

基于图深度学习的金融文本多标签分类算法

金雨澄^{1,2}, 王清钦^{1,2}, 高 剑³, 苗仲辰³, 林越峰³, 项雅丽^{1,2}, 熊 贇^{1,2}

(1. 复旦大学 计算机科学技术学院, 上海 210438; 2. 上海市数据科学重点实验室, 上海 200438;

3. 上海金融期货信息技术有限公司, 上海 200120)

摘 要: 金融文本多标签分类算法可以根据用户需求在海量金融资讯中实现信息检索。为进一步提升金融文本标签识别能力, 建模金融文本多标签分类中标签之间的相关性, 提出基于图深度学习的金融文本多标签分类算法。图深度学习通过深度网络学习局部和全局的图结构特征, 可以刻画节点之间的复杂关系。通过建模标签关联实现标签之间的知识迁移, 是构造具有强泛化能力算法的关键。所提算法结合标签之间的关联信息, 采用基于双向门控循环网络和标签注意力机制得到的新闻文本对应不同标签的特征表示, 通过图神经网络学习标签之间的复杂依赖关系。在真实数据集上的实验结果表明, 显式建模标签之间的相关性能够极大地增强模型的泛化能力, 在尾部标签上的性能提升尤其显著, 相比 CAML、BIGRU-LWAN 和 ZACNN 算法, 该算法在所有标签和尾部标签的宏观 F1 值上最高提升 3.1% 和 6.9%。

关键词: 文本多标签分类; 深度学习; 图神经网络; 注意力网络; 金融文本

开放科学(资源服务)标志码(OSID):



中文引用格式: 金雨澄, 王清钦, 高剑, 等. 基于图深度学习的金融文本多标签分类算法[J]. 计算机工程, 2022, 48(4): 16-21.

英文引用格式: JIN Y C, WANG Q Q, GAO J, et al. Multi-label financial text classification algorithm based on graph deep learning[J]. Computer Engineering, 2022, 48(4): 16-21.

Multi-label Financial Text Classification Algorithm Based on Graph Deep Learning

JIN Yucheng^{1,2}, WANG Qingqin^{1,2}, GAO Jian³, MIAO Zhongchen³, LIN Yuefeng³,
XIANG Yali^{1,2}, XIONG Yun^{1,2}

(1. School of Computer Science and Technology, Fudan University, Shanghai 200438, China;

2. Shanghai Key Laboratory of Data Science, Shanghai 200438, China;

3. Shanghai Financial Features Information Technology Co., Ltd., Shanghai 200120, China)

[Abstract] Multi-label financial text classification can retrieve relevant information from massive financial news according to user needs. To further improve the performance of multi-label financial text classification, this study proposes an algorithm to model the correlation between labels based on graph deep learning. Graph deep learning can describe the complex relationships between nodes by learning local and global graph structure features through deep neural networks. Modeling the correlation between labels can realize knowledge transfer between labels, which is key to constructing an algorithm with strong generalization ability. Therefore, this study utilizes graph neural network to learn the complex dependency between labels based on statistical information along with feature representations extracted using the bi-directional gated recurrent network and label attention mechanism. Experimental results on real world datasets show that modeling label correlations can significantly improve the classification performance, especially on tail labels. Compared with CAML, BIGRU-LWAN and ZACNN algorithms, the proposed algorithm improves the macro F1 values of all labels and tail labels up to 3.1% and 6.9%.

[Key words] multi-label text classification; deep learning; graph neural network; attention network; financial text

DOI: 10.19678/j.issn.1000-3428.0061221

基金项目: 国家自然科学基金(U1636207, U1936213)。

作者简介: 金雨澄(1998—), 男, 硕士研究生, 主研方向为数据挖掘; 王清钦, 硕士研究生; 高 剑(通信作者), 硕士; 苗仲辰、林越峰, 博士; 项雅丽, 硕士研究生; 熊 贇, 教授、博士。

收稿日期: 2021-03-22 **修回日期:** 2021-05-18 **E-mail:** 20212010042@fudan.edu.cn

0 概述

在移动互联网时代,金融新闻资讯成为人们高效获取市场情报的主要途径。然而随着新闻数量的爆发式增长,如何准确地对金融文本进行分类用以精准推荐或辅助决策,成为亟待解决的问题。由于单条新闻文本常常同多个标签相关联,因此多标签文本分类问题受到广泛关注^[1-3]。

