



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目：融合网络结构信息及文本内容的标签推荐方法
作者：车冰倩，周栋
收稿日期：2020-08-20
网络首发日期：2020-10-22
引用格式：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[J/OL]. 计算机应用. <https://kns.cnki.net/kcms/detail/51.1307.TP.20201021.0852.006.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

融合网络结构信息及文本内容的标签推荐方法

车冰倩,周 栋*

(湖南科技大学 计算机科学与工程学院,湖南 湘潭 411201)

(* 通信作者电子邮箱 dongzhou1979@hotmail.com)

摘 要:为文本推荐合适的标签是更好地组织和使用文本内容的一项有效手段,目前大部分标签推荐方法主要通过挖掘文本内容进行推荐。然而,大部分数据信息并非独立存在,如语料库中的文本间词共现关系可形成复杂的网络结构。以往研究表明,文本间的网络结构信息和文本内容信息可以分别从两个不同的角度对同一文本的语义进行概括,并且从两方面提取的信息可以互为补充和解释。基于此,提出了一种同时对文本网络结构信息和文本内容信息进行建模的标签推荐方法。该方法首先使用图卷积神经网络(GCN)提取文本间网络结构信息,然后使用循环神经网络(RNN)提取文本内容信息,最后使用注意力机制结合文本间网络结构信息和文本内容信息进行标签的推荐。与基于图卷积神经网络的标签推荐方法(GCN)、基于主题注意力的长短时记忆神经网络的标签推荐方法(TLSTM)等基线方法相比,提出的使用注意力机制结合网络结构信息与文本内容的标签推荐方法具有更好的性能。如在 Mathematics Stack Exchange 数据集上准确率、召回率和 F1 值相较最优基线方法分别提高了 2.3%、3.8%、7.0%。

关键词:标签推荐;循环神经网络;图卷积神经网络;注意力机制;网络结构信息;文本内容

中图分类号:TP391 **文献标志码:**A

Tag recommendation method combining network structure information and text content

CHE Bingqian, ZHOU Dong*

(School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan Hunan 411201, China)

Abstract: Recommend appropriate tags for text is an effective means to better organize and use text content. At present, most tag recommendation methods mainly recommend text by mining content. However, most of the data information does not exist independently, for example, the co-occurrence of words between texts in a corpus can form a complex network structure. Previous studies have shown that the network structure information and text content information between texts can summarize the semantics of the same text from two different angles, and extracting information from two aspects can complement and explain each other. Based on this, a tag recommendation method was proposed to simultaneously model the structure information of text network and the content information of text. Firstly, Graph Convolutional neural Network (GCN) was used to extract network structure information between texts, then the Recurrent Neural Network (RNN) was used to extract the text content information, and finally the attention mechanism was used to recommend tags by combining the network structure information between texts and the text content information. Compared with baseline methods, such as tag recommendation method based on GCN and tag recommendation with Topical attention-based Long Short-Term Memory (TLSTM), the proposed tag recommendation method using attention mechanism combining network structure information and text content has better performance. For example, on the Mathematics Stack Exchange dataset, the Precision, Recall and F1 are improved by 2.3%, 3.8%, and 7.0% respectively compared with the optimal baseline method.

Key words: tag recommendation; Recurrent Neural Network (RNN); Graph Convolutional neural Network (GCN); attention mechanism; network structure information; text content

0 引言

随着数字资源的快速增长,标签推荐越来越受到人们的关注。最新的研究表明,除标题、描述和用户评论等其他文本功能外,标签能够对资源进行分类和建立索引,是组织和检索

内容的有效方式^[1],被广泛应用于各种自然语言处理任务中,例如情感分类^[2]、内容推荐^[3]等为在线内容推荐合适的标签是自然语言处理中的一项重要任务。从内容创建者的角度来看,标签推荐有助于为创建者提高所选标签质量和用户体验;从内容消费者角度来看,标签推荐有助于为消费者的搜索和

收稿日期:2020-08-20;修回日期:2020-09-28;录用日期:2020-10-14。
基金项目:国家自然科学基金资助项目(61876062)。
作者简介:车冰倩(1998—),女,内蒙古呼伦贝尔人,硕士研究生,CCF 会员,主要研究方向:信息检索、自然语言处理;周栋(1979—),男,湖南长沙人,教授,博士,主要研究方向:信息检索、自然语言处理。

检索请求提供更好的服务^[4]。然而,手动标记通常既费时又费力。因此,为了提高数据管理效率和简化标注过程,有必要用到自动标签推荐系统。

目前已有研究者提出了许多自动标签推荐方法。一些研究利用传统的方法(TF-IDF, N-grams等)人工提取文本内容特征进行标签推荐^[5-6]。此外,潜在狄利克雷分布(Latent Dirichlet Allocation, LDA)模型^[7-8]或其扩展的主题模型^[9-10]用于从文本内容中提取隐含的主题信息,基于这些主题信息计算文本之间的相似性,再通过文本间的语义相似度进行标签推荐。近年来,使用深度学习模型对文本内容中的隐藏信息建模(如词序信息、上下文信息等),进行标签推荐也成为了当下研究热点之一^[11]。常用的模型包括卷积神经网络(Convolutional Neural Networks, CNN)^[12]和循环神经网络(Recurrent Neural Network, RNN)^[13]等。但是这些方法主要是通过挖掘文本的内容信息,从资源本体的角度来设计标签推荐方法。实际上,很多文本并非独立存在。例如,文本数据可通过文本间的词共现关系而联系到一起,文本中的词与文本之间具有隐含的网络结构信息。研究表明,文本内容层面的信息和网络结构层面的信息独立存在,但又相互关联和影响,从不同的角度共同对一个文本的语义特征进行概括会得到更好的结果^[14]。尤其在某方面信息缺乏时,内容和结构两方面信息可以互为补充,从而改进推荐的效果^[15]。因此,需要充分利用文本之间的网络结构信息和文本内容信息进一步提高推荐的准确性。

基于上述分析,本文提出一种融合网络结构信息和文本内容、基于 G-RNN (Graph Attention-Based Recurrent Neural Network)模型的标签推荐方法。该方法首先将文档与词汇作为节点构建异构网络图;然后,基于图卷积神经网络(Graph Convolutional neural Network, GCN)^[16]建立基于这些节点的文本异构图,捕获高阶邻域信息,获取文本间的网络结构信息。接下来基于 RNN 模型获取文本的内容信息;最后通过注意力机制对文本网络结构信息和文本内容信息进行交互建模。在三个真实数据集上的实验结果表明,本文的方法明显优于其他基线方法。本文的主要工作如下:

1) 本文提出了一种基于注意力机制的混合标签推荐模型。该模型融合了文本间网络结构信息和文本内容信息,以提取更好的文本语义特征。该模型可以解决现有标签推荐模型中文本语义特征提取不充分的问题。

2) 本文同时考虑到循环神经网络 RNN 的四种变体对提取文本内容信息的影响,并用实验进行了对比证明。

3) 实验结果表明,与现有的几种基线方法相比,本文提出的融合网络结构信息和文本内容信息的标签推荐方法具有更好的性能。

1 相关工作

标签推荐技术已经引起了学者的广泛关注和研究,并根据具体问题,提出了多种推荐方法。总的来说可以将这些推荐方法分为基于传统语义的方法和基于深度语义的方法。其

中基于传统语义的方法主要分为两类^[17]:基于协同过滤的方法和基于内容的方法。

基于协同过滤的方法核心思想是利用历史评分信息进行标签推荐。Feng等^[18]将社会标签系统建模为多类型图,通过学习图中节点和边的权重来推荐标签。Fang等^[19]提出了一种新的个性化标签推荐方法,该方法可视为典型张量分解的非线性扩展。Zhao等^[20]将标签数据中的关系建模为异构图,并提出了一个标签推荐的排序算法框架。

与基于协同过滤的方法相比,基于内容的方法是将内容作为输入,因此可以用于为新内容推荐标签,从而避免基于协同过滤方法的冷启动问题。典型的基于内容的标签推荐技术是使用生成模型。Krestel等^[7]使用 LDA,利用主题分布找出与目标资源具有相同主题的资源作为扩展标签集。Si等^[21]提出了一种基于扩展 LDA 的标签推荐方法,该模型通过在生成过程中添加标签变量来扩展 LDA 模型。Ding等^[22]提出了一种针对微博推荐标签的主题翻译模型,该模型将标签推荐建模为从内容到标签的翻译过程。Godin等^[23]设计和实现了基于朴素贝叶斯技术的二进制分类器,该分类器区分英语和非英语推文来推荐标签。

基于深度语义的方法通过基于深度神经网络的模型学习文本表示用以捕获文本内容,然后将该文本表示分类给不同的标签用来实现标签推荐。Weston等^[12]提出了一种基于 CNN 的推荐模型,该模型将单词以及整个文本内容表示嵌入在 CNN 体系结构的中间层中以学习语义嵌入。Gong等^[24]采用了具有注意力机制的 CNN 模型来执行标签推荐任务。Li等^[25]利用基于注意力机制的长短时记忆网络(Long Short Term Memory, LSTM)进行标签推荐。Huang等^[26]在文献^[24]和文献^[25]的基础上,提出了一种基于注意力机制的记忆网络的标签推荐模型,用记忆网络代替 CNN 和 LSTM 进行标签推荐。Li等^[27]提出了一种将 CNN 和 LSTM 相结合的标签推荐模型。Tang等^[28]采用了具有注意力机制的门控循环单元(Gated Recurrent Unit, GRU)模型,同时建模文本内容和标签进行标签推荐。Shi等^[29]提出了一种基于注意力机制的双向长短时记忆网络(Bi-directional Long Short Term Memory, Bi-LSTM)标签推荐模型。从这些研究结果来看,更准确的获取文本隐含语义信息有助于促进标签推荐的准确性。然而,许多真实数据集是以图或者网络形式出现的。图是一种特殊的数据格式,可以用来表示现实世界中的各种网络信息。在图中,节点由边连接,通过深度学习模型可以将图中的节点映射到向量空间中,充分挖掘向量空间中的节点进行分类。然而,到目前为止,很少有研究关注将图卷积神经网络应用于标签推荐,以充分挖掘文本中的词与文本本身之间的隐含网络结构信息,进一步提升标签推荐的性能。

最近,Yao等^[30]利用图卷积神经网络进行文本分类。受此启发,本文提出了一种融合文本间网络结构信息和文本内容的方法,该方法采用的标签推荐模型为 G-RNN 模型。与以往的工作相比,G-RNN 模型通过在 RNN 提取文本内容信息的基础上通过注意力机制结合了使用 GCN 提取的文本间的网络结构信息,使用更丰富的文本信息挖掘标签,能够有效提升推荐标签的准确度。

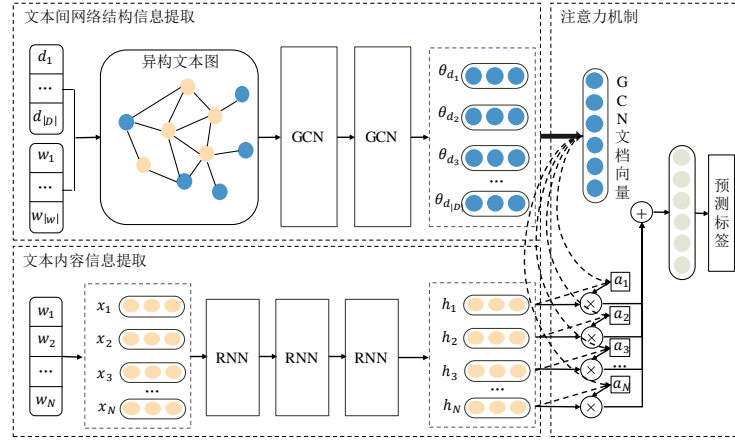


图1 G-RNN模型

Fig. 1 G-RNN model

2 自动标签推荐方法

首先定义问题,然后详细介绍本文提出的自动标签推荐方法。图1为该方法采用的标签推荐模型G-RNN,表1为本文的常用符号及其说明。

表1 常用符号及其含义说明

Tab. 1 Common symbols and their meanings

符号	含义说明
D	表示语料库中全部文本
d_i	表示文本内容
N	表示最大文本长度
L_w	词向量嵌入矩阵
G	异构文本图
V	图的节点集合
E	图的边集合
A	邻接矩阵
M	度矩阵
$\#S$	滑动窗口数量
g_i	输入门
f_i	遗忘门
o_i	输出门
r_i	重置门
z_i	更新门

