A joint approach to label classification and tag recommendation. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2017. 210−216.

[27] Rawat YS, Kankanhalli MS. ConTagNet: Exploiting user context for image tag recommendation. In: Proc. of the ACM Int'l Conf. on Multimedia. 2016. 1102−1106.

[28] Huang H, Zhang Q, Gong Y, *et al*. Hashtag recommendation using end-to-end memory networks with hierarchical attention. In: Proc. of the Int'l Conf. on Computational Linguistics: Technical Papers. 2016. 943−952.

[29] Rendle S, Schmidt-Thieme L. Pairwise interaction tensor factorization for personalized tag recommendation. In: Proc. of the ACM Int'l Conf. on Web Search and Data Mining. 2010. 81−90.

[30] Rendle S, Marinho LB, Nanopoulos A, *et al*. Learning optimal ranking with tensor factorization for tag recommendation. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2009. 727−736.

[31] Wang X, He X, Wang M, *et al*. Neural graph collaborative filtering. In: Proc. of the Int'l ACM SIGIR Conference on Research and Development in Information Retrieval. 2019. 165−174.

[32] Berg R, Kipf TN, Welling M. Graph convolutional matrix completion. arXiv: 1706.02263, 2017.

[33] Wang S, Lo D, Vasilescu B, *et al*. EnTagRec++: An enhanced tag recommendation system for software information sites. Empirical Software Engineering, 2018, 23(2): 800−832.

[34] Wang S, Lo D, Vasilescu B, *et al*. EnTagRec: An enhanced tag recommendation system for software information sites. In: Proc. of the IEEE Int'l Conf. on Software Maintenance and Evolution. 2014. 291−300.

[35] Zhou P, Liu J, Yang Z, *et al*. Scalable tag recommendation for software information sites. In: Proc. of the IEEE Int'l Conf. on Software Analysis, Evolution and Reengineering. 2017. 272−282.

[36] Maity SK, Panigrahi A, Ghosh S, *et al*. DeepTagRec: A content-cum-user based tag recommendation framework for stack overflow. In: Proc. of the European Conf. on Information Retrieval. 2019. 125−131.

[37] Zhou PY. Research on label recommendation method in software information station [Ph.D. Thesis]. Wuhan: Wuhan University, 2019 (in Chinese with English abstract).

[38] Gao J, He Y, Wang Y, *et al*. STAR: Spatio-temporal taxonomy-aware tag recommendation for citizen complaints. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management. 2019. 1903−1912.

[39] Ness S, Theocharis A, Tzanetakis G, *et al*. Improving automatic music tag annotation using stacked generalization of probabilistic SVM outputs. In: Proc. of the ACM Int'l Conf. on Multimedia. 2009. 705−708.

[40] Zhao Z, Wang X, Xiang Q, *et al*. Large-scale music tag recommendation with explicit multiple attributes. In: Proc. of the ACM Int'l Conf. on Multimedia. 2010. 401−410.

[41] Wang H, Chen B, Li W. Collaborative topic regression with social regularization for tag recommendation. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence. 2013. 2719−2725.

[42] Guan Z, Bu J, Mei Q, *et al*. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In: Proc. of the Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2009. 540−547.

[43] Feng W, Wang J. Incorporating heterogeneous information for personalized tag recommendation in social tagging systems. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2012. 1276−1284.

[44] Geng LL, Cui CR, Shi C, *et al*. Social image tagging and group joint recommendation based on deep multi-tasking learning. Computer Science, 2020, 47(12): 177−182 (in Chinese with English abstract).

[45] Wang J, Hong L, Davison BD. Tag recommendation using keywords and association rules. In: Proc. of the ECML/PKDD Discovery Challenge Workshop. 2009.

[46] Toderici G, Aradhye H, Pasca M, *et al*. Finding meaning on YouTube: Tag recommendation and category discovery. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2010. 3447−3454.

[47] Wang H, Shi X, Yeung DY. Relational stacked denoising autoencoder for tag recommendation. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2015.

[48] Yamasaki T, Hu J, Sano S, *et al*. FolkPopularityRank: Tag recommendation for enhancing social popularity using text tags in content sharing services. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence. 2017. 3231−3237.

[49] Nguyen H, Wistuba M, Grabocka J, *et al*. Personalized deep learning for tag recommendation. In: Proc. of the Pacific-Asia Conf. on Knowledge Discovery and Data Mining. 2017. 186−197.

[50] Wang X, Zhang Y, Yamasaki T. User-aware folk popularity rank: User-popularity-based tag recommendation that can enhance social popularity. In: Proc. of the ACM Int'l Conf. on Multimedia. 2019. 1970−1978.