二元关联是解决多标签问题最常用的思路^[1-3]。二元关联把多标签分类转化为多个二分类问题^[4],每次针对一个标签类别开展,即每次判断样本是否属于某个类别。不同的标签与文本中不同的特征有关。文献[1]提出的CAML算法使用标签注意力机制为每个二分类问题进行特征提取。然而,CAML使用固定窗口大小的卷积网络融入上下文信息,导致模型只能提取固定长度的局部短语信息。文献[2-3]提出的MSATT-KG和BiGRU-LWAN算法分别使用密集连接卷积层和双向门控循环网络代替一维卷积层,从而关注到文本中不同粒度的上下文信息。文献[5]提出的ZACNN算法在CAML的基础上进一步融入标签语义特征等先验知识,以提升模型在小样本场景下的性能表现。然而,这些工作都忽略了标签之间的复杂依赖关系,二元关联中多个二分类问题并不是相互独立的。

图深度学习^[5-7]通过深度网络刻画了图节点之间的关联,其在蛋白质分子属性推断^[8]、交通流量预测^[9]、金融欺诈检测^[10]、新闻文本分类^[11]等领域得到了成功应用。在多标签文本分类中,也有一些工作通过图深度学习刻画了标签层级结构。文献[12-13]分别提出使用Tree-LSTM(Tree-structured Long Short Term Memory)网络和图卷积神经网络(Graph Convolution Network,GCN)对标签的层级结构和标签语义描述进行编码表示。上述方法都依赖于预先定义的标签层级结构和丰富的标签语义信息。然而,在金融领域,这样的标签结构信息获取依赖于专业领域人员,对标签划分层级结构的代价很大,并且由于尾部标签出现频次较低,尾部标签对应的二分类问题中正负样本数量严重失衡,但上述工作难以直接处理这种数据的不均衡性问题。

本文通过建模金融文本多标签分类中标签之间的相关性,提出基于图深度学习的金融文本多标签分类算法,在不依赖于标签层级结构等先验知识的前提下,学习语义信息以建模标签之间的复杂依赖关系。该算法根据原始数据集中的标签分布构建标签关联图,对新闻文本使用双向门控循环网络进行上下文语义嵌入,并将通过标签注意力机制得到的文本特征表示作为图上对应标签节点的属性信息,进一步使用图神经网络融合标签之间的关联与文本

信息得到新闻的特征表征,利用多个线性层预测新闻文本在标签空间中的概率分布。同时,为解决二元关联中尾部标签正负样本严重不均衡的问题,选用非对称损失函数^[14]作为优化目标。

1 问题定义

给定一段新闻文本,经过分词、去除停用词等处理,将原始输入文本转换为字符序列 $d=[w_1, w_2, \dots, w_m]$,其中: w_i 表示输入序列中的第*i*个单词; m 表示输入文本长度。标签空间被定义为集合 $\{l_1, l_2, \dots, l_L\}$,其中: L 代表标签空间中的标签数量。每个新闻文本可能属于一个或多个标签,表示为标签集合 C 。本文多标签文本分类任务是学习一个映射函数 $f(\cdot)$,将字符序列映射为一个标签集合 C ,即 $f(d)=C, 0 \leq |C| \leq L$ 。具体地,输入新闻样本 d ,对于每个标签 l_i ,模型输出该样本属于标签 l_i 的概率 \hat{p}_i ,最终根据输入样本关于 L 个标签的概率构造输出标签集合 C 。

2 基于图深度学习的多标签分类算法

本节将详细介绍本文提出的基于图深度学习的多标签文本分类模型FMLG,其中所使用的重要符号定义如表1所示。

表1 重要符号定义

符号	含义
$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$	训练集, x_i 和 y_i 为文本和对对应标签
$[w_1, w_2, \dots, w_m]$	分词得到的字符序列
h_i^{init}	字符 w_i 的初始嵌入表示
h_i^c	融入上下文信息的嵌入表示
d_{init}	初始嵌入表示维度
d_c	融入上下文信息的嵌入表示维度
L	标签数量
q_i	注意力头
f_i^a	由标签注意力机制提取的文本特征
f_i^{inter}	融入标签关联的文本特征
$v_i^{(t)}$	图上节点在第 t 层的特征表示