2.1 问题定义

本文将标签推荐任务作为多分类问题进行处理。为更好地提取特征进行文本表示,本文考虑了注意力机制,并提出了一种融合网络结构信息和文本内容信息的标签推荐方法。

通常,语料库中全部文本可以表示为 $D = \{d_1, d_2, \dots, d_i, \dots, d_{|D|}\}$,其中 $|D|$ 指语料库中包含的文本数量。语料库中全部词汇可以表示为 $W = \{w_1, w_2, \dots, w_i, \dots, w_{|W|}\}$,其中 $|W|$ 指语料库中包含的词汇数量。单个文本内容可以表示为 $d_i = \{w_1, w_2, \dots, w_i, \dots, w_{|d_i|}\}$,其中 w_i 指文本的第 i 个词汇, $|d_i|$ 指文本包含的词汇数目。对于一个新文本,标签推荐的任务是为这个新文本推荐功能语义相近的标签。本文提出的标签推荐模型由三部分组成:1)文本间网络结构信息提取。本文首先将文档与词汇作为节点构建文本异构图,使用GCN捕获高阶邻域信息,从而获取文本间网络结构信息。2)文本内容信息提取。本文使用四种不同的RNN变体对文本内容中

的隐藏信息(如词序信息、上下文信息等)进行编码提取特征。3)注意力机制。本文使用注意力机制将文本网络结构信息与文本内容信息相融合,通过softmax层获取最终的推荐标签。

在本节的其余部分,将详细介绍这三个部分。所有这三个部分的基础是每个单词都被表示为低维,连续的实值向量,也称为单词嵌入^[31-32]。所有词向量都堆叠在词嵌入矩阵 $L_w = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{|W|}\} \in \mathbb{R}^{\dim \times |W|}$,其中 \mathbb{R} 表示实数集向量空间, \dim 是词向量的维数, \mathbf{x}_i 是 w_i 对应的词向量。使用嵌入学习算法对来自文本语料库的单词向量的值进行预训练,以更好地利用单词的语义和语法关联信息^[32]。给定文本内容,对文本中每个词进行嵌入,得到输入层向量。为保证文本向量输入长度一致,设定单个文本最长文本长度为 N ,文本 d_i 可用 $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ 表示。

2.2 文本间网络结构信息提取

语料库中文本间网络结构信息的提取主要是基于文档及文档间词共现信息构建异构文本图,通过GCN实现建模。本节将分为两步介绍该过程。第一步,构建异构文本图;第二步,使用GCN对异构文本图进行建模计算获取获得文档嵌入向量。

2.2.1 异构文本图构建

该部分将介绍如何从语料库中构建异构文本图。具体任务为,给定语料库 D ,构建异构图 $G = (V, E)$,其中 V 是图的节点集合, E 是图的边集合。节点指文档以及语料库中去重后的全部词汇,边是基于文本间词的共现和整个语料库中词的共现来构建的,其中文档与共现词进行连接,共现词与共现词之间进行连接。文档节点和单词节点之间的边的权重是文档中该词的TF-IDF值,其中词频(TF)为单词在文档中出现的次数,逆文档频率(IDF)是包含该词的文档数量的倒数的对数。为了更好地收集全局词共现信息,本文对语料库中的所有文档使用固定大小的滑动窗口收集信息。单词节点与单词节点之间的边的权重使用点互信息(Point-wise Mutual Information, PMI)来计算。任意节点 i 与节点 j 之间的边的权值定义为:

$$A_{ij} = \begin{cases} \text{PMI}(i, j) & i, j \text{ 为词, } \text{PMI}(i, j) > 0 \\ \text{TF-IDF}_{i, j} & i \text{ 为文档, } j \text{ 为词} \\ 1 & i, j \text{ 为词, } i = j \\ 0 & \text{其他} \end{cases} \quad (1)$$

接下来,单词对 i, j 的PMI值的计算如下:

$$\text{PMI}(i, j) = \ln \frac{p(i, j)}{p(i)p(j)} \quad (2)$$

$$p(i, j) = \frac{\#S(i, j)}{\#S} \quad (3)$$

$$p(i) = \frac{\#S(i)}{\#S} \quad (4)$$

其中, $\#S(i)$ 是语料库中包含单词 i 的滑动窗口的数量, $\#S(i, j)$ 是包含单词 i 和 j 的滑动窗口的数量, $\#S$ 是语料库中滑动窗口的总数。正 PMI 值意味着语料库中单词的高语义相关性, 而负 PMI 值表示语料库中很少或没有语义相关性。因此, 本文仅在 PMI 为正值时词对之间添加边。

2.2.2 图编码器

Kipf 等人^[16]在 2017 年提出了一种图卷积神经网络 GCN。GCN 是一个多层神经网络, 它是对图形结构数据的传统卷积算法的变体。它直接对图进行操作, 并根据节点的邻域属性归纳出节点的嵌入向量。

在构建了异构文本图后, 可获取到异构图的邻接矩阵 A 及其度矩阵 M , 其中 $M_{ii} = \sum_j A_{ij}$, 由于节点自循环, A 的对角线元素为 1。同时本文将特征矩阵设置为单位矩阵 I 。将邻接矩阵 A 与特征矩阵送入两层 GCN 中进行建模和卷积运算, 形成词和文档的嵌入表示向量, 第二层的节点嵌入与标签集的大小相同。

$$Z = \rho(\tilde{A} \text{ReLU}(\tilde{A}W_0)W_1) \quad (5)$$

其中 $\tilde{A} = M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$ 为归一化对称邻接矩阵, W_0 和 W_1 为权重矩阵, ρ 为激活函数。使用梯度下降法可以训练 W_0 和 W_1 的权重参数。在公式 5 中, $F_1 = \tilde{A}W_0$ 为第一层 GCN 的文档嵌入表示矩阵, $F_2 = \tilde{A} \text{ReLU}(\tilde{A}W_0)W_1$ 为第二层 GCN 的文档嵌入表示矩阵, 每一行为每个文档对应的文档向量。

2.3 文本内容信息提取

一般而言, 局部序列特征的提取是对文本进行编码, 将文本中嵌入的一系列单词转换成固定的二维向量的过程。有许多研究使用 RNN 的变体^[13, 27-29]来进行文本编码。GCN 可以有效地提取全局特征, 而 RNN 的变体 (如 LSTM, Bi-LSTM, GRU, Bi-GRU) 适合对文本进行编码提取局部序列特征。因此, 本文将使用 RNN 的四个变体对文本进行编码提取局部序列特征。