[51] Lima E, Shi W, Liu X, *et al*. Integrating multi-level tag recommendation with external knowledge bases for automatic question answering. ACM Trans. on Internet Technology, 2019, 19(3): 1−22.

[52] Tonge A, Caragea C. Privacy-aware tag recommendation for accurate image privacy prediction. ACM Trans. on Intelligent Systems and Technology, 2019, 10(4): 1−28.

[53] Chen X, Yu Y, Jiang F, *et al*. Graph neural networks boosted personalized tag recommendation algorithm. In: Proc. of the Int'l Joint Conf. on Neural Networks. 2020. 1−8.

[54] Khezrian N, Habibi J, Annamoradnejad I. Tag recommendation for online Q&A communities based on BERT pre-training technique. arXiv: 2010.04971, 2020.

[55] Ramage D, Hall D, Nallapati R, *et al*. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. 2009. 248−256.

[56] Zha H, He X, Ding C, *et al*. Bipartite graph partitioning and data clustering. In: Proc. of the Int'l Conf. on Information and Knowledge Management. 2001. 25−32.

[57] Mikolov T, Sutskever I, Chen K, *et al*. Distributed representations of words and phrases and their compositionality. arXiv: 1310.4546. 2013.

[58] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. 2014. 1532−1543.

[59] Sabour S, Frosst N, Hinton G. Dynamic routing between capsules. arXiv: 1710.09829, 2017.

[60] Yang M, Zhao W, Ye J, *et al*. Investigating capsule networks with dynamic routing for text classification. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. 2018. 3110−3119.

[61] Shi L, Wang Y, Cheng Y, *et al*. Review of attention mechanism in natural language processing. Data Analysis and Knowledge Discovery, 2020, 4(5): 1−14 (in Chinese with English abstract).

[62] Blei D, Ng A, Jordan M. Latent dirichlet allocation. Journal of Machine Learning Research, 2003, 3: 993−1022.

[63] Lai S, Xu L, Liu K, *et al*. Recurrent convolutional neural networks for text classification. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2015. 2267−2273.

[64] Zhou C, Sun C, Liu Z, *et al*. A C-LSTM neural network for text classification. arXiv: 1511.08630, 2015.

[65] Yang Z, Yang D, Dyer C, *et al*. Hierarchical attention networks for document classification. In: Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016. 1480−1489.

[66] He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 770−778.

[67] Che BQ, Zhou D. A tag recommendation method for fusing network structure information and text content. Computer Application, 2021, 41(4): 976−983 (in Chinese with English abstract).

[68] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, Vol.25. 2012. 1097−1105.

[69] Bao HZ, Zhou D, Wu T. Tag recommendation method fusing multi-source heterogeneous network information. Journal of Shandong University (Natural Science Edition), 2019, 54(3): 56−66 (in Chinese with English abstract).

[70] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.

[71] Lu J, Yang J, Batra D, *et al*. Hierarchical question-image co-attention for visual question answering. In: Advances in Neural Information Processing Systems, Vol.29. 2016. 289−297.

[72] Kowald D, Pujari SC, Lex E. Temporal effects on hashtag reuse in Twitter: A cognitive-inspired hashtag recommendation approach. In: Proc. of the Int'l Conf. on World Wide Web. 2017. 1401−1410.

[73] Zhang SW. Research on multi-modal content tag recommendation technology in social networks [MS. Thesis]. Nanjing: Nanjing University, 2020 (in Chinese with English abstract).

[74] Hershey S, Chaudhuri S, Ellis DPW, *et al*. CNN architectures for large-scale audio classification. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. 2017. 131−135.

[75] Arora S, Liang Y, Ma T. A simple but tough-to-beat baseline for sentence embeddings. In: Proc. of the Int'l Conf. on Learning Representations. 2017.

[76] Hu D. An introductory survey on attention mechanisms in NLP problems. In: Proc. of the SAI Intelligent Systems Conf. 2019. 432−448.

[77] Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems. 2014. 3104−3112.

[78] Wehrmann J, Barros RC, Dôres SN, *et al*. Hierarchical multi-label classification with chained neural networks. In: Proc. of the Symp. on Applied Computing. 2017. 790−795.

[79] Miller GA. WordNet: An Electronic Lexical Database. Cambridge: MIT, 1998.

[80] Wu Z, Palmer M. Verb semantics and lexical selection. In: Proc. of the Annual Meeting of the Association for Computational Linguistics. 1994. 133−138.

