

姜焱松, 马敬东, 赵冬, 等. 基于长短期记忆模型的患者投诉自动分类研究[J]. 中华医学图书情报杂志, 2018, 27(6): 16-21.

DOI: 10.3969/j.issn.1671-3982.2018.06.003

· 研究与探讨 ·

基于长短期记忆模型的患者投诉自动分类研究

姜焱松, 马敬东, 赵冬, 罗玮, 倪维斌, 夏晨曦

[摘要] 患者投诉是改善医疗服务管理的珍贵资源。针对患者投诉分析需要大量人力、时间成本以及分类存在主观影响的问题, 使用一种科学的分类标准对收集到的患者投诉进行人工标注, 设计了一种基于长短期记忆模型的多个二元分类器结合的方法对患者投诉语料进行分类, 探索了分类器对于单标签语料、多标签语料和仿真语料的分类预测的性能, 为患者投诉分类提供了一种全新高效的方法, 为更好地理解患者投诉打下坚实基础。

[关键词] 患者投诉; 自动分类; 长短期记忆模型; 人工标注; 标签

[中图分类号] R197.323

[文献标志码] A

[文章编号] 1671-3982(2018)06-0016-06

LSTM model-based automatic classification of patient complaints

JIANG Yao-song, MA Jing-dong, ZHAO Dong, LUO Wei, NI Wei-bin, XIA Chen-xi

(Central China University of Science and Technology Tongji Medical College Medical and Health Management School, Wuhan 430030, Hubei Province, China)

Corresponding author: XIA Chen-xi

[Abstract] Patient complaints are a valuable resource for improving medical service management. A LSTM model-based automatic classification method of patient complaints with multiple binary classifiers was thus designed in accordance with the vast amount of manpower, time, costs and subjective influencing factors required for the analysis of patient complaints, and was used in classification of the words and phrases in patient complaints. The performance of classifiers was studied in classification of single-labeled words and phrases, multi-labeled words and phrases, and simulated words and phrases in order to provide a brand-new and effective method for the classification of patient complaints and to lay a solid foundation for the better understanding of patient complaints.

[Key words] Patient complaints; Automatic classification; LSTM model; Artificial annotation; Tags

患者投诉是指患者参与医疗活动时, 由于医务人员医疗机构未能满足他们的预期而产生的抱怨^[1]。患者投诉中包含了大量与医疗服务质量和

医院管理水平相关的信息, 如果利用得当, 能够改善医疗服务质量, 提升医院管理水平, 有效避免医患纠纷^[2-5]。日常工作中, 医疗机构会收到大量患者投诉, 这些投诉会针对不同的问题^[4]。处理患者投诉的第一步就是进行患者投诉分类, 以便将不同的问题分发给不同的科室进行处理。然而人工处理这些患者投诉会消耗大量的人力与时间成本。另外处理人员由于主观观念的差异, 对于同一个问题有不同看法, 导致对患者投诉内容的理解存在偏差, 进一步影响处理效率。在实际工作中, 一条患者投诉可能会描述多个问题, 即有多个分类标签, 如果仍然局限于从一个角度理解的患者投诉分类, 必然导致患者

[基金项目] 中央高校基本科研业务费资助华中科技大学自主创新基金项目“面向社交网络的情感分析与观点挖掘方法研究”(0118516036)

[作者单位] 华中科技大学同济医学院医药卫生管理学院, 湖北 武汉 430030

[作者简介] 姜焱松(1993-), 男, 辽宁本溪市人, 在读硕士研究生。

[通讯作者] 夏晨曦(1975-), 男, 湖北武汉市人, 博士, 讲师, 发表论文 20 余篇, 研究方向为数据挖掘、自然语言处理。E-mail: xcxdy@hust.edu.cn

投诉内容无法被充分理解,进而导致处理患者投诉问题时出现错误。

针对以上问题,本文引进了一套科学的患者投诉分类标准,并根据实际数据调整了该分类标准。参照该标准对收集到的患者投诉语料进行人工标注,解决人员主观性对分类的影响。利用深度学习的方法,构建基于长短期记忆模型(Long Short Term Memory, LSTM)的多个二元分类器,旨在实现高效的患者投诉自动分类,并探究分类器在医院实际应用的情况,为更好地理解患者投诉打下坚实基础。

