

基于机器学习的文学作品英译自动评价^①



孙李丽¹, 郭琳², 张文诺¹, 文旭³

¹(商洛学院 人文学院, 商洛 726000)

²(商洛学院 电子信息与电气工程学院, 商洛 726000)

³(西南大学 外国语学院, 重庆 400715)

通讯作者: 郭琳, E-mail: guolin0303@163.com

摘要: 为了提高文学英译作品自动评价的水平, 引入基于机器学习的智能算法模型成为当前最有效的方法. 首先研究文学作品的翻译规则和特殊性, 建立基于变量特征的翻译评价指标体系; 然后利用 Python 语言平台, 英译文本经 Stanford Parser、NLTK 等工具包过滤预处理之后, 采取 VSM 向量空间模型获得特征编码和特征度, 再输入到 Random-RF、Original-RF 和 AHP-RF 算法模型中训练学习, 完成翻译质量评价与分析. 实验结果表明, 融合层次分析法、灰色关联法和随机森林算法的 AHP-RF 模型的分类效果优于其它 2 种, 同时人工译本相较于其它 4 种机器译本, 质量评分高、分类错误率小, 评价结果与实际翻译情况吻合.

关键词: 机器学习; 随机森林; AHP-RF; 自动评价

引用格式: 孙李丽, 郭琳, 张文诺, 文旭. 基于机器学习的文学作品英译自动评价. 计算机系统应用, 2021, 30(3): 196-201. <http://www.c-s-a.org.cn/1003-3254/7823.html>

Automatic Evaluation for English Translation of Literary Works Based on Machine Learning

SUN Li-Li¹, GUO Lin², ZHANG Wen-Nuo¹, WEN Xu³

¹(School of Humanities, Shangluo University, Shangluo 726000, China)

²(Electronic Information and Electrical Engineering College, Shangluo University, Shangluo 726000, China)

³(College of International Studies, Southwest University, Chongqing 400715, China)

Abstract: The intelligent algorithm model based on machine learning has become the most effective method at present to improve the automatic evaluation for the English translation of literary works. First, the translation rules and particularity of literary works are studied, and the index system of translation evaluation based on the variable features is established. Then, with the aid of the Python language platform, after the English translation is filtered and preprocessed by tools such as Stanford Parser and NLTK, the feature codes and feature degree are obtained with the Vector Space Model (VSM). Furthermore, the results are input into the Random-RF, Original-RF, and AHP-RF algorithm models for training and learning. Thus, the evaluation and analysis of translation quality are completed. The experimental results show that the AHP-RF model combining the analytic hierarchy process, the grey correlation method, and the random forest algorithm has better classification than the other two. Meanwhile, compared with the other four machine translation versions, the manual translation has a high quality score and a low classification error, and the corresponding evaluation results are consistent with the actual translation.

Key words: machine learning; random forest; AHP-RF; automatic evaluation

① 基金项目: 商洛文化暨贾平凹研究中心开放课题 (17SLWH09); 商洛学院服务地方项目 (18SKY-FWDF009); 国家自然科学基金重大项目 (15ZDB099)

Foundation item: Open Fund of Shangluo Culture and Jia Pingwa Research Centre (17SLWH09); Service Local Project of Shangluo University (18SKY-FWDF009); Major Program of National Social Science Foundation of China (15ZDB099)

收稿时间: 2020-07-09; 修改时间: 2020-08-11, 2020-08-19; 采用时间: 2020-08-21; csa 在线出版时间: 2021-03-03