2.1 模型框架

图1展示了FMLG模型的基本框架。输入新闻文本 d 。首先,将原始字符序列转换为嵌入表示,并进一步表示为融合上下文信息的语义嵌入向量,由于不同标签会侧重不同的文本特征,因此多标签注意力网络使用多个注意力头提取标签相关的文本特征;然后,为了捕捉标签之间的依赖关系并学习更丰富的语义信息,在构建标签关联图的基础上,模型使用门控图神经网络实现标签间的语义特征交互;最后,多个线性层被用于预测标签的概率分布。

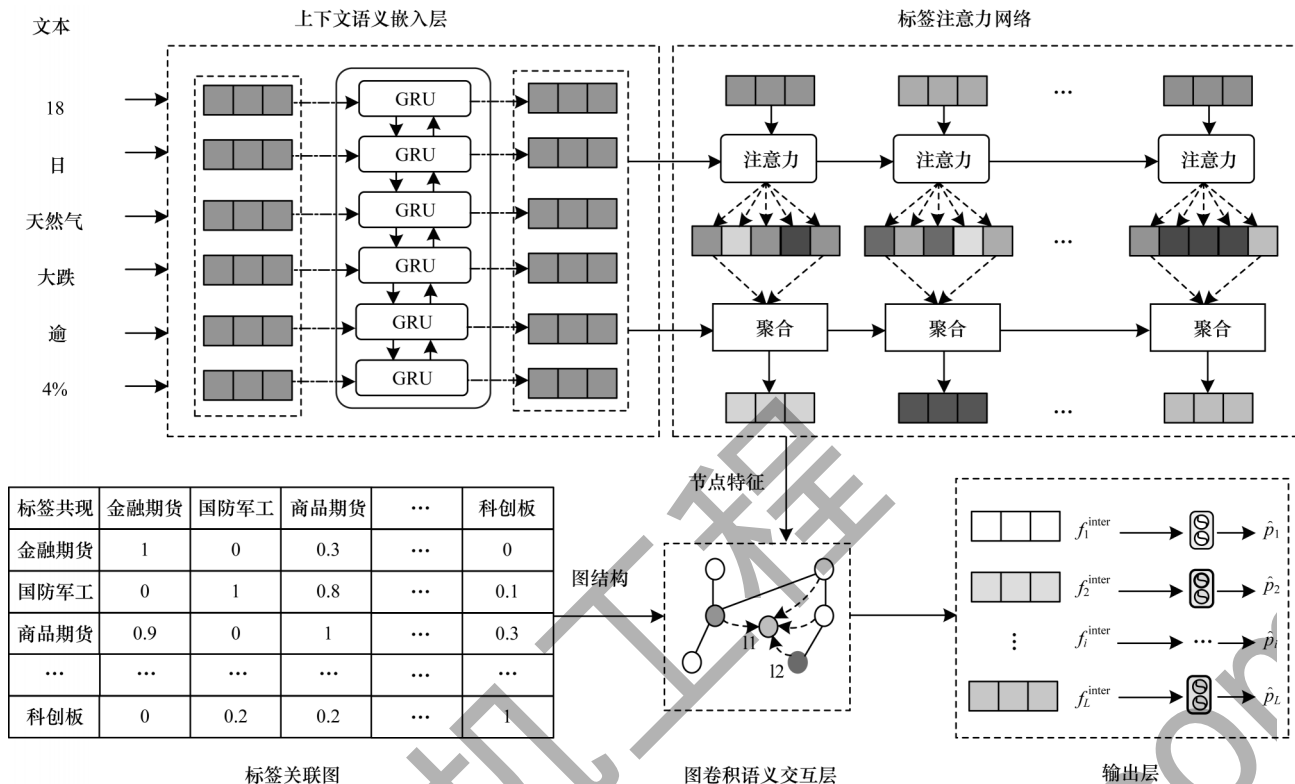


图1 FMLG模型框架

Fig.1 Framework of FMLG model

2.2 上下文语义嵌入

JOHNSON等^[15]指出,预训练的词嵌入表示可以使模型训练过程更加稳定,因此,使用word2vec^[16]进行嵌入表示学习。单词序列经过嵌入表示后得到向量序列 $\mathbf{H} = \langle \mathbf{h}_0^{\text{init}}, \mathbf{h}_1^{\text{init}}, \dots, \mathbf{h}_m^{\text{init}} \rangle$, $\mathbf{h}_i^{\text{init}} \in \mathbb{R}^{d_{\text{vec}}}$ 。为了在嵌入表示中融入上下文信息,进一步使用双向门控循环网络对向量序列 \mathbf{H} 进行编码,得到一系列文本的上下文语义嵌入表示 $\langle \mathbf{h}_0^c, \mathbf{h}_1^c, \dots, \mathbf{h}_m^c \rangle$, $\mathbf{h}_i^c \in \mathbb{R}^{d_c}$ 。

2.3 多标签注意力网络

由于文本中常常包含大量冗余信息且每个标签侧重不同角度的文本特征,因此FMLG使用标签注意力 $\{q_1, q_2, \dots, q_L\}$ 进行特征提取,对每个注意力向量有 $q_i \in \mathbb{R}^{d_c}$ 。注意力权重计算方式如下:

$$\alpha_{ij} = \frac{\exp(q_i^T \mathbf{h}_j^c)}{\sum_{k=1}^m \exp(q_i^T \mathbf{h}_k^c)} \quad (1)$$

$$\mathbf{f}_i^a = \sum_{j=1}^m \alpha_{ij} \mathbf{h}_j^c \quad (2)$$

其中: α_{ij} 代表文本中第 j 个字符与标签 l_i 的相关程度; \mathbf{f}_i^a 代表同标签 l_i 相关的文本向量表示。最终可以得到文本的向量表示 $\{\mathbf{f}_1^a, \mathbf{f}_2^a, \dots, \mathbf{f}_L^a\}$ 。

2.4 图语义交互层

尽管FMLG模型利用了标签注意力来提取与各个标签相关的文本特征,但这样的做法只是在标签与文本之间建立联系,仍然无法捕捉标签之间的关系。本文先利用统计关系对标签进行构图,再通过门控神

经网络^[17]对其进行特征交互。这不仅可以在模型中显式地融入标签之间的关联,而且其中的门控机制也可以自适应地捕捉到更丰富的相关文本特征。

本文模型首先根据训练集构建标签关联图 G ,图 G 为有向带权图,其中节点集 V 由数据集中的所有标签组成,边权 A_{ij} 被定义为标签 l_i 和 l_j 之间的条件概率,条件概率通过训练数据集估算得到,即:

$$A_{ij} = p(l_j | l_i) \quad (3)$$

$$p(l_j | l_i) \approx \frac{\sum_{k=1}^N I(l_i, l_j \in y_k)}{\sum_{k=1}^N I(l_i \in y_k)} \quad (4)$$

其中: I 为指示函数。给定图 G ,模型使用门控图神经网络进行特征交互。图上节点的初始向量表示设置为注意力层提取到的特征,即 $\mathbf{v}_i^{(0)} = \mathbf{f}_i^a$ 。首先通过图上的信息流动从邻节点聚合信息:

$$\tilde{\mathbf{v}}_i^{(k)} = \frac{\sum_j A_{ji} \mathbf{v}_j^{(k)}}{\sum_j A_{ji}} \quad (5)$$