本文首先通过预训练得到给定文本 (x_1, x_2, \dots, x_N) , 使用三层 RNN 的变体按顺序进行序列建模。RNN 的变体按顺序对其进行处理, 获取到所有的隐藏向量 $[h_1, h_2, \dots, h_N]$ 。

1) 基于 LSTM 的序列编码器。

LSTM^[33]在每个单元使用输入门, 忘记门和输出门来控制信息沿着序列的传递。文本中每个单词按序列分别输入到一个 LSTM 单元。对于 t 时刻 LSTM 单元的输入 x_t , 给定 $t-1$ 时刻的输出 h_{t-1} 和单元状态 c_{t-1} , LSTM 单元使用输入门 g_t , 遗忘门 f_t , 输出门 o_t 和描述当前输入的单元状态 \tilde{c}_t 获取到 t 时刻的输出 h_t 和单元状态 c_t 。在对文本进行编码时, t 时刻的 LSTM 单元的更新公式如下所示:

$$g_t = \sigma(W_g x_t + W_g h_{t-1} + b_g) \quad (6)$$

$$f_t = \sigma(W_f x_t + W_f h_{t-1} + b_f) \quad (7)$$

$$\tilde{c}_t = \tanh(W_c x_t + W_c h_{t-1} + b_c) \quad (8)$$

$$c_t = f_t c_{t-1} + g_t \tilde{c}_t \quad (9)$$

$$o_t = \sigma(W_o x_t + W_o h_{t-1} + b_o) \quad (10)$$

$$h_t = o_t \tanh(c_t) \quad (11)$$

其中, σ 为 sigmoid 函数, W_g, W_f, W_c 和 W_o 是加权矩阵, 而 b_g, b_f, b_c 和 b_o 是偏移量。

2) 基于 Bi-LSTM 的序列编码器。

Bi-LSTM 是基于众所周知的 LSTM 模型^[33]开发的双向神经网络。一个 Bi-LSTM 单元包含两个 LSTM 单元, 用于捕获文本中的前向和后向信息。在本文中, $t-1$ 时刻的前向 LSTM 单元的输出记为 \vec{h}_{t-1} , $t-1$ 时刻的后向 LSTM 单元的输出记为 \overleftarrow{h}_{t-1} 。因此, t 时刻的 Bi-LSTM 单元的更新公式如下所示:

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \quad (12)$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t-1}) \quad (13)$$

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (14)$$

3) 基于 GRU 的序列编码器。

GRU^[34]在每个单元使用重置门, 更新门来控制信息沿着序列的传递。同 LSTM 的序列编码器一样, 文本中每个单词分别输入到一个 GRU 单元。对于 t 时刻 GRU 单元的输入 x_t , 给定 $t-1$ 时刻的输出 h_{t-1} , GRU 使用重置门 r_t , 更新门 z_t 和当前时刻待输入的隐藏状态信息 \tilde{h}_t 获取到 t 时刻的输出 h_t 。在对文本进行编码时, t 时刻的 GRU 的更新公式如下所示:

$$r_t = \sigma(W_r x_t + W_r h_{t-1} + b_r) \quad (15)$$

$$z_t = \sigma(W_z x_t + W_z h_{t-1} + b_z) \quad (16)$$

$$\tilde{h}_t = \tanh(W_h x_t + W_h h_{t-1} + b_h) \quad (17)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (18)$$

其中, σ 为 sigmoid 函数, W_z, W_r, W_h 是加权矩阵, 而 b_z, b_r, b_h 是偏移量。

4) 于 Bi-GRU 的序列编码器。

Bi-GRU 是基于众所周知的 GRU 模型^[34]开发的双向神经网络。一个 Bi-GRU 单元包含两个 GRU 单元, 用于捕获文本中的前向和后向信息。在本文中, $t-1$ 时刻的前向 GRU 单元的输出记为 \vec{h}_{t-1} , $t-1$ 时刻的后向 GRU 单元的输出记为 \overleftarrow{h}_{t-1} 。因此, t 时刻的 Bi-GRU 单元的更新公式如下所示:

$$\vec{h}_t = \text{GRU}(x_t, \vec{h}_{t-1}) \quad (19)$$

$$\overleftarrow{h}_t = \text{GRU}(x_t, \overleftarrow{h}_{t-1}) \quad (20)$$

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (21)$$

综上所述, RNN 层可得到文本序列隐藏向量序列 $[h_1, h_2, \dots, h_N]$ 。

2.4 注意力机制

本文引入了注意力机制对网络结构信息和文本内容信息进行整合。给定文本 d_i , 使用 GCN 获取到矩阵 F_2 中对应文本 d_i 的文档嵌入向量 $\theta_{d_i} \in \mathbb{R}^{k \times 1}$, k 为语料库中标签数量, 使用 RNN 按顺序对文本 d_i 进行处理, 获取到所有的隐藏向量 $[h_1, h_2, \dots, h_N]$ 。注意力层将为每个文本连续输出向量 $\text{vec} \in \mathbb{R}^{\mu \times 1}$ 。输出向量为每个隐藏状态 h_j 的加权和:

$$\gamma_j = l_a^T \tanh(W_a \cdot x_{d_i} + U_a \cdot h_j) \quad (22)$$

$$a_j = \text{softmax}(\gamma_j) = \frac{\exp(\gamma_j)}{\sum_{j'=1}^N \exp(\gamma_{j'})} \quad (23)$$

$$\text{vec} = \sum_{j=1}^N a_j h_j \quad (24)$$

其中, μ 为 RNN 隐藏层维数, $W_a \in \mathbb{R}^{\mu \times K}$, $U_a \in \mathbb{R}^{\mu \times \mu}$ 和 $l_a \in \mathbb{R}^{\mu \times 1}$ 是权重矩阵。 $a_j \in [0, 1]$ 是 h_j 的注意力权重且 $\sum_j a = 1$ 。

接下来将输出向量 $\text{vec} = [v_1, v_2, \dots, v_K]$ 馈送到线性层, 该线性层的输出长度为所有标签的数量。通过 softmax 层输出选出概率高标签作为推荐标签:

$$\text{softmax}(v_i) = \frac{\exp(v_i)}{\sum_{i'=1}^K \exp(v_{i'})} \quad (25)$$