机器翻译效率高、成本低,随着人工智能技术发展得到广泛应用,同时人们对翻译质量的诉求也越来越高.衡量一个翻译作品的质量主要看是否忠于源语,以及目标语的语法、语义错误和优美程度,特别是地域文化和方言气息浓厚的文学作品,译文不仅要准确描述内容情节,还要真实表达人设、历史、社会和文化环境,因此翻译评价和机器翻译类似,需要对作品解码和重新编码,通过融合和训练,合理科学评价作品翻译质量.智能翻译评价系统需要算法或者模型来完成文本变量特征的特征^[1],再通过规则构造评价权重集,最后用匹配模型实现译文的评价和分析.目前已有专家开始探索定性或定量的翻译质量评价方法.哈尔滨工业大学的赵铁军等最先提出人工评价和自动评价思路,利用句子相似度评价 EBMT 系统翻译质量^[2];其后苏州大学的周国栋等总结自动评价的 3 个方面,包括语言学检测点、字符串匹配和机器学习^[3];厦门大学的周敏康等提出用层次分析法和模糊数学模型定量评价机器译文翻译质量^[4];中国科学院的刘群等提出融合策略,多角度综合评价方法,比如在 Blend 上,对比使用 SVM 和 FFNN 两种机器算法^[5];北京交通大学的刘明童等提出利用跨语句注意力机制构建 BP 复述识别模型,提高翻译准确性^[6].关于文学作品机器翻译和评价出现较晚,相关翻译评价模型,需要结合作品语言特点改进机器评价方法,提高自动翻译评价水平.本文提出一种融合策略,充分利用层次分析法中专家打分的主观能动性和灰色关联度法中相似行为映射结果的客观科学性,以及随机森林算法中的集成学习策略和分类评价优势,建立面向文学作品翻译的评价指标体系和模型,并通过地方戏剧译本实验验证.

1 机器翻译评价方法

翻译评价以实现作品翻译的准确性和真实性为目的,对机器翻译中固有或潜在的影响因素进行分析,掌握译文发生畸变及其和原文之间相似程度,从而制定出评价指标体系和评价模型.本文研究一种基于机器学习的英译文评价方法,评价整体流程如图 1 所示.先对输入的译文样本进行数据预处理,通过建立的评价指标体系得到文本变量特征;由软件抽样把训练样本送至 3 种随机森林决策树中,通过其中的分类器进行性能评估,得到分类结果和评价错误率,同时由建立的 AHP-GRA 模型计算评价结果.

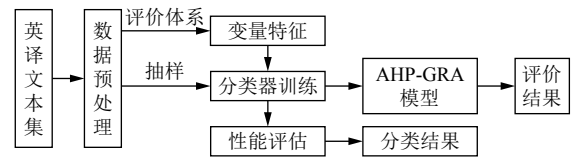


图 1 评价整体流程图

2 翻译评价指标体系

2.1 评价指标体系的建立

提出的翻译评价指标体系是建立在文学作品英译本基础上,进行定量和定性地评价作品翻译质量,此种评价方法不仅要考虑常规文本翻译方法,还要考虑文学作品的语言因素,兼顾翻译技巧的一般性和文学特殊性.陕西的乡土小说《带灯》^[7]经改编之后的商洛花鼓现代剧剧本,带有复杂交织的人物感情和差异性地域特色,里面出现大量方言和充满影射、隐晦的口语,以及原作的创造性地方特色语言,包括人物对话、俚语和言外之意等.本文以地方戏剧为例,通过对《带灯》的实地调研和翻译实践,分析了戏剧翻译的技法和特色,同时开始研究人工翻译和机器翻译的联系和区别,结合相关翻译评论和作者已有研究^[8-10],设计文学作品英译质量评价体系,如表 1 所示,其翻译质量影响包括一般翻译技法-静态变量特征群 $g[T]$ 、作者定位-动态变量特征群 $h[T]$ 、作品定位-辅助变量特征群 $p[T]$ 和译者主体性-参考变量特征群 $s[T]$ 四类一级指标,以及他们所属的 19 个二级指标.一般翻译技法分析了文学作品翻译的常规影响变量,包括语法、句型、词汇、语言、相关性和优美性;作者定位分析了设定人物的传统伦理、物质精神条件、人与自然社会的关系;作品定位分析了作品需要表达的历史、社会、人性和当地自然因素;译者主体性分析了译者对原文的喜爱与信任、理解和侵入、吸收和传播、偏差和补偿等.