然而从邻节点聚合到的信息可能存在噪声,因此模型使用门控机制自适应选择有效信息,信息聚合与传递的具体过程如下:

$$\mathbf{r}_i^{(k)} = \sigma(\mathbf{W}_{r1} \tilde{\mathbf{v}}_i^{(k)} + \mathbf{W}_{r2} \mathbf{v}_i^{(k)} + \mathbf{b}_r) \quad (6)$$

$$\mathbf{z}_i^{(k)} = \sigma(\mathbf{W}_{z1} \tilde{\mathbf{v}}_i^{(k)} + \mathbf{W}_{z2} \mathbf{v}_i^{(k)} + \mathbf{b}_z) \quad (7)$$

$$\mathbf{n}_i^{(k)} = \tanh(\mathbf{W}_{n1} \tilde{\mathbf{v}}_i^{(k)} + \mathbf{b}_{n1} + \mathbf{r}_i^{(k)} * (\mathbf{W}_{n2} \mathbf{v}_i^{(k)} + \mathbf{b}_{n2})) \quad (8)$$

$$\mathbf{v}_i^{(k+1)} = (\mathbf{1} - \mathbf{z}_i^{(k)}) * \mathbf{n}_i^{(k)} + \mathbf{z}_i^{(k)} * \mathbf{v}_i^{(k)} \quad (9)$$

其中: σ 代表Sigmoid函数; $*$ 代表Hadamard乘积; $\mathbf{r}_i^{(k)}$ 和 $\mathbf{z}_i^{(k)}$ 代表更新门和重置门; \mathbf{W}_* 和 \mathbf{b}_* 分别代表可训练权重和偏置。重复上述步骤 K 次,得到语义交互后的文本向量表示 $\{\mathbf{v}_1^{(K)}, \mathbf{v}_2^{(K)}, \dots, \mathbf{v}_L^{(K)}\}$ 。为了防止深层图神经网络出现的过平滑现象,模型进一步加入了残差连接,语义交互层最终得到文本表示 $\{\mathbf{f}_1^{\text{inter}}, \mathbf{f}_2^{\text{inter}}, \dots, \mathbf{f}_L^{\text{inter}}\}$ 。同标签 l_i 相关联的特征 $\mathbf{f}_i^{\text{inter}}$ 定义如下:

$$\mathbf{f}_i^{\text{inter}} = \mathbf{v}_i^{(K)} \parallel \mathbf{v}_i^{(0)} \quad (10)$$

其中: \parallel 代表拼接操作。

最终,对于每个类别,模型训练一个二分类器:

$$\hat{p}_i = \sigma(\mathbf{w}_{p_i}^T \mathbf{f}_i^{\text{inter}} + b_{p_i}) \quad (11)$$

其中: \hat{p}_i 为标签 l_i 的预测输出概率; \mathbf{w}_{p_i} 和 b_{p_i} 是对应的二分类器可训练参数权重和偏置; σ 代表sigmoid函数。

2.5 损失函数

由于大部分标签只在少数样本中出现,因此采用二元关联解决多标签分类会导致二分类问题中正负样本比例的严重失衡。本文采用非对称损失函数(Asymmetric Loss)^[14]处理非均衡分布问题。非对称损失通过权重衰减因子和置信度阈值平衡正负样本在损失中的占比。预测概率 \hat{p}_i 同真实概率 p_i 之间的非对称损失函数 L_{asy} 计算如下:

$$L_{\text{asy}}(\hat{p}_i, p_i) = -p_i L^+ - (1 - p_i) L^- \quad (12)$$

$$L^+ = \ln(\hat{p}_i) \quad (13)$$

$$L^- = \ln(1 - \hat{p}_i) \tilde{p}_i \quad (14)$$

$$\tilde{p}_i = \max(\hat{p}_i - m, 0) \quad (15)$$

其中: L^+ 和 L^- 分别代表样本作为正类和负类时带来的损失。权重衰减因子 γ 用于为易分负样本的损失赋予更小的权重,置信度阈值 m 用于忽略预测置信度较高负样本带来($\hat{p}_i < m$)的损失。上述2个参数可以减少负样本在损失中的占比,使模型更注重正样本产生的优化信息。

3 实验与结果分析

本节通过实验验证FMLG模型的有效性,并分析讨论实验中的场景数据以及相关的参数设置。

3.1 数据集

实验使用的金融新闻数据集中共包含84 707条中文金融新闻文本,文本的平均长度为598个中文字符。每条文本由新闻标题和内容两部分组成。数据集每条文本都被打上多个金融类话题标签,共包含115个标签。数据集中每条文本平均被打上1.5个标签,35%的文本标签数量大于1,13%的标签在数据集中出现次数少于100次。

3.2 对比算法

下面介绍本文采用的对比算法:

1)CAML^[1]。由于文本中存在大量冗余信息且不同标签对应着不同的文本特征,因此该方法提出在卷积神经网络的基础上,进一步使用注意力机制为每个标签进行特征提取。

2)BiGRU-LWAN^[3]。由于卷积神经网络只能提取局部信息,而双向GRU可以捕捉长距离依赖,因此该方法将CAML的卷积神经网络替换为双向GRU。BiGRU-LWAN在多个多标签文本分类数据集上都取得了极佳的效果,是一个极具竞争力的对比算法。

3)ZACNN^[5]。RIOS等提出在CAML中融入标签的语义信息以进一步提升模型在尾部低频标签上的性能表现。

3.3 实验设置

实验预处理阶段使用jieba(<https://github.com/fxsjy/jieba>)对文本进行分词,并取文档频率大于4次的字符组成词表。预训练嵌入表示维度设置为100。随机划分70%、15%、15%的数据分别作为训练集,测试集和验证集,并将在训练集中出现次数少于100次的标签视为尾部少样本标签。

对于本文提出的FMLG算法,经双向GRU编码后的特征维度 d_e ,门控图神经网络隐藏层维度都设置为300。非对称损失函数中的 γ 和 m 分别被设置为2和0.05。对于CAML算法,CNN的卷积核大小设置为3,特征维度 d_e 设置为100。BiGRU-LWAN的特征维度设置为300。ZACNN的卷积核大小设置为3。

为了使得模型有着更好的泛化能力,实验中对于上述所有模型的词嵌入层向量表示和注意力系数使用概率为0.2的Dropout^[18]。训练阶段优化器选用Adam^[19],学习率设置为 10^{-4} ,ZACNN训练轮数设为30轮,其余模型设置为10轮。本文实验中选择广泛使用的评价指标精度、召回率和F1值来评估模型的性能对比,并基于验证集选择最优的模型进行测试。

3.4 实验结果

在FMLG模型上分别进行2组实验。FMLG/BCE和FMLG/ASL分别表示使用交叉熵和非对称损失函数训练的FMLG模型。表2、表3展示了模型在所有标签和尾部标签上的性能表现,其中,加粗数据表示最优结果。

表2 模型在所有标签上的性能表现

指标	微观			宏观		
	精度	召回率	F1值	精度	召回率	F1值
CAML	88.60	84.62	86.56	75.74	66.89	71.04
BIGRU-LWAN	90.17	86.60	88.21	79.31	72.76	75.89
ZACNN	81.59	69.87	75.28	59.07	41.89	49.02
FMLG/BCE	89.35	88.24	88.79	80.67	75.29	77.89
FMLG/ASL	87.35	91.06	89.17	78.28	79.66	78.97

表3 模型在尾部标签上的性能表现						
Table 3 Model performance on tail labels						%
指标	微观			宏观		
	精度	召回率	F1值	精度	召回率	F1值
CAML	70.68	52.89	60.50	67.85	52.78	59.37
BIGRU-LWAN	75.72	60.51	67.27	73.75	60.06	66.21
ZACNN	57.78	30.02	39.51	48.32	29.85	36.90
FMLG/BCE	76.92	64.67	70.26	78.42	64.50	70.78
FMLG/ASL	72.48	72.98	72.73	74.38	71.91	73.13

从表2、表3所列出的实验数据可以看出：

1)ZACNN模型效果最差。ZACNN中标签的嵌入表示通过标签名称中字符的嵌入表示取平均得到,但是由于标签名称同对应文本特征存在较大的语义鸿沟,因此导致模型表达能力欠缺。

2)BIGRU-LWAN效果优于CAML,表明卷积网络只能提取固定长度的局部信息的特点具有局限性,双

向GRU可以更好地建模上下文之间的长距离依赖。

3)相比BIGRU-LWAN,FMLG/BCE取得了更好的结果,且在尾部标签上的提升更为明显,这表明显式地建模标签之间的关联能够提升模型性能。

4)对比FMLG/BCE和FMLG/ASL可以发现,使用非对称损失函数能够大幅提升模型在正负样本失衡条件下的召回率与F1值。

为了更加直观,在表4中进一步列出部分测试样例预测结果,其中文本中的下划线部分代表新闻中的重要信息。从中可以发现,相比性能最优的对比算法BIGRU-LWAN算法,FMLG可以预测出更为完整的标签集合。以样例1为例,由于标签“A股策略”和“研判优选”在训练集中存在共现关系,它们在标签关联图中存在连边,FMLG可以通过图语义交互层学习两者之间的依赖关系,从而实现更为准确的预测输出。

