



基于矩阵分解和注意力多任务学习的客服投诉工单分类

宋勇¹, 严志伟², 秦玉坤², 赵东明³, 叶晓舟¹, 柴园园¹, 欧阳晔¹

(1. 亚信科技(中国)有限公司, 北京 100193; 2. 亚信科技(南京)有限公司, 江苏 南京 210013;
3. 中国移动通信集团天津有限公司, 天津 300020)

摘要: 投诉工单自动分类是通信运营商客服数字化、智能化发展的要求。客服投诉工单类别有多层, 每一层有多个标签, 层级之间有所关联, 属于典型的层次多标签文本分类问题, 现有解决方法大多数基于分类器同时处理所有的分类标签, 或者对每一层级分别使用多个分类器进行处理, 忽略了层次结构之间的依赖。提出了一种基于矩阵分解和注意力的多任务学习的方法(MF-AMLA), 处理层次多标签文本分类任务。在通信运营商客服场景真实投诉工单分类数据下, 与该场景常用的机器学习算法和深度学习算法的 Top1 F1 值相比分别最大提高了 21.1% 和 5.7%。已在某移动运营商客服系统上线, 模型输出的正确率 97% 以上, 对客服坐席单位时间的处理效率提升 22.1%。

关键词: 层次多标签分类; 注意力机制; 多任务学习; 客服工单分类

中图分类号: TP183

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2022031

Customer service complaint work order classification based on matrix factorization and attention multi-task learning

SONG Yong¹, YAN Zhiwei², QIN Yukun², ZHAO Dongming³,

YE Xiaozhou¹, CHAI Yuanyuan¹, OUYANG Ye¹

1. AsiaInfo Technologies (China) Co., Ltd., Beijing 100193, China

2. AsiaInfo Technologies (Nanjing) Co., Ltd., Nanjing 210013, China

3. China Mobile Communications Group Tianjin Co., Ltd., Tianjin 300020, China

Abstract: The automatic classification of complaint work orders is the requirement of the digital and intelligent development of customer service of communication operators. The categories of customer service complaint work orders have multiple levels, each level has multiple labels, and the levels are related, which belongs to a typical hierarchical multi-label text classification (HMTC) problem. Most of the existing solutions are based on classifiers to process all classification labels at the same time, or use multiple classifiers for each level, ignoring the dependence between hierarchies. A matrix factorization and attention-based multi-task learning approach (MF-AMLA) to deal with hierarchical multi-label text classification tasks was proposed. Under the classification data of real complaint work orders in the customer service scenario of communication operators, the maximum Top1 F1 value of MF-AMLA is increased by 21.1% and 5.7% respectively compared with the commonly used machine learning algo-



rithm and deep learning algorithm in this scenario. It has been launched in the customer service system of one mobile operator, the accuracy of model output is more than 97%, and the processing efficiency of customer service agent unit time has been improved by 22.1%.

Key words: hierarchical multi-label classification, attention mechanism, multi-task learning, customer service work order classification

0 引言

近年来,通信运营商的客户数量稳步上升,业务范围增大,面临的投诉问题日趋繁多和复杂^[1]。快速准确地处理投诉工单是提升客户满意度和运营商效益的重要方面。针对每个投诉工单,首先需要判断其所属的业务类别。传统的依赖客服坐席标记投诉工单所属类别需要耗费大量的人力成本和时间成本,而且不能保证很高的准确率。因此,对投诉工单文本进行自动化分类是运营商朝着数字化、智能化方向发展的要求^[2]。

通信运营商客服场景的投诉工单文本相较于一般的文本,通常对应很多层级的业务类别,并且各类别之间存在一定的关联关系^[3],每个层级的业务类别串联形成最终的总类别,属于典型的层次多标签文本分类(hierarchical multi-label text classification, HMTc)问题。传统的文本分类方法很少考虑类别层级的关联关系,文本特征表示相对较弱,分类效果仍有提升空间。因此,为了实现运营商投诉工单文本的自动分类以及保证高的准确率,需要采取更有针对性的有效分类方法。本文创新性地提出一种基于矩阵分解(matrix factorization, MF)和注意力的多任务学习(attention-based multi-task learning approach, AMLA)的方法,被称为MF-AMLA,处理HMTc任务,帮助运营商客服快速准确地对多层级的投诉工单文本进行分类。