2.2 文本特征度的提取

利用 Python 语言平台^[11],把机器和人工翻译的文学作品英译文输入到相关模型工具包,过滤不完整或有冗余噪音的文本数据,分析语法、语义和语境,然后采取向量空间模型 VSM 方法(把文本内容处理为向量空间中的向量运算,并以空间相似度表达语义相似度)^[12],用特征向量表征每一段文本,同时包含特征项 T_x 和特征度,特征度是影响系统评价结果的关键因素.为了提高特征度获取的精度和速度,先采用 Stanford Parser (Stanford NLP 提供的一种词性语法分析工具)^[13]和 NLTK(一种基于 Python 的自然语义处理库)^[14]进行

文本初过滤和提取处理(包括词性和句型标注、平均、局部最大值及最小值、词频加权、位置加权、句法分析等);然后利用 Doc2Vec 方法(一种随机文本获得固定长度特征的无监督算法工具)^[15,16]提取静态变

量特征,利用 K-means(一种迭代求解的聚类分析算法)^[17]提取动态变量特征,利用 LDA 模型(一种基于三层贝叶斯概率的文档主题生成模型)^[18]提取文本辅助和参考变量特征,获得文本的编码和特征度。

表 1 文学作品英译质量影响变量特征库 $W[i]$

文本库	编号	特征	特征描述	参考特征度	权重
g[T]一般翻译技法-静态变量	T_1	句型	句型设计符合原文的程度	0~100%	x_1
	T_2	语法	语法设计符合原文的程度	0~100%	x_2
	T_3	地方词汇	地方词汇翻译准确程度	0~100%	x_3
	T_4	地方语言	地方语句翻译准确程度	0~100%	x_4
	T_5	对话相关性	对话语句的相关合理性	1~5	x_5
	T_6	对话优美性	对话语句的相关优美性	1~5	x_6
h[T]作者定位-动态变量	T_7	传统伦理	人物传统思想观念的符合程度	0~100%	x_7
	T_8	人物物质条件	人物生活条件描述的符合程度	0~100%	x_8
	T_9	人物精神条件	人物精神状态描述的符合程度	0~100%	x_9
	T_{10}	人与自然关系	人物在努力改变自然环境,或被环境同化	Y/N	x_{10}
	T_{11}	人与社会关系	人物在影响、同化别人,或被影响、同化	Y/N	x_{11}
p[T]作品定位-辅助变量	T_{12}	历史定位因素	表达作品历史背景突出程度	0~100%	x_{12}
	T_{13}	社会定位因素	表达作品社会文化突出程度	0~100%	x_{13}
	T_{14}	人性定位因素	表达作品人性差别突出程度	0~100%	x_{14}
	T_{15}	自然定位因素	表达作品自然环境迥异程度	0~100%	x_{15}
s[T]译者主体性--参考变量	T_{16}	喜爱和信任	译者对原文喜爱和信任程度	0~100%	x_{16}
	T_{17}	理解和侵入	译者对原文理解与认知程度	0~100%	x_{17}
	T_{18}	吸收和传播	译文吸收原文精髓信息程度	0~100%	x_{18}
	T_{19}	偏差和补偿	译者对译文质量的诠释说明	1~5	x_{19}

3 文学作品英译评价模型

3.1 AHP-GRA 模型

AHP-GRA 模型在层次分析法的基础上融合了灰色关联度算法^[19,20],是一种无监督学习模型.层次分析法将半定性、半定量问题转化为定量问题,对专家主观打分的依赖性强,而灰色关联度算法是根据序列几何形状的相似性来确定序列重要关系,强调行为结果的客观性,AHP-GRA 模型兼顾了二者优点.通过层次分析法计算出所有指标的基础权重 x_i ,然后构建评价指标重要性判断矩阵 V ,然后确定比较集列和最优指标集,再对指标进行离散性的规范量化,然后通过式(1)计算翻译评价指标的关联系数,经过加权求和得到翻译译文的加权关联度.式(1)中 δ 为翻译分辨系数,本文取 0.5; δ_j 为翻译质量关联系数.通过式(2)得到灰色关联系数矩阵 R ,再结合前面的权重矩阵 V ,由式(3)计算得到灰色关联度值 S_j , T 为文本特征.