表4 部分测试样例预测结果				
Table 4 Prediction results of some test samples				
样例	文本	BIGRU-LWAN 预测结果	FMLG 预测结果	标签之间有无图上连边
样例1	今日,在兴业证券2020年度策略会上,兴业证券研究院首席策略分析师王德伦表示,在国家重视、居民配置、机构配置、全球配置等指引下,真正属于中国的权益时代有望正式开启,A股正在经历第一次“长牛”机会。行业配置上,短中期看好高股息资产的配置价值,以金融、地产、交运、电力公用等行业龙头为代表的高股息标的有……	A股策略	A股策略,研判优选	有
样例2	【国内商品期市收盘涨跌互现能化品多数上涨】财联社11月21日讯,国内商品期市收盘涨跌互现,能化品多数上涨,燃油涨超3%,橡胶、郑醇、塑料涨逾1%;基本金属涨跌不一,沪锡涨逾2%,沪镍、沪铅跌近1%;农产品多数下跌,鸡蛋跌逾2%;黑色系多数下跌,螺纹钢、线材跌逾1%……	商品期货	商品期货,钢铁	有
样例3	昨日晚间,新希望(000876.SZ)发布公告,称根据生猪发展战略,公司拟斥资近10亿元收购两家养殖企业。<p><p>此次收购,意味着公司离相关目标更进一步。上市公司在接受财联社记者采访时表示:“目前公司‘2022年实现2500万头生猪出栏量’这一中长期规划正在按部就班地推进……	农林牧渔	农林牧渔,猪肉相关	有

3.5 消融实验

为表明语义交互层带来的提升不完全是由于更深层的网络结构导致,实验中还将图神经网络中的邻接矩阵替换为单位阵。从表5实验结果可以发现,使用单位阵代替标签关联图会导致性能下降。这表明在模型中通过构建标签关联图的方式显式建模标签之间的关联具有重要意义。

表5 不同邻接矩阵对实验结果的影响				
Table 5 Influence of adjacency matrix to experimental result				
指标	所有标签		尾部标签	
	微观F1值	宏观F1值	微观F1值	宏观F1值
单位阵	88.95	78.72	70.51	71.56
标签关联图	89.17	78.97	72.73	73.13

此外,还通过实验探究不同图神经网络结构对实验结果的影响,实验结果见表6。可以发现,门控

神经网络取得了更好的效果,进一步验证了从邻节点聚合到的信息往往存在噪声,门控机制可以从自适应选择有效信息。

表6 不同图神经网络对实验结果的影响				
Table 6 Influence of different graph neural networks to experimental result				
指标	所有标签		尾部标签	
	微观F1值	宏观F1值	微观F1值	宏观F1值
图卷积网络 ^[20]	88.55	77.68	66.68	69.46
门控图神经网络	89.17	78.97	72.73	73.13

4 结束语

本文提出基于图深度学习的多标签文本分类算法FMLG,通过标签统计信息构建关联图,并利用门控图神经网络挖掘标签之间的关系。在与各个算法的对比实验中,FMLG在所有标签和尾部标签的宏观F1值上最高取得了2.0%和4.5%的提升,这表明显式

建模标签之间的关系可以大幅提升模型的泛化能力。为了进一步解决二元关联中正负样本不均衡的问题,FMLG使用非对称损失函数作为优化目标。相比现有算法,FMLG在所有标签和尾部标签宏观F1值上最高取得了3.1%和6.9%的提升。由于现实场景中标签数量较多,数据集中的样本常常只被打上部分标签,使用存在标签缺失的样本训练模型会对性能造成较大的负面影响。后续将在本文工作的基础上使用PU learning^[21-22](Positive-Unlabelled learning)进一步模型在标签缺失场景下的鲁棒性。