1 患者投诉相关研究

国内外对患者投诉进行了大量研究。患者投诉对某一个领域的医生与医疗质量的影响是该领域研究的一个重要方向。Catron TF 等^[6]研究了患者投诉与外科手术事件的关系,得出了患者投诉与外科手术不良事件发生率成正相关的结果;Hiivala N 等^[7-8]分别使用牙科病人的投诉进行了牙科病人安全事件的预防干预与有安全隐患的牙医诊所的检测。患者投诉管理系统也是该领域研究的重要部分。Levin CM 等^[9]展示了斯坦福医疗保健相关部门设计的一套病人投诉数据管理系统,通过这套系统可以简化患者的投诉收集和解决流程,从而提高医疗质量;Mirzoev T 等^[10]收集并分析了相关文献,得出了有效干预是提高患者投诉管理系统的重要举措。更多的研究者则着眼于患者投诉的分类标准研究。Montini T 等^[11]开发了一个由 22 个病人投诉代码和 5 个提供者代码组成的患者投诉分类标准;Reader TW 等^[4]汇总了来自 59 项研究的投诉编码分类方法,从中选取 729 个经过特殊处理的代码,并将这些代码细化后概念化为一份含有 3 个大类及 7 个小类的患者投诉分类标准;Harrison R 等^[12]证明了 Reader T W 等设计的患者投诉分类同样适用于严重投诉案件的分类。

国内对于患者投诉的研究主要集中于收集患者投诉并进行人工分类与分析。如收集儿科门急诊患者投诉并进行分析,通过对患者投诉内容有针对性地改进,满足了患儿与家长的需求^[13];对 174 例门诊采血患者投诉的分析^[14];分析 278 例住院患者的医疗投诉^[15]等。国内针对患者投诉自动分类的研究更加稀少。

2 数据与方法

2.1 数据来源

选取某大型三甲医院 2012 年到 2017 年的 8 000 条患者投诉,经过去重合并,去除无意义条目后最终得到实验语料 7 872 条。

2.2 方法

本文的技术路线如图 1 所示。

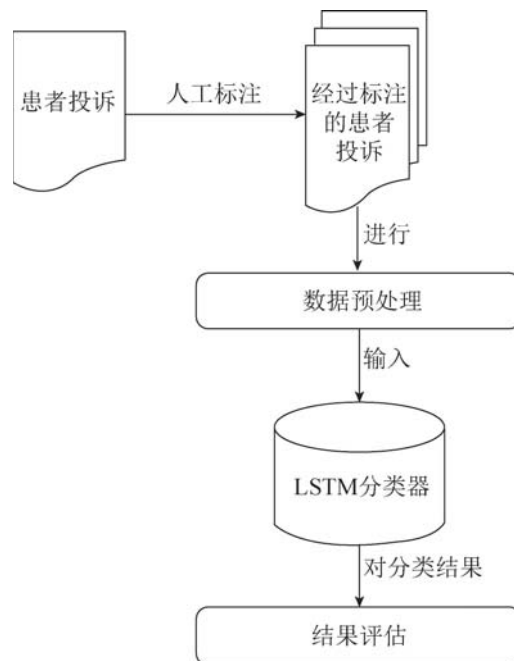


图 1 本文技术路线图

2.2.1 人工标注

本研究根据经过调整的 Reader TW 等^[4]人的患者投诉分类标准,由两组经过严格培训的专业人员对样本数据进行人工标注之后,对标记结果进行一致性检验,并对不一致的标记进行审核校正。本文选用 Kappa 系数^[16]检验两组专业人员标注的一致性。kappa 系数的计算结果最小值是 0,最大值是 1,通常用 5 个区间来代表一致性的不同级别,分别是最低的一致性系数区间为 0 到 0.2、一般的一致性系数区间为 0.21 到 0.4、中等水平的一致性系数区间为 0.41 到 0.6、较高的一致性系数区间为 0.61 到 0.8 以及最高的一致性系数区间为 0.81 到 1。

2.2.2 数据预处理

本研究采用了 python 平台的 jieba 分词包,并引入自定义词表进行中文分词、去除停用词,通过抓取医学文献摘要、百科语料、医学专业书籍等形成 1.95

亿的词汇语料库,使用 word2vec 进行训练,最后得到 200 维高质量的词向量。将分词的结果与词向量进行映射,得到患者投诉文本的分布式表示,作为 LSTM 神经网络的矩阵输入。