$$\delta_j(i) = \frac{\min_j \min_i |x_0(i) - x_j(i) + \delta \max_j \max_i |x_0(i) - x_j(i)|}{\max_j \max_i |x_0(i) - x_j(i) + \delta \max_j \max_i |x_0(i) - x_j(i)|} \quad (1)$$

$$R_{ij} = \begin{bmatrix} r_{i0} \cdots r_{1j} \\ \vdots \\ r_{i0} \cdots r_{ij} \end{bmatrix} \quad (2)$$

$$S_j(i) = \frac{\sum_{i=1}^T R_{ij}(V_i)^2}{\sqrt{\sum_{i=1}^T (V_i)^2}} \quad (3)$$

根据灰色关联度值对翻译准确度进行分级评价,当评价指标关联度值为 0.8~1 时认为译文水平优秀,为 0.6~0.8 时认为优良,为 0.4~0.6 时认为一般,为 0~0.4 时认为译文不准确.

3.2 随机森林算法

随机森林算法是一种基于集成学习 Bagging 算法建立的包含多个决策树的随机分类器,准确度高、处理能力强,适合于分类和变数评估等问题^[21].原始 Original-RF 森林算法,是一种基于 Boosting 算法的依赖串行生成序列化方法,先初始训练得到基学习器,然后调整样本训练下一个基训练器,如此重复达到基训练器预期数目,最后将所有基训练器加权结合得到分

类结果; 随机抽样 Random-RF 森林算法是基于 Bagging 算法的改进版, 产生相对独立和差异化的基训练器集合, 通过 Bootstrap 自助采样, 如图 2 所示, 引入决策树结构, 从根节点开始将数据样本根据特征进行分类, 每个类别决策树通过 Bootstrap 抽样产生一个训练集, 重复随机抽取 n 次的 N 个样本数据. 决策树数量根据所选取的译文长短确定, 随后在生长过程中以指数最小原则选出符合评价指标体系中若干特征变量的最优集合, 通过构建的 N 个决策树形成随机森林. 将测试样本集输入到随机森林, 由最大投票数的分类结果作为翻译评价的输出结果.

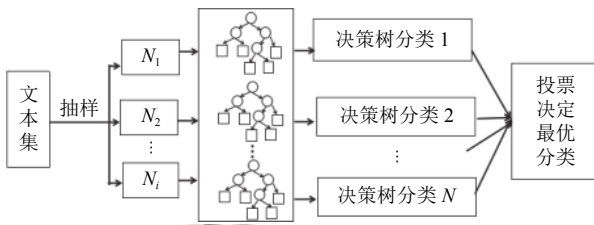


图 2 随机森林算法示意图

3.3 AHP-RF 评价模型

通过研究 AHP-GR 模型和随机森林算法的优点, 建立融合算法的文学英译作品评价模型 AHP-RF, 如图 3 所示, 由随机森林算法得到分类结果和分类错误率, 再由 AHP-GR 模型得到译文翻译质量评价. 作品英译文的训练样本集通过工具包预处理后, 得到文本特征数据 $T_1 \sim T_{19}$, 通过 Bootstrap 再从对应训练集中抽取 N 个样本构成 N 个决策树, 不剪枝完全自然生长得到随机森林分类器, 通过多数投票表决得到分类结果和分类错误率; 最后将测试样本集输入到模型 AHP-GR 中, 经过分层加权关联度计算, 得到译文翻译质量评价.