参考文献

- [1] MULLENBACH J, WIEGREFFE S, DUKE J, et al. Explainable prediction of medical codes from clinical text [C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Stroudsburg, USA: ACL, 2018: 1101-1111.
- [2] XIE X C, XIONG Y, YU P S, et al. EHR coding with multi-scale feature attention and structured knowledge graph propagation [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2019: 649-658.
- [3] CHALKIDIS I, FERGADIOTIS E, MALAKASIOTIS P, et al. Large-scale multi-label text classification on EU legislation [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: ACL, 2019: 6314-6322.
- [4] 马慧芳,贾美惠子,李晓红,等. 一种基于标签关联关系的微博推荐方法[J]. 计算机工程, 2016, 42(4): 197-201, 208.
MA H F, JIA M H Z, LI X J, et al. A microblog recommendation method based on label correlation relationship [J]. Computer Engineering, 2016, 42(4): 197-201, 208. (in Chinese)
- [5] RIOS A, KAVULURU R. Few-shot and zero-shot multi-label learning for structured label spaces [C]//Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2018: 3132-3142.
- [6] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs [C]//Proceedings of 2014 International Conference on Learning Representations. Banff, Canada: [s. n.], 2014: 1-14.
- [7] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs [C]//Proceedings of the 33rd International Conference on Machine Learning. New York, USA: JMLR, 2016: 2014-2023.
- [8] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks [C]//Proceedings of 2018 International Conference on Learning Representations. Vancouver, Canada: [s. n.], 2018: 1-12.
- [9] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2017: 1025-1035.
- [10] GUO S N, LIN Y F, FENG Y, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting [C]//Proceedings of 2019 AAAI Conference on Artificial Intelligence. Hawaii, USA: AAAI, 2019: 922-929.
- [11] 刘月,翟东海,任庆宁. 基于注意力CNLSTM模型的新闻文本分类[J]. 计算机工程, 2019, 45(7): 303-308, 314.
LIU Y, ZHAI D H, REN Q N. News text classification based on CNLSTM model with attention mechanism [J]. Computer Engineering, 2019, 45(7): 303-308, 314. (in Chinese)
- [12] WANG D X, LIN J B, CUI P, et al. A semi-supervised graph attentive network for financial fraud detection [C]//Proceedings of 2019 IEEE International Conference on Data Mining. Washington D. C., USA: IEEE Press, 2019: 598-607.
- [13] XIE P T, XING E. A neural architecture for automated ICD coding [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, USA: ACL, 2018: 1066-1076.
- [14] BEN-BARUCH E, RIDNIK T, ZAMIR N, et al. Asymmetric loss for multi-label classification [EB/OL]. [2021-04-01]. <https://arxiv.org/abs/2009.14119>.
- [15] JOHNSON R, ZHANG T. Convolutional neural networks for text categorization: shallow word-level vs. deep character-level [EB/OL]. [2021-04-01]. <https://arxiv.org/abs/1609.00718>.
- [16] LE Q V, MIKOLOV T. Distributed representations of sentences and documents [C]//Proceedings of the 31st International Conference on Machine Learning. Stockholm, Sweden: JMLR, 2018: 1188-1196.
- [17] LI Y J, TARLOW D, BROCKSCHMIDT M, et al. Gated graph sequence neural networks [C]//Proceedings of 2016 International Conference on Learning Representations. San Juan, USA: [s. n.], 2016: 273-283.
- [18] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [19] KINGMA D P, BA J. Adam: a method for stochastic optimization [C]//Proceedings of 2015 International Conference on Learning Representations. San Diego, USA: [s. n.], 2015: 1-15.
- [20] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [C]//Proceedings of 2016 International Conference on Learning Representations. San Juan, USA: [s. n.], 2016: 1-14.
- [21] KIRYO R, NIU G, PLESSIS M C, et al. Positive-unlabeled learning with non-negative risk estimator [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. [S. l.]: Curran Associates, 2017: 1674-1684.
- [22] SHU S, LIN Z, YAN Y, et al. Learning from multi-class positive and unlabeled data [C]//Proceedings of 2020 IEEE International Conference on Data Mining. Washington D. C., 2020: 1256-1261.