1 相关工作

1.1 多标签文本分类

文本分类是自然语言处理(NLP)中的一个

经典任务,旨在为给定的文本序列分配预定义类别^[4-5]。早期的文本分类算法主要基于浅层学习,如朴素贝叶斯(naive bayes)、支持向量机(support vector machine, SVM)、 K 最邻近分类(K -nearest neighbor, KNN)和随机森林等,这些方法存在高维度、高稀疏问题,分类效果差。近年来,深度学习的快速发展使得神经网络在文本分类领域大放异彩。例如,文本卷积神经网络(text convolutional neural network, TextCNN)^[6]能够捕捉局部相关性,在短文本分类上取得较好效果。循环神经网络(recurrent neural network, RNN)^[7-8]考虑了上下文信息的语义关联。长短期记忆(long short-term memory, LSTM)网络^[9]可以学习长距离依赖信息,解决了传统RNN梯度消失的问题。门控循环单元(gated recurrent unit, GRU)^[10]作为LSTM的一种变体,结构更加简单。Transformer的自注意力机制在处理长期依赖的问题时相较于LSTM有更好的表现^[11]。基于此,GPT(generative pre-training)预训练词向量模型^[12]、BERT(bidirectional encoder representations from transformers)预训练语言模型^[13],依次刷新各NLP任务效果。

多标签文本分类是文本分类的重要分支,相比于单标签分类任务,多标签通常有复杂的依赖性,每个文本不只属于一个类别,而是被赋予多个标签。Tsoumakas等^[14]将多标签分类的方法分为两类,一类是问题转换的方法,另一类是算法适应的方法。前者将多标签分类问题转化为一个或多个单标签分类问题,代表算法有Binary Relevance^[15]、Random k -labelsets^[16]、Calibrated Label Ranking^[17]、Label Power-set^[18]。这类方法面对类

别多、数据量大的数据集时存在计算复杂度高的缺陷。后者通过拓展单标签学习算法,直接处理多标签的数据,代表算法有 Rank-SVM^[19]、ML-KNN^[20]。

而在许多实际问题中,文本建模的预测输出是结构化的类别,即类别之间存在依赖关系,例如表现为树状分层结构或有向无环图^[21]。这种多标签分类属于 HMTc^[22]。HMTc 分类任务,主要方法可以分为两类:一类为局部方法,根据类别标签的层次结构,对预测输出空间的每个分类建立模型,将 HMTc 分类转化为对多个单一标签的分类,最后组合各个模型获得最终的标签^[23]。例如, Cesa-Bianchi 等^[24]提出了 Hierarchical SVM,仅在父类标签被激活时进行 SVM 学习;另一类为全局方法,预测输出空间建立单一模型预测所有类别,相比于局部方法,可挖掘、使用各标签之间的依赖关系,学习到隐含特征获得更好的分类效果^[25]。例如, Vens 等^[23]提出 Clus HMC,使用单个决策树处理整个层次类别结构。Borges 等^[26]提出了一种基于竞争人工神经网络的全局方法预测层次结构中的所有类别。然而,上述这些研究只关注到了层次结构的局部区域或整体结构,而忽略了层次结构不同层级之间的依赖关系。

1.2 多任务学习

多任务学习 (multi-task learning, MTL) 是对多个任务共享模型,同时学习,提升模型性能的算法^[27]。MTL 有许多优点,例如可以提高数据计算效率,通过共享表示可以减少过度拟合,利用辅助信息可以快速学习等^[28]。MTL 同时考虑多个相关任务的学习过程,能够利用任务之间的关联提高单一任务学习的泛化能力。例如, Liu 等^[29]利用 16 个不同的文本分类任务进行 MTL,相较于单任务显著提升了分类效果。MTL 现有方法主要有两类:硬参数共享^[28,30]和软参数共享^[31-32]。前者在不同任务的隐藏层共享模型权重,使每个