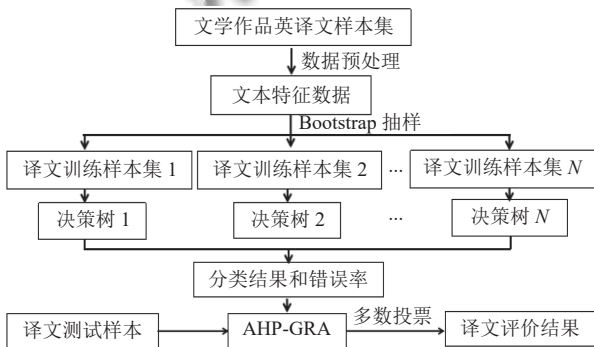


图 3 AHP-RF 作品翻译评价模型

4 实验结果分析

本文以地方戏剧《带灯》剧本为源语文本, 以作者的人工英译本和百度、谷歌、有道、搜狗 4 种在线机器英译文为评价样本集. 机器翻译方法和人工翻译类似, 是对小说解码和重新编码的过程, 并非机械化般无情, 而是中西方语言文化的一次融合和训练, 从英语译文中找到原作的词义、语义和语境. 在实验过程中, 树节点变量数目和决策树数目非常关键, 决定了系统的评价错误率. 例如图 4 示例所示, 当决策树节点变量数目为左边的 5 时, 译文语义符合原文, 模型对于文本的误判率均值不同, 需要选择误判率均值最低时的变量数目. 当错误率趋于稳定不变, 此时决策树数目设定不变.

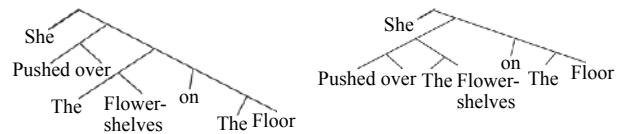


图 4 相同词序列的不同句法结构示例

通过 Bootstrap 从表 1 中的 4 个一级特征和 19 个二级特征中抽取 19 个文本特征, 带入本文的 AHP-RF 模型, 得到 19 个分类结果, 如图 5 所示.

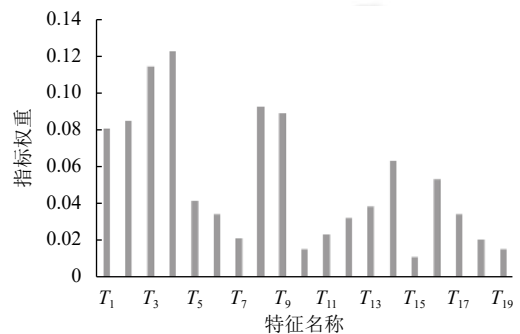


图 5 AHP-RF 二阶特征项权重

图 5 中各个特征的权重值, 其中最重要的第一层次特征项是地方语言 (0.123) 和地方词汇 (0.115), 接下来第二层次是人物物质条件 (0.093) 和人物精神条件 (0.090), 第三层次是语法 (0.085) 和句型 (0.081), 第四层次是人性因素 (0.064) 和译者喜爱和信任 (0.054), 之后的特征项权重小于平均值, 说明以上变量特征是影响实验样本翻译水平的重要指标; 其中最低层次是人与自然的关系 (0.016)、偏差和补偿 (0.016) 和自然定位因素 (0.012), 说明此类指标在翻译评价体系中的影

响最小. 然后用 AHP-GRA 算法把测试样本进行分级评价, 得到 M 组分段样本的 $19 \times M$ 个灰色关联度, 再分级评价得到离散化的数据文本. 通过实验得到机器 1~4 的译文评价价值分别为 0.53, 0.57, 0.44 和 0.62, 人工英译本的评价价值为 0.83, 其中机器 4 的译文水平达到优良, 翻译质量高于其他 3 种, 而人工译文水平达到优秀, 翻译质量明显优于机器翻译.

通过原始 Original-RF 森林算法、随机抽样 Random-RF 森林算法和本文的 AHP-RF 算法, 分别计算输出分类结果性能指标错误率 (error rate), 如图 6 至图 8 所示的不同决策树数量下译本集的各种模型算法分类错误率对比, 随着决策树数目的增加, 分类错误率明显下降, 并且当决策树数目增长到 200 附近时, 分类错误率趋于稳定, 所以针对选取的实验样本, 确定 200 是其最优的分类决策树数量.