2.2.3 分类方法

患者投诉分类是文本分类的子问题,其应用的方法与使用的理论基础与文本分类基本相同。由于不需要人工提取特征且解决了循环神经网络长期依赖导致的信息遗忘问题^[17],基于深度学习^[18]的 LSTM 模型在文本分类任务中具有更好的性能^[19-21],研究者逐渐把这种高效的方法应用在文本分类领域。

Shih, CH 等^[22]使用一种基于连体 LSTM 方法对 IMDB 和 20-Newsgrups 中的文本进行了分类实验; Jiang, MY 等^[23]则提出了一种基于深信念网络和 softmax 回归的混合文本分类模型。赵明等^[24]使用 LSTM 模型对饮食健康文本进行了分类,邓三鸿等^[25]研究了 LSTM 模型在中文图书的标签分类中的应用,两者均取得了相较于传统机器学习更好的结果。可见 LSTM 模型在文本分类领域有着广泛的应用和更好的效果。

故而本研究选择 LSTM 作为分类模型,并根据 LSTM 模型构建相应的分类器来对患者投诉语料进行自动分类。本研究采用基于 Java 语言的 Deeplearning4j 平台进行所有基于 LSTM 模型的分分类器的构建与训练。

具体分类方法如下:首先基于 LSTM 模型构建患者投诉分类器(以下简称分类器),再将患者投诉分为训练数据集与测试数据集,将训练数据集输入构建好的基于 LSTM 模型的分分类器中进行训练,通过测试数据集反映分类器的性能。经过多轮训练,分类器性能趋于稳定并达到一个较高的水平时,就可以使用该分类器进行患者投诉自动分类。

2.2.4 分类结果评估指标

本研究通过准确率(Precision, P)、召回率(Recall, R)以及 F 值(F-measure)3 类指标对实验测试结果进行评价。三个指标值均在 0-1 之间,结果越接近 1,说明分类模型性能越好。

3 患者投诉分类标准与人工标注

3.1 患者投诉分类标准

为保证分类标准的科学性,采用 Reader TW 等人基于 59 项研究的投诉编码分类方法构建的具有 3 个大类(临床问题、管理问题、关系问题)及 7 个小类(质量、安全、环境、管理制度、沟通、尊重和患者权利、倾听)的患者投诉分类标准^[4]。由于使用的患者投诉语料实际情况不同,对该患者投诉分类标准进行了微调。根据人工标注结果,本研究使用的患者投诉语料中分到原分类标准中“关系”大类下“倾听(listening)”小类的投诉只有 96 例,并且这些语料表达的内容与“尊重和患者权利”基本一致,故将“倾听”小类合并到“尊重和患者权利”小类中。具体分类标准如表 1 所示。

表 1 患者投诉分类标准

投诉大类	投诉小类	投诉小类的相关内涵
临床问题(与医护人员提供的临床和护理质量和安全有关的问题)	质量	医务人员未及时关注并改善患者身体状况 医务人员未按照临床行为标准为患者提供医疗服务 治疗效果不理想或出现副作用
	安全	医务人员在诊断、用药等临床过程中出现错误 由于医务人员的失误或者能力不足等问题导致患者处于危险状态
管理问题(与医疗服务环境和医院管理制度有关的问题)	环境	医院病房住宿环境差 医院硬件设施与人员等医疗资源不足 医院安保条件差
	管理制度	病人未能及时入院或获得治疗 医院管理流程出现问题 与病历管理有关的问题
	沟通	医务人员与患者没有沟通或沟通不及时 医务人员与患者沟通中出现错误
关系问题(与医院工作人员对病人或其家属行为有关的问题)	尊重和患者权利	医院工作人员不尊重患者 医院工作人员侵犯患者隐私等权利

本研究根据分类标准中的 6 个投诉小类(以下简称类别)对收集到的患者投诉进行自动分类。

3.2 人工标注结果与分析

通过计算两组标注人员的 kappa 系数比较标注的一致性。经计算,本研究标注的 kappa 系数如表 2 所示。

表 2 标注人员标注的 kappa 系数

	质量	安全	环境	制度	沟通	尊重	总体
kappa 系数	0.68	0.80	0.88	0.81	0.81	0.89	0.81

除临床大类下的“质量”类别以外,其他类别标注的 kappa 系数都达到较高的一致性,总体标注的一致性也达到较高水平,说明标注结果达到了相当满意的标注一致性。对标注不一致的语料进行审核校正,使所有语料都有确定的唯一的标注结果。患者投诉语料标注的结果如表 3 所示。