[81] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv: 1609.02907, 2016.

[82] Xiao L, Huang X, Chen B, *et al*. Label-specific document representation for multi-label text classification. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing and the Int'l Joint Conf. on Natural Language Processing. 2019. 466−475.

[83] Chen B, Huang X, Xiao L, *et al*. Hyperbolic capsule networks for multi-label classification. In: Proc. of the Annual Meeting of the Association for Computational Linguistics. 2020. 3115−3124.

[84] Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2016. 855−864.

[85] Yang Y, Di YD, Liu JH, *et al*. Research on sort learning based on tensor decomposition in personalized tag recommendation. Computer Science, 2020, 47(S2): 515−519 (in Chinese with English abstract).

[86] Lu YN, Du DF. Tag recommendation based on pair interaction tensor decomposition. Journal of University of Science and Technology of China, 2019, 49(1): 31−39 (in Chinese with English abstract).

[87] Cui Y, Jia M, Lin T, *et al*. Class-balanced loss based on effective number of samples. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2019. 9268−9277.

[88] Zhang D, Li T, Zhang H, *et al*. On data augmentation for extreme multi-label classification. arXiv: 2009.10778, 2020.

[89] Lin T, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 2980−2988.

[90] Zoph B, Vasudevan V, Shlens J, *et al*. Learning transferable architectures for scalable image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 8697−8710.

[91] Zhou B, Cui Q, Wei XS, *et al*. BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2020. 9719−9728.

[92] Liu HF, Jing LP, Yu J. Survey of matrix factorization based recommendation methods by integrating social information. Ruan Jian Xue Bao/Journal of Software, 2018, 29(2): 340−362 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5391.htm [doi: 10.13328/j.cnki.jos.005391]

[93] Devlin J, Chang MW, Lee K, *et al*. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv: 1810.04805, 2018.

[94] Lan Z, Chen M, Goodman S, *et al*. Albert: A lite BERT for self-supervised learning of language representations. arXiv: 1909.11942, 2019.

[95] Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. arXiv: 1906.05474, 2019.

[96] Tuarob S, Pouchard LC, Giles CL. Automatic tag recommendation for metadata annotation using probabilistic topic modeling. In: Proc. of the ACM/IEEE-CS Joint Conf. on Digital Libraries. 2013. 239−248.

附中文参考文献:

[37] 周平义. 软件信息站中标签推荐方法研究[博士学位论文]. 武汉: 武汉大学, 2019.

[44] 耿蕾蕾, 崔超然, 石成, 等. 基于深度多任务学习的社交图像标签和分组联合推荐. 计算机科学, 2020, 47(12): 177−182.

[61] 石磊, 王毅, 成颖, 等. 自然语言处理中的注意力机制研究综述. 数据分析与知识发现, 2020, 4(5): 1−14.

[67] 车冰倩, 周栋. 融合网络结构信息及文本内容的标签推荐方法. 计算机应用, 2021, 41(4): 976−983.

[69] 包恒泽, 周栋, 吴谈. 融合多源异构网络信息的标签推荐方法. 山东大学学报(理学版), 2019, 54(3): 56−66.

[73] 张素威. 社交网络多模态内容标签推荐技术研究[硕士学位论文]. 南京: 南京大学, 2020.

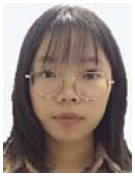
[85] 杨洋, 邸一得, 刘俊晖, 等. 基于张量分解的排序学习在个性化标签推荐中的研究. 计算机科学, 2020, 47(S2): 515−519.

[86] 鲁亚男, 杜东舫. 基于成对交互张量分解的标签推荐. 中国科学技术大学学报, 2019, 49(1): 31−39.

[92] 刘华锋, 景丽萍, 于剑. 融合社交信息的矩阵分解推荐方法研究综述. 软件学报, 2018, 29(2): 340−362. http://www.jos.org.cn/1000-9825/5391.htm [doi: 10.13328/j.cnki.jos.005391]

徐鹏宇(1997－), 男, 博士生, CCF 学生会员, 主要研究领域为多标签学习, 标签推荐.

刘华锋(1994－), 男, 博士, CCF 学生会员, 主要研究领域为信息检索, 深度生成模型.



刘冰(2000－), 女, 学士, 主要研究领域为信息检索, 多标签分类.

景丽萍(1978－), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为机器学习, 高维数据表示及其在人工智能领域中的应用.

于剑(1969－), 男, 博士, 教授, 博士生导师, CCF 会士, 主要研究领域为人工智能, 机器学习.