本文将标签推荐建模为多分类任务, 通过最小化标签分类的交叉熵误差, 以监督的方式训练模型。损失函数如下:

$$\text{loss} = -\frac{1}{|Y|} \sum_{(y, T) \in Y \times T} \log p(T_i | y) \quad (26)$$

其中 Y 是训练集, y 和 T 是训练集文档及其对应标签集, T_i 是正在预测的当前标签, $p(T_i | y)$ 是为输入文档 y 选择标签 T_i 的概率。

表2 数据集详细信息
Tab. 2 Detail of datasets

数据集	帖子总数	训练集帖子数	测试集帖子数	词汇总数	节点总数	标签总数
AU	65 653	52 522	13 131	18 508	71 030	600
SO	57 984	46 387	11 597	22 730	69 117	200
Math	64 983	51 986	12 997	26 214	78 200	717

3 实验与分析

本文将提出方法中的标签模型 G-RNN 应用于标签推荐任务以评估性能。本章首先描述了使用的数据集和评测标准, 其次介绍了基线和实验设置, 最后进行了性能评估和实例研究。

3.1 数据集

在本文的实验中, 使用了三个数据集: Mathematics Stack Exchange (Math), Ask Ubuntu (AU) 和 Stack Overflow (SO)。所有数据集均已正式发布并公开可用。每个数据集包含多个帖子, 每个帖子包含一段文本内容(标题和正文)以及相应的标签。实验对数据集执行以下预处理步骤。首先, 删除所有标点符号, 然后将其余的字符串拆分为单个单词, 进行去除停用词和词干还原处理。其次, 删除文本内容中的一些低频词以减少噪声, 计算每个单词的频率并将词汇表中每个数据集的最频繁出现的单词保留在词汇表中。最后, 由于很少使用低频标签, 因此同时对标签的频率进行计数, 并保留最常用的标签。本文随机选择 AU 数据集和 SO 数据集的一个子集, AU、SO、Math 分别包含 65 653、57 984、64 983 篇帖子, 18 508、22 730、26 214 个去重词汇, 71 030、69 117、78 200 个节点, 600、200、717 个标签。对于每个真实数据集, 随机选择 80% 的数据作为训练集, 并在训练集中随机选择 10% 的数据作为验证集, 其余的 20% 作为测试集, 详细数据集信息汇总如表 2

所示。

3.2 评测标准

本文采用准确率、召回率以及 F1 值对所提出方法的效果进行评估。

准确率表示推荐的标签集中相关标签所占的比例, 其计算如下所示:

$$\text{Precision}@k = \frac{|\text{tag}\lambda_{d_i} \cap \text{tag}\lambda_k|}{|\text{tag}\lambda_k|} \quad (27)$$

其中: $\text{tag}\lambda_k$ 表示推荐的标签集合, $\text{tag}\lambda_{d_i}$ 表示文本 d_i 真实标签集合。

召回率表示推荐的相关标签占有所有相关标签的比例, 其计算如下所示:

$$\text{Recall}@k = \frac{|\text{tag}\lambda_{d_i} \cap \text{tag}\lambda_k|}{|\text{tag}\lambda_{d_i}|} \quad (28)$$

其中: $\text{tag}\lambda_k$ 表示推荐的标签集合, $\text{tag}\lambda_{d_i}$ 表示文本 d_i 真实标签集合。

F1 值融合了召回率和准确率, 是它们两者的调和平均值, 其计算如下所示:

$$\text{F1}@k = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (29)$$

3.3 基线和实验设置

3.3.1 基线系统

实验选择以下基线方法进行比较。

1) Maxide^[35]: 该方法使用传统的矩阵补齐模型对文本内容进行建模推荐标签, 是一种传统的标签推荐多标签学习方法。

2) TagSpace^[12]: 该方法使用 CNN 对文本内容信息建模得到文本的嵌入表示, 通过标签得分函数对候选标签进行排序得到推荐标签。

3) LSTM^[25]: 该方法使用 LSTM 对文本内容信息进行建模得到隐藏向量, 添加 softmax 层以输出所有候选标签的推荐概率分布。

4) TLSTM^[25]: 该方法将 LSTM 模型的隐藏向量与 LDA 模型得到的主题向量通过注意力机制进行结合, 添加 softmax 层以输出所有候选标签的推荐概率分布。

5) GCN^[30]: 该模型直接使用 GCN 进行标签推荐, 将文档节点的特征向量视为文档向量, 添加 softmax 层以输出所有候选标签的推荐概率分布。注意, 该模型并未对文本内容进行建模。

6) G-LSTM: 本文提出的 G-RNN 模型的第一个变体, 本文将 GCN 层的文档嵌入向量与 LSTM 最后一层的隐藏向量通过注意力机制进行结合, 添加 softmax 层以输出所有候选标签的推荐概率分布。

7) G-BiLSTM: 本文将 GCN 层的文档嵌入向量与 Bi-LSTM 最后一层的隐藏向量通过注意力机制进行结合, 添加 softmax 层以输出所有候选标签的推荐概率分布。

8) G-GRU: 本文将 GCN 层的文档嵌入向量与 GRU 最后一层的隐藏向量通过注意力机制进行结合, 添加 softmax 层以输出所有候选标签的推荐概率分布。

9) G-BiGRU: 本文将 GCN 层的文档嵌入向量与 Bi-GRU 最后一层的隐藏向量通过注意力机制进行结合, 添加 softmax 层以输出所有候选标签的推荐概率分布。