表 3 患者投诉语料标注结果

类别	语料数量	类别	语料数量
质量	673	质量安全环境	3
安全	660	质量安全沟通	5
环境	2265	质量安全尊重	7
制度	1056	质量环境制度	1
沟通	741	质量环境沟通	3
尊重	1618	质量环境尊重	4
质量安全	30	质量制度沟通	1
质量环境	29	质量制度尊重	1
质量制度	14	质量沟通尊重	10
质量沟通	58	安全环境制度	2
质量尊重	41	安全环境沟通	2
安全环境	48	安全环境尊重	2
安全制度	22	安全制度沟通	1
安全沟通	22	安全制度尊重	2
安全尊重	96	安全沟通尊重	5
环境制度	117	环境制度尊重	6
环境沟通	20	环境沟通尊重	2
环境尊重	67	制度沟通尊重	3
制度沟通	12	质量安全制度尊重	2
制度尊重	75	质量安全沟通尊重	2
沟通尊重	143	质量环境制度尊重	1
总计			7872

从表 3 可以看出,患者投诉语料总计 7 872 条,被分到了 42 个分类之中。其中单标签患者投诉 7 013 条,占 89.09%;多标签语料 859 条,占 10.91%;具有两个标签的患者投诉 794 条,占 10.09%;具有三个标签的患者投诉 60 条,占 0.76%;具有四个标签的患者投诉 5 条,占 0.06%。

4 实验结果与分析

标注结果中有些语料被标注为多个标签,即一条语料并非只被标记为一个标签,这是由于一条患者投诉提到了分类标准中的多个问题。实验语料中多标签患者投诉只占总体实验数据的 10.91%,多标签患者投诉在 42 个分类中的分布是不均匀的,且每个类的分布数量都非常少,有的类别甚至只有一条语料。若将全部类别考虑进去,模型很难学习到有用的信息,因此构建一个多元分类器来进行多标签患者投诉的分类不可行。故本研究针对每个类别的数据都构建了一个单独的基于 LSTM 模型的二元分类器来判断语料是否属于该类,即构建了 6 个二元分类器来对每条语料进行单独判断。由于各个类别样本数量不均衡,会影响学习结果,所以在构建分类器时使用了欠采样^[26]的方法进行语料的采集,即降低负类的数量。具体方法是将所有属于该类别的单标签语料作为正向语料,并从剩下 5 个类别的单标签语料中按比例抽取总数与正类相当的语料作为负向语料,进行训练与测试。

首先,本研究使用 7013 条单标签患者投诉语料作为实验语料,每个分类器使用各自类别的欠采样语料数据进行训练。表 4 展示了 6 个分类器在各自测试集上的测试结果,测试结果使用精确度,召回率和 F 值作为评价指标。

表 4 各类别的分类器性能测试情况

类别	精确度/%	召回率/%	F 值/%
质量	83.57	86.02	84.78
安全	83.44	91.30	87.20
环境	90.83	91.44	91.13
制度	92.27	88.26	90.22
沟通	91.77	94.77	93.25
尊重	97.12	95.60	96.35

从表 4 中可以看出,除“质量”和“安全”外,其他 4 个分类器的 F 值均超过了 90%,各种评价指标基本也均在 90% 以上,说明本研究构建的分类器具有应用价值与现实意义。

然后使用本研究构建的 6 个二元分类器,对 859 条多标签语料进行分类预测。预测结果如表 5 所示。

表 5 多标签语料预测结果

	至少预测出 一个类别	预测出了 所有类别	预测完全 正确	完全没有 预测出来
数量	835	518	276	24
所占百分比/%	97.20	60.30	32.13	2.80

对于多标签语料,有 97.20% 的语料至少有一个类别被预测了出来,有 60.30% 的语料所有的类别都被预测了出来。其中 32.13% 的语料预测结果与标注结果完全一致,只有 2.80% 的语料没有类别预测出来。