权重都经过训练,共同最小化多个损失函数。后者针对不同的任务有特定的任务模型,在参数空间讨论需要交互哪些信息,设定不同的权重,通过将不同模型参数之间的距离加入联合目标函数进行优化。

综合以上相关研究,基于 HMTc 的特点,本文创新性地提出了 MF-AMLA。矩阵分解 (matrix factorization, MF) 是一种协同过滤推荐算法^[33],通过将用户评分矩阵分解以学习用户和物品之间的交互关系。这里,本文引入 MF 模块,增强文本和分层多标签的语义表示,以捕捉文本与层次标签的浅层隐藏关系。然后对不同分层的多标签进行基于注意力的 MTL,既考虑局部层级特征以及全局信息,又考虑不同层级之间的关联关系,挖掘、使用各层级标签之间的依赖关系,学习更深层的隐含特征。MF 和 MTL 联合损失函数,在各子任务间互相修正错误,提升模型对层次多标签文本的分类效果。

2 模型结构

MF-AMLA 模型结构如图 1 所示,主要涉及 3 部分内容:文本表征层、嵌入层、AMLA 层。

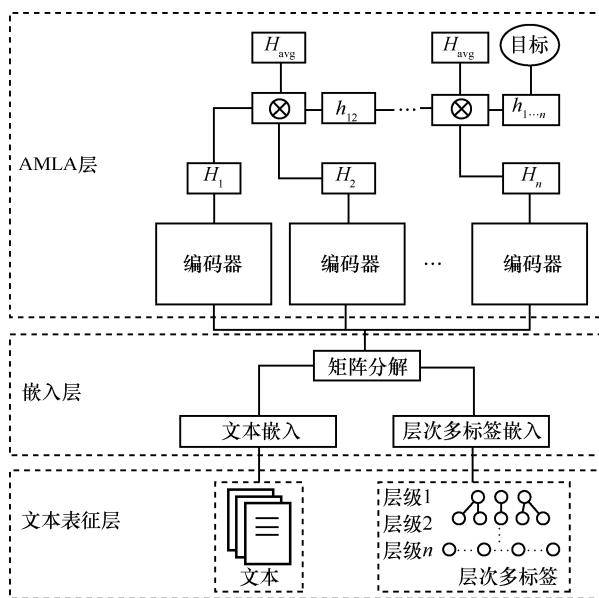


图 1 MF-AMLA 模型结构



2.1 文本表征层

对于 HMTL 任务, 涉及 Text 和层次多标签 (hierarchical multi-label, HML), Text 表示为:

$$t_i = (t_1, t_2, \dots, t_i) \quad (1)$$

其中, i 代表文本的条数, t_i 代表第 i 条文本。

HML 表示为:

$$l_i = (l^1, l^2, \dots, l^j) \quad (2)$$

其中, $l^j = \{l'_1, l'_2, l'_3, \dots, l'_j\} \in \{0, 1\}^{|l^j|}$, l^j 表示层级, l'_j 表示分层多标签。

2.2 嵌入层

在基于注意力机制的多任务学习的第一阶段, 为了使文本和分层多标签类别的向量表示统一, 本文首先在嵌入 (embedding) 层编码文本和分层多标签类别结构, 然后使用 MF 层增强文本语义表示的编码。

嵌入层对 t_i 和 l_i 进行编码, 采取的编码方式使用 Bert 预训练模型对文本进行特征表示, 输出抽取后的嵌入序列。 t_i 经过嵌入后表示为 $T_i = (T_1, T_2, \dots, T_i)$, 其中, T_i 代表第 i 条文本的嵌入向量。 l_i 经过嵌入后表示为 $L_i = (L^1, L^2, \dots, L^j)$, 其中, $L^j = \{E_1, E_2, \dots\} \in \{0, 1\}^{|L^j|}$, L^j 和 E 分别表示层级和该层级多标签的嵌入向量。