表3 三个数据集性能比较
Tab. 3 Performance comparison of three datasets

数据集	模型	top-1			top-3			top-5		
		准确率	召回率	F1 值	准确率	召回率	F1 值	准确率	召回率	F1 值
Math	Maxide	0.126 0	0.054 0	0.075 6	0.097 0	0.121 0	0.107 7	0.081 0	0.167 0	0.109 1
	Tagspce	0.144 0	0.065 0	0.089 6	0.115 0	0.141 0	0.126 7	0.091 0	0.186 0	0.122 2
	LSTM	0.543 6	0.268 5	0.359 5	0.304 5	0.425 3	0.354 9	0.208 3	0.541 2	0.300 8
	TLSTM	0.596 8	0.282 3	0.383 3	0.323 3	0.457 0	0.378 7	0.237 4	0.563 4	0.334 0
	GCN	0.552 0	0.312 7	0.399 2	0.312 0	0.489 0	0.381 0	0.218 2	0.573 4	0.316 1
	G-BiGRU	0.598 3	0.331 5	0.426 6	0.336 2	0.502 1	0.402 7	0.244 0	0.588 8	0.345 0
AU	Maxide	0.126 0	0.054 0	0.075 6	0.097 0	0.121 0	0.107 7	0.081 0	0.167 0	0.109 1
	Tagspce	0.084 0	0.030 0	0.044 2	0.072 0	0.079 0	0.075 3	0.064 0	0.119 0	0.083 2
	LSTM	0.347 4	0.170 2	0.228 5	0.202 2	0.292 2	0.239 0	0.165 7	0.401 5	0.234 5
	TLSTM	0.444 9	0.180 1	0.256 4	0.279 1	0.340 6	0.306 8	0.222 4	0.432 1	0.293 6
	GCN	0.378 6	0.186 0	0.249 5	0.245 8	0.342 5	0.286 2	0.196 4	0.442 0	0.271 9
	G-BiGRU	0.422 4	0.210 6	0.281 1	0.256 2	0.369 8	0.302 7	0.225 9	0.459 5	0.302 9
SO	Maxide	0.121 0	0.049 0	0.069 8	0.088 0	0.110 0	0.097 8	0.075 0	0.159 0	0.101 9
	Tagspce	0.126 0	0.051 0	0.072 6	0.087 0	0.113 0	0.098 3	0.076 0	0.161 0	0.103 3
	LSTM	0.535 6	0.305 5	0.389 1	0.312 5	0.542 8	0.396 6	0.216 6	0.625 6	0.321 8
	TLSTM	0.597 5	0.314 0	0.411 7	0.356 2	0.563 1	0.436 3	0.248 5	0.653 6	0.360 1
	GCN	0.557 5	0.343 6	0.425 2	0.353 1	0.583 0	0.439 8	0.238 4	0.681 0	0.353 1
	G-BiGRU	0.598 1	0.367 3	0.455 1	0.359 4	0.605 1	0.451 0	0.245 0	0.694 3	0.362 2

3.3.2 实验设置

本文使用了两层GCN^[16]和三层RNN^[28]。对于GCN部分,本文设置第一个卷积层的嵌入大小为300,滑动窗口大小为20。对于RNN部分,本文将最大文本长度设置为40,使用词向量模型Word2Vec对单词进行预训练,单词嵌入维数设置为100。对于LSTM/GRU的神经元,其隐藏状态的维数设置为500。在训练模型时,采用Adam优化器^[36]进行模型训练。为了防止过度拟合,本文将学习率设置为0.02,dropout为0.5,损失权重为0,并在损失不再减少时停止训练过程。

本文实验平台为Windows 10系统下的PyCharm Community2019,所有跟踪实验均在配置为Intel Core i7-9750CPU,显卡为NVIDIA GeForce GTX 1650Ti,内存为16GB的计算机上完成,采用Keras深度学习框架。

3.4 性能评估

本节将通过与其他基线方法比较,证明所提方法的有效性。表3为基线模型和本文模型在不同评价指标下在三个数据集的性能比较,图2~4对比显示了使用GCN和RNN变体同时获取网络结构信息和内容信息进行推荐的性能。为了对比有效性,本文对不同数据集采用推荐top-1,top-3和top-5个标签进行对比。

表3的结果表示了本文提出的最优变体方法与Maxide、Tagspace、LSTM、TLSTM、GCN这5种方法推荐性能的比较。从图中可以看出,本文提出的模型总体上在三个数据集上的性能优于基线方法。在召回率指标上,G-BiGRU模型基于三个数据集平均提高了5.5%;在F1值指标上,G-BiGRU模型基于三个数据集平均提高了5.2%。与三种评价标准最优的基线进行比较,在Math数据集上,G-BiGRU模型准确率平均提高2.3%,召回率平均提高3.8%,F1值平均提高7.0%;在AU数据集上,G-BiGRU模型召回率平均提高8.4%,F1值平均提高3.8%;在SO数据集上,G-BiGRU模型召回率平均提高4.2%,F1值平均提高4.8%。总体上,G-BiGRU模型的准确率提高有限,原因是Math数据集文本较长,可获取更多的文本

内容信息,而其他两个数据集文本较短,对文本信息的提取效果较差。G-BiGRU模型的召回率提高较多,G-BiGRU模型的F1值较其他基线都有提升。

图2~4分别表示了本文提出的四种变体模型在不同评估标准下的推荐性能。使用GCN和Bi-GRU结合的G-BiGRU模型整体性能最高。GCN和LSTM、GRU结合的G-LSTM和G-GRU模型在融合文本间网络结构信息基础上同时考虑了文本内容信息,所以比单纯使用GCN效果好。GCN和Bi-LSTM、Bi-GRU结合的G-BiLSTM和G-BiGRU模型在融合网络结构信息基础上同时考虑了文本内容上下文信息,所以效果较好。由于Bi-GRU单元较Bi-LSTM单元更加简洁、高效和收敛速度快,较GRU单元更全面提取了文本的内容信息,所以使用RNN变体中的Bi-GRU模型与GCN模型相结合是推荐标签的理想选择。

综上所述,对于本文采用的3种评估标准,本文提出的方法比其他基线方法效果好,原因是该方法同时使用了文本间网络结构信息和文本内容信息,这表明本文的方法可以推荐更加准确的标签。虽然TLSTM使用LSTM和LDA同时提取文本的内容的隐藏信息和主题信息来推荐标签,但是仍没有考虑到文本与文本之前隐含的网络结构信息,而这些信息可以用来提高推荐的性能。虽然GCN通过提取文本间网络结构信息来推荐标签,但是没有考虑到文本内容信息,而这些信息对于推荐性能的提高也是必不可少的。当推荐标签个数逐渐增加时,所有方法的召回率都不断增加,这是因为推荐的真实标签数目越来越多,同时,由于文本的真实标签数目有限,可以预测召回率随着推荐标签个数的增多逐渐趋于稳定。与此相反,基本上所有方法随着标签个数的增加准确率逐渐变小,这是因为推荐了越来越多的不准确的标签。尽管本文的模型在AU、SO数据集上准确率没有超过每个基线模型,但是总体性能仍然非常令人满意。这证明了融合网络结构信息和文本内容信息在标签推荐中的有效性。