但在实际应用的环境中,单标签与多标签往往混合出现,而且根据标注结果,单标签患者投诉会远多于多标签患者投诉。针对这种情况,我们进行了模拟实验,即从全部数据集中随机抽取 20%,总计 1 575 条的语料作为测试集用以测试分类器在实际情况下的应用情况。测试集数据分布如表 6 所示。

表 6 模拟实验测试集语料分布

类别	语料数量	类别	语料数量
环境	427	质量安全	6
尊重	344	质量环境	4
制度	227	质量制度	4
安全	153	环境沟通	4
沟通	144	安全制度	3
质量	129	制度沟通	3
沟通尊重	27	质量安全沟通	3
环境制度	25	质量沟通尊重	3
安全尊重	15	环境制度尊重	2
制度尊重	12	安全沟通	1
环境尊重	11	质量安全环境	1
质量尊重	9	质量环境尊重	1
质量沟通	8	安全环境制度	1
安全环境	7	安全制度尊重	1
总计			1575

使用训练好的分类器对 1 575 条模拟实验语料进行分类预测实验。实验结果如表 7 所示。

表 7 模拟实验预测结果

	至少预测 出一个类	预测出了 所有类别	预测完全 正确	完全没有 预测出来
数量	1538	1451	910	40
百分比/%	97.65	92.13	57.78	2.54

对于多分类语料,有 97.65% 的语料至少有一个类被预测了出来,有 92.13% 的语料所有的类别都被预测了出来。其中 57.78% 的语料预测结果与标注结果完全一致,只有 2.54% 的语料没有类别预测出来,说明分类器有一定的实际应用价值。

但在实际实验当中,我们发现“质量”类别的分类器的效果较差,极大地影响了整体分类器的性能与实验的结果,故我们将“质量”类别的语料与分类器排除在外,进行了第二次实验。实验结果如表 8 所示。

表 8 去除“质量”类别后的预测结果

	至少预测 出一个类	预测出了 所有类别	预测完全 正确	完全没有 预测出来
数量	1387	1324	1038	20
百分比/%	98.57	94.10	73.77	1.42

在去掉“质量”类别之后,分类器的整体性能得到了极大提升,各个指标在各自的水平上均有很大提高,尤其是“完全预测正确”提升了 16%,去掉将“质量”类别可以明显提高分类器性能与实验结果。

在实验过程的各个环节中,“质量”类别的表现都不理想。其原因可能为分到“质量”类别的语料相对于其他类别特征不够明显,语料数量也最少。其他类别均有比较明显的特征,例如“安全”类别的语料很可能提到“失误”“出错”等词语,“环境”类别的语料则会提到“吵闹”“脏”等词语,提到“态度差”很有可能是“尊重”类别的语料。而“质量”类别涵盖范围比较广泛,涉及面较多,分类效果自然不及其他类别。

5 总结与讨论

本研究引入一种科学的患者投诉分类标准,并根据实际情况对分类标准进行了细微调整。针对患者投诉多标签、单标签混合存在的问题,本文设计了基于长短期记忆模型的多个二元分类器结合的方法,实现了高效的患者投诉自动分类,为更好地理解患者投诉打下坚实基础。

本研究不足之处在于首先是语料样本量不够充足,特别是训练效果较差的“质量”类别的语料;其次是患者投诉格式内容千差万别,没有定式;最后随着患者投诉量的不断增加,现有患者投诉分类标准未必适合更大样本量的患者投诉语料。

未来研究中会进一步扩大语料样本量,尤其是“质量”类别的语料,提升分类器性能;并设计一套结构化患者投诉收集系统,使得收集到的患者投诉更加规范,易于分析。后续将在现有患者投诉分类标准的基础上根据收集到的更多语料样本继续调整,使其适合更大样本的患者投诉语料。

【参考文献】

[1] Cai K, Spangler S, Chen Y, *et al.* Leveraging Sentiment Analysis for Topic Detection [C]//Sydney: Ieee/wic/acm International Conference on Web Intelligence and Intelligent Agent Technology, 2008. Wi-Iat. IEEE, 2009: 265-271.

[2] 李欣. 昆山某医院医疗投诉与医疗纠纷现状分析[D]. 苏州: 苏州大学, 2015.

[3] Wofford MM, Wofford JL, Bothra J, *et al.* Patient complaints about physician behaviors: a qualitative study[J]. *Academic Medicine*, 2004, 79(2): 134.