本文把文本和分层多标签类别组合成矩阵 S :

$$S = (S_{ij}) \in (T_i, L_j) \quad (3)$$

MF 模块用来增强文本和分层多标签类别的隐含信息, 对矩阵 S , 使 T_i 分类到 L_j , 求解目标为:

$$\text{Classification}(T, L) = r_{wl} = p_w^T q_l = \sum_f p_{w,k} q_{k,l} \quad (4)$$

其中, p_w 和 q_l 都为列向量。 $p_{w,k}$ 和 $q_{k,l}$ 是需要计算的参数, $p_{w,k}$ 度量第 w 个文本和第 k 个隐类标签的关系, $q_{k,l}$ 度量第 k 个隐类标签和第 l 个分层多标签类别的关系。

$$\text{Loss}_{\text{MF}} = \sum_{(w,k) \in K} (r_{wl} - p_w^T q_l)^2 \quad (5)$$

使用梯度下降算法降低损失:

$$\begin{aligned} p_{w,k} &= p_{w,k} - \eta(-e_{ul} q_{k,l}) = p_{w,k} + \eta e_{ul} q_{k,l} \\ q_{k,l} &= q_{k,l} - \eta(-e_{ul} p_{w,k}) = q_{k,l} + \eta e_{ul} p_{w,k} \end{aligned} \quad (6)$$

2.3 AMLA 层

AMLA 层包含两个模块: Transformer 和 MTL。

(1) Transformer

本文首先对由 MF 层输出的文本表示 r_{wl} 进行层次提取, 提取出的文本和标签的合成信息 X 作为输入该层级对应的 Transformer 做编码器 (encoder) 处理。

每个编码器包含两个模块, 分别为自注意力 (self-attention) 模块和前馈网络模块。在自注意力模块, 输入的嵌入 X 被线性变换成 3 个矩阵: 查询 Q 、键 K 、值 V 。缩放点积注意力为:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

多头注意力机制将 Q 、 K 、 V 拆分成 h 个头, 并行执行注意力操作, 再将多头的输出值串联, 投影后形成输出。经多头注意力机制后再接前馈网络层, 得到句子的特征向量 H 。

这里, 每个层级标签都对应一个 Transformer 的编码器层, 每个编码器层包含 6 个编码器模块。Transformer 层的输出为 n 个特征向量 $(H_1, H_2, \dots, H_n)$, n 为标签的层级数目。

(2) MTL

MTL 旨在通过各层级语义信息的交互获取各层级标签之间的隐藏关系, 最终输出总的层级多标签。如图 1 所示, 对于 MTL 模块, 其输入是前一层 Transformer 层输出的多个层级的特征向量 $(H_1, H_2, \dots, H_n)$, 以递归的方式进行注意力 (attention) 交互,

$$\begin{aligned} h_{1,2,\dots,i} &= \text{softmax}(h_{1,2,\dots,i-1} H_i^T) H_{\text{avg}} \\ h_{12} &= \text{softmax}(H_1 H_2^T) H_{\text{avg}} \end{aligned} \quad (8)$$

其中, $H_{\text{avg}} = \text{avg}(H_1, H_2, \dots, H_n)$ 。比如, 层级一对应的 Transformer 层输出的语义向量 H_1 与层级二的 H_2 进行点积 (dot-product) 注意力计算, 再乘上所有层级的平均向量, 得到第一步 MTL 的输出结果 h_{12} , 以此作为第二步 MTL 的一个输入, 与层级三的 H_3 再进行注意力交互。以此类推, 最后得到文本所有层级标签的 MTL 得分 $h_{1,2,\dots,n}$ 。在每一步 MTL 操作中, 点积考虑了局部的层级信息, 再融合 H_{avg} 保持了全局的信息, 综合利用局部和全局的信息可以更好地获取层级标签父类与子类之间相互关联又相互制约的特性。