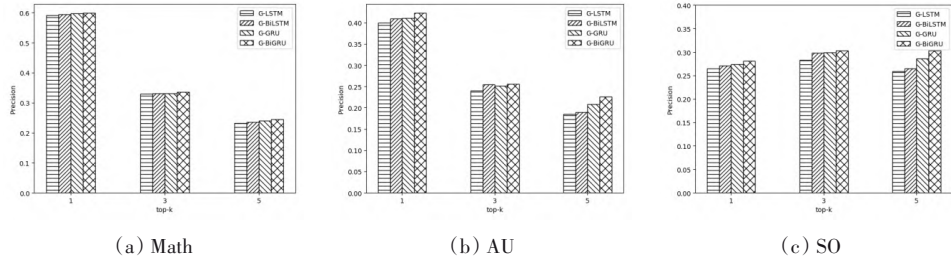


图2 各模型在三个数据集上的准确率比较

Fig. 2 Precision comparison of each model on three datasets

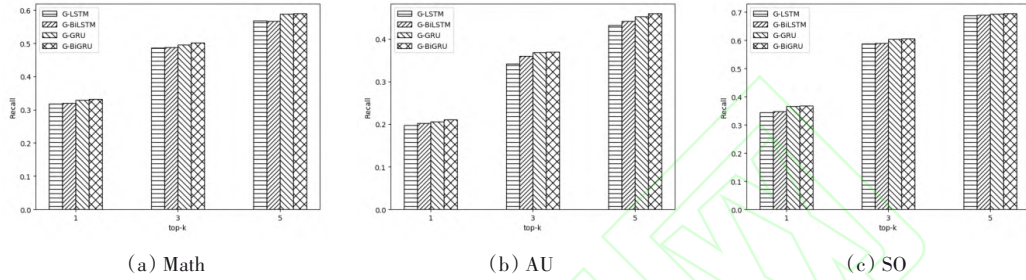


图3 各模型在三个数据集上的召回率比较

Fig. 3 Recall comparison of each model on three datasets

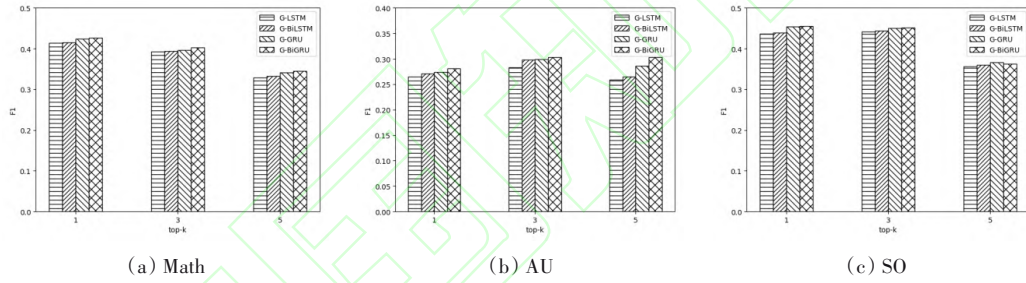


图4 各模型在三个数据集上的F1值比较

Fig. 4 F1 comparison of each model on three datasets

3.5 实例研究

为了更加清晰地体现实验的效果,首先,在测试数据集中随机挑选一篇帖子,该帖子的题目为“Implementing and Enforcing Coding Standards”,正确标签为“c#, standards, coding-style”,使用具有代表性的基线 TLSTM 模型和 GCN 模型与本文最优的变体模型 G-BiGRU 模型分别对该帖子进行标签推荐。

从表4可以看出使用 G-BiGRU 模型推荐出3个正确标签,而 TLSTM 模型和 GCN 模型仅推荐出2个正确标签。G-BiGRU 模型比 TLSTM 模型和 GCN 模型的效果好是因为 G-BiGRU 模型不仅使用了文本间的单词共现信息,还同时使用了文本内容信息。

4 结语

本文提出了一种融合网络结构信息和文本内容信息的标签推荐方法,该方法同时考虑了文本与文本之间的通过单词共现信息可获取到的隐藏网络结构信息和文本内容的隐藏信息。并使用注意力机制结合两种信息将标签推荐问题转化为一个多分类问题。结果表明,同时考虑文本的网络结构信息和内容信息可以有效地提高标签推荐的效果。目前的工作没

有考虑到标签与标签之间的语义联系,因此,在未来,本文将通过引入标签与标签之间的语义联系或其他有用的辅助信息来进行模型改进。

表4 对帖子推荐标签实例

Tab. 4 An example of tag recommendation for post

模型	得到的5个概率最大的标签		
	Number	Tag	True Tag?
TLSTM	1	standards	Yes
	2	PostgreSQL	No
	3	coding-style	Yes
	4	multithreading	No
	5	Upgrade	No
GCN	1	c#	Yes
	2	concurrency	No
	3	multithreading	No
	4	coding-style	Yes
	5	windows-vista	No
G-BiGRU	1	c#	Yes
	2	PostgreSQL	No
	3	multithreading	No
	4	standards	Yes
	5	coding-style	Yes

参考文献 (References)