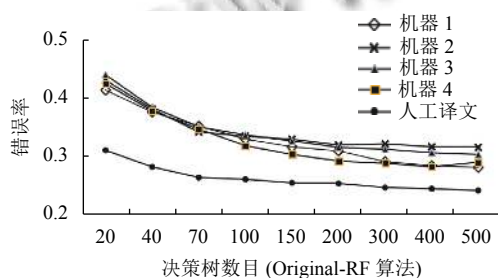


图 6 Original-RF 算法的错误率结果

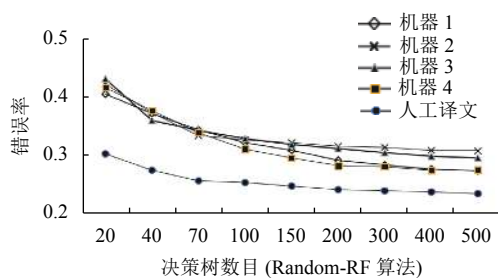


图 7 Random-RF 算法的错误率结果

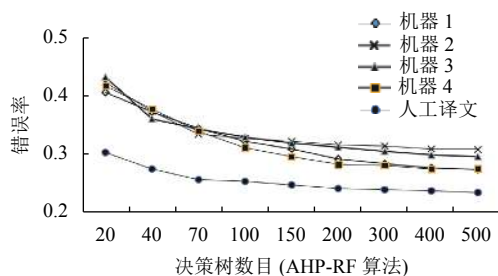


图 8 AHP-RF 算法的错误率结果

如图 9 所示, 在相同决策树数量 200 下各种译本集的 3 种随机森林算法分类错误率对比, 基于文本特征分层抽样的 AHP-RF 算法的分类错误率最低, Random-RF 算法次之, Original-RF 算法最高, 并且人工译文的错误率小于其他 4 种在线机器. 评价结果与实际翻译情况较为吻合, 说明文本提出的翻译评价方法是可行的.

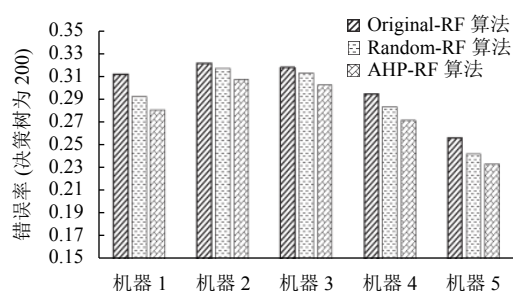


图 9 不同随机森林算法的错误率对比结果

5 结论

通过对翻译评价方法的研究, 建立了文学作品翻译的评价指标体系和 AHP-RF 评价模型, 并通过地方戏剧《带灯》的 5 种译本作为实例, 进行了翻译质量评价与分析, 实验结果表明, 融合层次分析法、灰色关联法和随机森林算法的 AHP-RF 模型, 输入样本有效完成了变量特征分类, 为翻译作品的质量评价奠定了基础. 按照权重大小排位前八的依次是地方语言、地方词汇、人物物质条件、人物精神条件、语法、句型、人性因素和译者喜爱和信任, 处于最低层的依次是人与自然的关系、偏差和补偿和自然定位因素; 人工译本的评价结果高于其它 4 种机器译本, 分类错误率小于其它 4 种机器译本, 评价结果与实际翻译情况吻合; AHP-RF 模型的分类效果优于 Random-RF 和 Original-RF, 并且当决策树数目为 200 时, 该实验样本的分类错误率下降趋于平衡状态.