最后通过 $h_{1,2,\dots,n}$ 计算最终的标签目标。

2.4 损失函数

本模型采取的多任务学习的共享方式是软参数共享, 将层次多标签的分类任务分成若干个递归注意力子任务。总的损失函数需要结合所有子任务的损失函数。在多任务的背景下, 每个子任务具有不确定性, 其权重不同。本文提出一种动态分配损失函数权重的方法, 达到在各子任务效果上的平衡。具体算法如下: 每训练固定步数, 计算一次各子任务损失函数值与总损失函数值的比例, 如果比值低于一定阈值 γ , 则调整一次损失函数权重, 使加权后各子任务损失函数值占总损失函数值的比例不低于阈值 γ , 从而使模型能够时刻关注每个子任务的变化。总的损失函数为:

$$L = \sum_i \alpha_i L_i \quad (9)$$

其中:

$$\alpha_i = \max(1, \gamma L / L_i) \quad (10)$$

为动态分配每个子任务损失函数的权重, γ 是可调阈值 (默认取 $1/N$, N 为任务数目), L_i 为子任务的交叉熵损失函数:

$$L_i(p, q) = -\sum_{j=1}^n p(x_j) \lg q(x_j) \quad (11)$$

其中, $p(x_j)$ 为该文本真实标签的概率, $q(x_j)$ 为预测标签的概率。

3 实验结果及分析

3.1 数据集介绍

本文使用的数据集来自运营商客服接听电话投诉后人工标注好的文本。投诉工单文本实例见表 1。数据筛选前的原始标签的长尾分布见表 2。考虑实际生产实践中着重关注占比高的处理方式, 本文按照 2/8 原则, 筛选前 80% 的头部标签数据进行后续的下游分类任务, 对于剩余 20% 的长尾标签数据则不采用。筛选后, 共有 283 904 条投诉工单文本, 原始的标签数量为 186 个。进行预处理后, 得到 7 个层级子标签 (见表 1 的样本实例), 一级到七级的标签数量分布为 4、5、45、49、6、83、64 (文本标签分布见表 3), 例如一级标签对应的类别有: 家庭业务、移动业务、集团业务、增值业务 4 个类别。之后对数据集划分, 按照原始标签 8:2 的比例将其划分为训练集和验证集。

表 1 投诉工单文本实例

文本	客户来电反映本机预约了宽带安装, 但是装维人员电话: 1xxxxxxxxx, 服务态度非常不好, 一直没有帮客户安装, 而且讲话非常不耐烦。客户对此不满, 来电投诉。请查询处理, 谢谢, 随时联系 1xxxxxxxxx
标签	(一级) 家庭业务 → (二级) 服务触点 → (三级) 装维人员 → (四级) 全局流转 → (五级) 售后服务 → (六级) 装维人员服务态度差 → (七级) 解释差错
文本	客户反映: 于 2019 年开始在昆明龙泉路, 某小区无法使用移动数据, 已补换卡均无法使用, 提示网络错误! 强烈要求我方工作人员上门处理, 并赔偿 (未详细说明赔偿事宜), 解释无效, 请老师帮忙处理
标签	(一级) 移动业务 → (二级) 网络质量 → (三级) 手机上网 (4G) → (四级) 有信号无法使用 → (五级) 功能使用 → (六级) 有信号无法浏览网页 → (七级) 室外

表 2 数据筛选前的文本原始标签的长尾分布

占比	70%	80%	90%	100%
标签数	73	186	889	3 692

表 3 文本标签分布

层级	原始	一级	二级	三级	四级	五级	六级	七级
数量	186	4	5	45	49	6	83	64



3.2 评估方法

因数据分布不均衡,为了更全面地评价模型的分类效果,本文采用对原始标签分类的精确率 P (Precision, $P = TP/(TP + FP)$)、召回率 R (Recall, $R = TP/(TP + FN)$) 以及它们的调和平均值 $F1$ ($F1 = 2PR/(P + R)$) 为评价指标对模型效果进行评价。其中, TP 为真正例, FP 为假正例, FN 为假负例。