- [1] 王梦恬, 魏晶晶, 廖祥文, 等. 融合评论标签的个性化推荐算法[J]. 计算机科学与探索, 2016, 10(10): 1429-1438. (WANG M T, WEI J J, LIAO X W, et al. Personalized recommendation algorithm fusing comment tag[J]. Journal of Frontiers of Computer Science & Technology, 2016, 10(10): 1429-1438.)
- [2] 孙松涛, 何炎祥. 基于CNN特征空间的微博多标签情感分类[J]. 工程科学与技术, 2017, 49(3): 162-169. (SUN S T, HE Y X. Multi-label sentiment classification of Weibo based on CNN feature space[J]. Journal of Engineering Science and Technology, 2017, 49(3): 162-169.)
- [3] 吴磊, 岳峰, 王含茹, 等. 一种融合科研人员标签的学术论文推荐方法[J]. 计算机科学, 2020, 47(2): 51-57. (WU L, YUE F, WANG H R, et al. A method for recommending academic papers based on tags of researchers [J]. Journal of Computer Science, 2020, 47(2): 51-57.)
- [4] BELÉM F M, ALMEIDA J M, GONÇALVES M A. A survey on tag recommendation methods [J]. Journal of the Association for Information Science and Technology, 2017, 68(4): 830-844.
- [5] MANISHA P, RUSHED K: Tag Recommendation by Link Prediction Based on Supervised Machine Learning[C]//Proceedings of the Sixth International Conference on Weblogs and Social Media. Menlo Park, CA: AAAI Press, 2012: 547-550.
- [6] MARTINS E F, FABIANO M B, ALMEIDA J M, et al. On cold start for associative tag recommendation [J]. Journal of the Association for Information Science and Technology, 2016, 67(1): 83-105.
- [7] KRESTEL R, FANKHAUSER P, NEJDL W. Latent dirichlet allocation for tag recommendation [C]//Proceedings of the third ACM conference on Recommender systems. New York: ACM, 2009: 61-68.
- [8] RAMAGE D, HALL D, NALLAPATI R, et al. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora [C]// Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2009, 248 - 256.
- [9] WU Y, YAO Y, XU F, et al. Tag2Word: Using Tags to Generate Words for Content Based Tag Recommendation[C]// Proceedings of the 25th ACM International Conference on Information and Knowledge Management. New York: ACM, 2016: 2287-2292.
- [10] WU Y, XI S, YAO Y, et al. Guiding supervised topic modeling for content based tag recommendation [J]. Neurocomputing, 2018, 314: 479-489.
- [11] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature 2015, 521(7553): 436-444.
- [12] WESTON J, CHOPRA S, ADAMS K. #tagspace: Semantic embeddings from hashtags [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2014, 1822 - 1827.
- [13] ZHANG Q, WANG J, HUANG H, et al. Hashtag Recommendation for Multimodal Microblog Using Co-Attention Network [C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2017: 3420-3426.
- [14] GUO Z, ZHANG H F, ZHU S H, et al. A Two-Level Topic Model Towards Knowledge Discovery from Citation Networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(4): 780-794.
- [15] JEVIN D W, IAN W S, CARL T. Bergstrom: A Recommendation System Based on Hierarchical Clustering of an Article-Level Citation Network [J]. IEEE Transactions on Big Data, 2016, 2 (2): 113-123.
- [16] KIPF T N, WELING M. Semi-Supervised Classification with Graph Convolutional Networks [EB/OL]. [2020-05-20]. <https://arxiv.org/abs/1609.02907>.
- [17] SHI X W, HUANG H Y, ZHAO S Y, et al. Tag Recommendation by Word-Level Tag Sequence Modeling [C]// Proceedings of the 2019 Database Systems for Advanced Applications. Cham: Springer, 2019: 420-424.
- [18] FENG W, WANG J Y. Incorporating heterogeneous information for personalized tag recommendation in social tagging systems [C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 1276-1284.
- [19] FANG X M, PAN R, CAO G X, et al. Personalized Tag Recommendation through Nonlinear Tensor Factorization Using Gaussian Kernel [C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2015: 439-445.
- [20] ZHAO W, GUAN Z Y, LIU Z. Ranking on heterogeneous manifolds for tag recommendation in social tagging services [J]. Neurocomputing, 2015, 148: 521-534.
- [21] SI X, SUN M. Tag-LDA for scalable real-time tag recommendation [J]. Journal of Computational Information Systems, 2008, 6(1): 1-8.
- [22] DING Z Y, QIU X P, ZHANG Q, et al. Learning Topical Translation Model for Microblog Hashtag Suggestion [C]// Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2013: 2078-2084.
- [23] FRÉDÉRIC G, SLAVKOVÍK J V, NEVE W D, et al. Using topic models for Twitter hashtag recommendation [C]//International World Wide Web Conferences Steering Committee. Rio de Janeiro, Brazil: ACM, 2013: 593-596.
- [24] GONG Y Y, ZHANG Q. Hashtag Recommendation Using Attention-Based Convolutional Neural Network [C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2016: 2782-2788.
- [25] LI Y, LIU T, JIANG J, et al. Hashtag Recommendation with Topical Attention-Based LSTM [C]//Proceedings of the 26th International Conference on Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2016: 3019-3029.
- [26] HUANG H R, ZHANG Q, GONG Y Y, et al. Hashtag Recommendation Using End-To-End Memory Networks with Hierarchical Attention [C]//Proceedings of the 26th International Conference on Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2016: 943-952.
- [27] LI J, XU H, HE X W, DENG J H, et al. Tweet modeling with LSTM recurrent neural networks for hashtag recommendation [C]// Proceedings of the 2016 International Joint Conference on Neural Networks. Piscataway: IEEE, 2016: 1570-1577.
- [28] TANG S J, YAO Y F, ZHANG S W, et al. An Integral Tag Recommendation Model for Textual Content [C]// Proceedings of

- the Thirty-Third AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2019: 5109-5116.
- [29] SHI M, TANG Y F, LIU J X. TA-BLSTM: Tag Attention-based Bidirectional Long Short-Term Memory for Service Recommendation in Mashup Creation [C]// Proceedings of the 2019 International Joint Conference on Neural Networks. Piscataway: IEEE, 2019: 1-8.
- [30] YAO L, MAO C S, LUO Y. Graph Convolutional Networks for Text Classification [C]// Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2019: 7370-7377.
- [31] BENGIO Y, DUCHARME R, VINCENT P, et al. A Neural Probabilistic Language Model [J]. Journal of Machine Learning Research, 2003, 3: 1137-1155.
- [32] MIKOLOV T, SUTSK I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality [C]// Proceedings of the 27th Annual Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2013: 3111-3119.
- [33] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [J]. Neural computation, 1997, 9(8):1735-1780.
- [34] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [EB/OL]. [2020-06-03]. <https://arxiv.org/pdf/1412.3555.pdf>.
- [35] XU M, JIN R, ZHOU Z H. Speedup Matrix Completion with Side Information: Application to Multi-Label Learning[J]. Advances in Neural Information Processing Systems, 2013:2301-2309.
- [36] KINGMA D, BA J. Adam: a method for stochastic optimization [EB/OL]. [2020-06-03]. <https://arxiv.org/pdf/1412.6980.pdf>.
- This work is partially supported by the National Natural Science Foundation of China (61876062).
- CHE Bingqian**, born in 1998, M. S. candidate. Her research interests include information retrieval, natural language processing.;
- ZHOU Dong**, born in 1979, Ph. D., professor. His research interests include information retrieval, natural language processing.