参考文献

- 1 吴彦文, 黄凯, 王馨悦, 等. 一种融合主题模型的短文本情感分类方法. 小型微型计算机系统, 2019, 40(10): 2082-2086. [doi: 10.3969/j.issn.1000-1220.2019.10.010]
- 2 姚建民, 周明, 赵铁军, 等. 基于句子相似度的机器翻译评价方法及其有效性分析. 计算机研究与发展, 2004, 41(7): 1258-1265.
- 3 李良友, 贡正仙, 周国栋. 机器翻译自动评价综述. 中文信息学报, 2014, 28(3): 81-91. [doi: 10.3969/j.issn.1003-0077.

- 2014.03.011]
- 4 孙逸群, 周敏康. 机器翻译质量综合评价方法研究. 中国科技翻译, 2017, 30(2): 20–24.
 - 5 马青松, 张金超, 刘群. 基于融合策略的机器翻译自动评价方法. 中文信息学报, 2018, 32(9): 11–19. [doi: 10.3969/j.issn.1003-0077.2018.09.003]
 - 6 刘明童, 张玉洁, 徐金安, 等. 基于句法结构的神经网络复述识别模型. 北京大学学报(自然科学版), 2020, 56(1): 45–52.
 - 7 贾平凹. 带灯. 北京: 人民文学出版社, 2013.
 - 8 孙李丽. 地方戏剧中的晋语英译——以商洛花鼓戏《带灯》为例. 商洛学院学报, 2018, 32(3): 42–46, 55.
 - 9 冯丽君, 张威. 贾平凹作品“走出去”之生态译介策略研究. 小说评论, 2019, (6): 63–68.
 - 10 张白桦, 杨茹. 乔治·斯坦纳翻译四步骤视域下韩译《高兴》的译者主体性. 长春大学学报, 2020, 30(1): 50–53.
 - 11 吴立金, 简阳, 张凯, 等. 基于 Python 语言的 GUI 自动化测试脚本技术研究. 计算机测量与控制, 2015, 23(10): 3330–3332, 3337.
 - 12 肖尚, 房至一, 董洪良, 等. 基于改进型 VSM-HowNet 融合相似度算法研究. 吉林大学学报(信息科学版), 2018, 36(6): 674–680.
 - 13 刘建华, 张智雄. 基于 Stanford Parser 的实体间关系识别. 现代图书情报技术, 2009, (5): 1–5.
 - 14 李晨, 刘卫国. 基于 NLTK 的中文文本内容抽取方法. 计算机系统应用, 2019, 28(1): 275–278. [doi: 10.15888/j.cnki.csa.006700]
 - 15 李鼎宇, 胡学钢. 面向短文本的跨领域情感分类算法. 小型微型计算机系统, 2018, 39(5): 1005–1009. [doi: 10.3969/j.issn.1000-1220.2018.05.025]
 - 16 高森, 严曙, 崔超远, 等. 基于联合分类器过滤噪声的微博主题发现. 计算机系统应用, 2018, 27(1): 132–136. [doi: 10.15888/j.cnki.csa.006141]
 - 17 刘海峰, 刘守生, 张学仁. 聚类模式下一种优化的 K-means 文本特征选择. 计算机科学, 2011, 38(1): 195–197. [doi: 10.3969/j.issn.1002-137X.2011.01.045]
 - 18 吕超镇, 姬东鸿, 吴飞飞. 基于 LDA 特征扩展的短文本分类. 计算机工程与应用, 2015, 51(4): 123–127. [doi: 10.3778/j.issn.1002-8331.1403-0448]
 - 19 杜思义, 聂鹏辉. 基于改进层次分析法的地基基础优化选型. 河南科学, 2019, 37(6): 968–974. [doi: 10.3969/j.issn.1004-3918.2019.06.017]
 - 20 郭正红, 马辛华, 兰安怡. 基于层次分析法权重和灰色服务器负载预测的云计算 on-line 迁移策略. 计算机测量与控制, 2015, 23(3): 1002–1004, 1007. [doi: 10.3969/j.issn.1671-4598.2015.03.095]
 - 21 王梓杰, 周新志, 宁芊. 基于 PCA 和随机森林的故障趋势预测方法研究. 计算机测量与控制, 2018, 26(2): 21–23, 26.