3.3 实验与分析

为了评估对运营商行业层次多标签文本的分类效果,本文基于第 3.1 节获取的数据集对比了本模型 MF-AMLA 与 5 种其他广泛应用的评价不错的模型: Hierarchical SVM、Clus HMC、GRU、TextCNN、Bert。前两种是基于机器学习的层次标签分类模型,后 3 种是基于深度学习的多标签分类方法。

为了多方面评估模型效果,本文引入 Top k 准确率评估方法,即置信度最高的 k 个结果中包含正确标签的占比。本文采用 Top1、Top5、Top10 3 个指标。经过对比实验,各模型的精确率 P 、召回率 R 的对比见表 4,各模型 F1 值对比如图 2 所示。整体来看,本文的 MF-AMLA 效果最优。其中, Hierarchical SVM 与 Clus HMC 出现较早,未采用深度学习的方法,效果较差,与之相比, MF-AMLA 在 Top1、Top5、Top10 的 F1 值分别最大提高了 21.1%、26.7%、22.1%; GRU、TextCNN 和 BERT 模型作为文本分类常见深度学习方法,与之相比, MF-AMLA 在 Top1、Top5、Top10 的 F1 值分别最大提高了 5.7%、4.8%、4.8%。

表 4 各模型的精确率 P 、召回率 R 的对比

模型	Top 1		Top 5		Top 10	
	P	R	P	R	P	R
Hierarchical SVM	0.413	0.507	0.642	0.681	0.733	0.759
Clus HMC	0.449	0.477	0.665	0.688	0.741	0.763
GRU	0.622	0.597	0.874	0.886	0.921	0.918
TextCNN	0.614	0.608	0.89	0.882	0.938	0.934
BERT	0.666	0.642	0.926	0.915	0.97	0.955
MF-AMLA	0.671	0.662	0.938	0.921	0.971	0.964

以上实验结果表明,本文模型通过 MF 模块将文本和分层多标签类别的向量表示统一,提升了模型的表征能力,再引入 AMLA 递归注意力多任务学习对层级标签分类后,提升了模型学习层级标签之间关联关系的能力。本文提出的 MF-AMLA 可以有效提升运营商行业中投诉工单文本的分类效果。

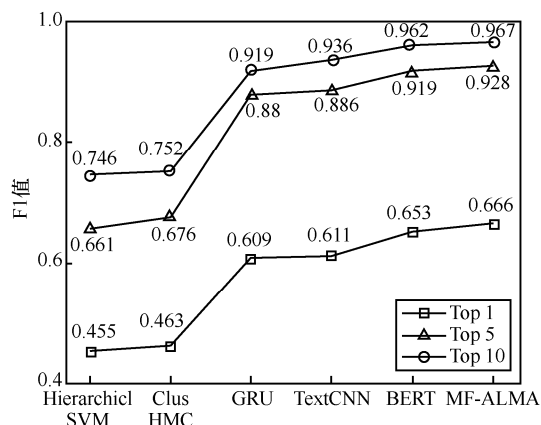


图 2 各模型 F1 值对比

4 结束语

运营商的投诉工单文本属于层次多标签的文本,目前一般的分类方法难以获得好的分类效果。本文提出一种基于矩阵分解和注意力多任务学习的层次多标签分类方法 (MF-AMLA),针对运营商领域的投诉工单文本含有层次多标签的特点,加入了 MF 模块增强文本和分层多标签的语义表示,其次,按层级关联关系进行了递归注意力多任务学习,解决 HMTC 问题,提高了投诉工单分类的准确率。该模型已在某移动运营商客服系统上线,对客服坐席接入的实时投诉数据, MF-AMLA 输出 Top 5 分类结果,辅助客服坐席快速定位用户投诉或咨询的问题,及时协助用户解决问题。客服坐席反馈模型输出的正确率在 97%以上,对客服坐席单位时间的处理效率提升 22.1% (平均一通投诉或通话时长由 58 s 降低为 49 s)。