[4] Reader TW, Gillespie A, Roberts J. Patient complaints in health-care systems: a systematic review and coding taxonomy[J]. *Bmj Quality & Safety*, 2014, 23(8): 678.

[5] 李凤琴, 周新. 某医院医疗投诉现状、成因及对策分析[J]. *中国农村卫生事业管理*, 2011, 31(3): 305-306.

[6] Catron TF, Guillaumondegui OD, Karrass J, *et al.* Patient Complaints and Adverse Surgical Outcomes[J]. *American Journal of Medical Quality the Official Journal of the American College of Medical Quality*, 2015, 25(1): 172.

[7] Hiivala N, Mussalorauhamaa H, Tefke HL, *et al.* An analysis of dental patient safety incidents in a patient complaint and health-care supervisory database in Finland[J]. *Acta Odontologica Scandinavica*, 2016, 74(2): 81-89.

[8] Hiivala N, Mussalorauhamaa H, Murtomaa H. Can patients detect hazardous dental practice? A patient complaint study[J]. *International Journal of Health Care Quality Assurance*, 2015, 28(3): 274.

[9] Levin CM, Hopkins J. Creating a Patient Complaint Capture and Resolution Process to Incorporate Best Practices for Patient-Centered Representation[J]. *Joint Commission Journal on Quality & Patient Safety*, 2014, 40(11): 484.

[10] Mirzoev T, Kane S. Key strategies to improve systems for managing patient complaints within health facilities—what can we learn from the existing literature? [J]. *Global Health Action*, 2018, 11(1): 1458938.

[11] Montini T, Noble AA, Stelfox HT. Content analysis of patient complaints[J]. *International Journal for Quality in Health Care Journal of the International Society for Quality in Health Care*, 2008, 20(6): 412.

[12] Harrison R, Walton M, Healy J, *et al.* Patient complaints about hospital services: applying a complaint taxonomy to analyse and respond to complaints[J]. *Int J Qual Health Care*, 2016, 28(2): w3.

[13] 唐燕, 奚益群, 李锦康, 等. 儿童专科医院门诊急诊投诉原因分析及对策[J]. *中国社会医学杂志*, 2017(6): 630-633.

[14] 吴志琴, 付艳. 门诊采血患者投诉原因分析及对策[J]. *中国护理管理*, 2012, 12(8): 78-79.

[15] 郑群, 竺红宇, 刘红, 等. 278例住院患者医疗投诉原因分析[J]. *重庆医学*, 2013(33): 4073-4074.

[16] 郭铁斌, 郭威, 秦宇辰, 等. 基于 Kappa 系数的一致性检验及其软件实现[J]. *中国卫生统计*, 2016, 33(1): 169-170.

[17] S H, J S. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.

[18] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436.

[19] Sundermeyer M, Ney H. From feedforward to recurrent LSTM neural networks for language modeling [J]. *IEEE/ACM Transactions on Audio Speech & Language Processing*, 2015, 23(3): 517-529.

[20] Palangi H, Deng L, Shen Y, *et al.* Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval [J]. *IEEE/ACM Transactions on Audio Speech & Language Processing*, 2015, 24(4): 694-707.

[21] 万圣贤, 兰艳艳, 郭嘉丰, 等. 用于文本分类的局部化双向长短时记忆 [J]. *中文信息学报*, 2017, 31(3): 62-68.

[22] Shih CH, Yan BC, Liu SH, *et al.* Investigating Siamese LSTM networks for text categorization [C]//Honolulu: Asia-Pacific Signal and Information Processing Association Summit and Conference. IEEE, 2018: 641-646.

[23] Jiang M, Liang Y, Feng X, *et al.* Text classification based on deep belief network and softmax regression [J]. *Neural Computing & Applications*, 2016(7): 1-10.

[24] 赵明, 杜会芳, 董翠翠, 等. 基于 word2vec 和 LSTM 的饮食健康文本分类研究 [J]. *农业机械学报*, 2017, 48(10): 202-208.

[25] 邓三鸿, 傅余洋子, 王昊. 基于 LSTM 模型的中文图书多标签分类研究 [J]. *现代图书情报技术*, 2017, 1(7): 52-60.

[收稿日期: 2018-04-20]

[本文编辑: 施沅坤]