未来本文将考虑更多文本分类任务的实践,如电商售后服务咨询内容的自动分类,根据用户反馈

的内容,给出用户的真实意图并归类,快速解决用户的诉求;用于智能对话机器人中,通过对问答内容的精准分类,快速定位用户意图,给出正确的交流策略;以及新闻自动归类,对新闻网站的内容,进行自动归类和分折,支持舆情分析,为公共安全和会治理做出贡献等,验证其他类似的任务是否有进一步的效果提升。

参考文献:

- [1] TUNG F C. Customer satisfaction, perceived value and customer loyalty: the mobile services industry in China[J]. *African Journal of Business Management*, 2013, 7(18): 1730-1737.
- [2] LIN L, ZHU B, WANG Q, et al. A novel 5G core network capability exposure method for telecom operator[C]//*Proceedings of 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking*. Piscataway: IEEE Press, 2020: 1450-1454.
- [3] WEHRMANN J, CERRI R, BARROS R. Hierarchical multi-label classification networks[C]//*Proceedings of International Conference on Machine Learning*. [S.l.:s.n.], 2018: 5075-5084.
- [4] KOWSARI, MEIMANDI J, HEIDARYSAFA, et al. Text classification algorithms: a survey[J]. *Information*, 2019, 10(4): 150.
- [5] MINAE S, KALCHBRENNER N, CAMBRIA E, et al. Deep learning: based text classification[J]. *ACM Computing Surveys*, 2021, 54(3): 1-40.
- [6] ZHANG Y, WALLACE B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[J]. *arXiv preprint arXiv:1510.03820*, 2015.
- [7] MIKOLOV T, KOMBRINK S, BURGET L, et al. Extensions of recurrent neural network language model[C]//*Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing*. Piscataway: IEEE Press, 2011: 5528-5531.
- [8] SOCHER R, LIN C C-Y, NG A Y, et al. Parsing natural scenes and natural language with recursive neural networks[C]//*Proceedings of the 28th International Conference on Machine Learning*. [S.l.:s.n.], 2011.
- [9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [10] DEY R, SALEM F M. Gate-variants of Gated Recurrent Unit (GRU) neural networks[C]//*Proceedings of 2017 IEEE 60th International Midwest Symposium on Circuits and Systems*. Piscataway: IEEE Press, 2017: 1597-1600.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//*Advances in neural information processing systems*. [S.l.:s.n.], 2017: 5998-6008.
- [12] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding with unsupervised learning[EB]. 2018.
- [13] DEVLIN J, CHANG M-W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.
- [14] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Mining multi-label data[M]//*Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2009: 667-685.
- [15] BOUTELL M R, LUO J B, SHEN X P, et al. Learning multi-label scene classification[J]. *Pattern Recognition*, 2004, 37(9): 1757-1771.
- [16] TSOUMAKAS G, VLAHAVAS I. Random k-labelsets: an ensemble method for multilabel classification[C]//*Machine Learning: ECML*. [S.l.:s.n.], 2007.
- [17] FÜRNKRANZ J, HÜLLERMEIER E, LOZA MENCÍA E, et al. Multilabel classification via calibrated label ranking[J]. *Machine Learning*, 2008, 73(2): 133-153.
- [18] MADJAROV G, KOCEV D, GJORGJEVIKJ D, et al. An extensive experimental comparison of methods for multi-label learning[J]. *Pattern Recognition*, 2012, 45(9): 3084-3104.
- [19] ELISSEEFF A, WESTON J. A kernel method for multi-labelled classification[J]. *Advances in neural information processing systems*, s.l.: The MIT Press, 2001(14): 681-687.
- [20] ZHANG M L, ZHOU Z H. ML-KNN: a lazy learning approach to multi-label learning[J]. *Pattern Recognition*, 2007, 40(7): 2038-2048.
- [21] BI W, KWOK J T. Multilabel classification on tree-and dag-structured hierarchies[C]//*Proceedings of the 28th International Conference on Machine Learning*. [S.l.:s.n.], 2011.
- [22] GONG J B, TENG Z Y, TENG Q, et al. Hierarchical graph transformer-based deep learning model for large-scale multi-label text classification[J]. *IEEE Access*, 2020(8): 30885-30896.
- [23] VENS C, STRUYF J, SCHIETGAT L, et al. Decision trees for hierarchical multi-label classification[J]. *Machine Learning*, 2008, 73(2): 185-214.
- [24] CESA-BIANCHI N, GENTILE C, ZANIBONI L. Incremental algorithms for hierarchical classification[J]. *The Journal of Machine Learning Research*, 2006(7): 31-54.
- [25] LEVATÍĆ J, KOCEV D, DŽEROSKI S. The importance of the label hierarchy in hierarchical multi-label classification[J]. *Journal of Intelligent Information Systems*, 2015, 45(2): 247-271.
- [26] BORGES H B, NIEVOLA J C. Multi-Label Hierarchical Classification using a Competitive Neural Network for protein function prediction[C]//*Proceedings of 2012 International Joint Conference on Neural Networks (IJCNN)*. Piscataway: IEEE Press, 2012: 1-8.



- [27] CARUANA R. Multitask learning[J]. Machine learning, 1997, 28(1): 41-75.
- [28] COLLOBERT R, WESTON J. A unified architecture for natural language processing: deep neural networks with multitask learning[C]//Proceedings of the 25th international conference on Machine learning-ICML '08. New York: ACM Press, 2008: 160-167.
- [29] LIU X D, HE P C, CHEN W Z, et al. Multi-task deep neural networks for natural language understanding[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019.
- [30] EATON E, DESJARDINS M, LANE T. Modeling transfer relationships between learning tasks for improved inductive transfer[C]//Machine Learning and Knowledge Discovery in Databases. [S.l.:s.n.], 2008.
- [31] DUONG L, COHN T, BIRD S, et al. Low resource dependency parsing: cross-lingual parameter sharing in a neural network parser[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2015.
- [32] YANG Y, HOSPEDALES T M. Trace norm regularised deep multi-task learning[J]. arXiv preprint arXiv:1606.04038, 2016.
- [33] BENNETT J, LANNING S. The netflix prize[C]//Proceedings of KDD Cup and Workshop. New York: ACM Press, 2007: 35.

[作者简介]



宋勇 (1989-), 男, 亚信科技(中国)有限公司通信人工智能实验室通信业务与应用算法研究部负责人, 主要研究方向为 NLP、知识图谱、AIOps、推荐等。



严志伟 (1994-), 男, 博士, 亚信科技(南京)有限公司通信人工智能实验室算法工程师, 主要研究方向为 NLP、AIOps。



秦玉坤 (1987-), 男, 亚信科技(南京)有限公司通信人工智能实验室算法工程师, 主要研究方向为 NLP、AIOps、知识图谱。

赵东明 (1984-), 男, 博士, 中国移动通信集团天津有限公司技术专家, 天津移动 AI 实验室/天津移动博士后科研工作站负责人, 主要研究方向为知识图谱、智能语音情感、认知概念网络。

叶晓舟 (1980-), 男, 博士, 亚信科技(中国)有限公司通信人工智能实验室资深总监、首席科学家, 主要研究方向为通信网络与人工智能。

柴园园 (1980-), 女, 博士, 亚信科技(中国)有限公司通信人工智能实验室首席算法科学家, 主要研究方向为深度学习、人工智能、数据科学及管理。

欧阳晔 (1981-), 男, 博士, 亚信科技(中国)有限公司首席技术官、高级副总裁, 主要研究方向为移动通信、人工智能、数据科学、科技研发创新与管理。