

国家社科基金学科类别自动判定模型构建研究

沈思[†], 翁小颖¹, 孙豪¹, 王东波²

(1. 南京理工大学 经济管理学院, 南京 210094; 2. 南京农业大学 信息科学学院, 南京 210095)

摘要:在把所获取的国家社科基金项目标题按照词表示成训练和测试语料的基础上, 基于条件随机场模型和双向长短时记忆模型对所构建的国家社科基金项目学科类别判定模型, 进行了多个角度和层面的验证, 并与支持向量机模型的实验结果进行对比. 基于相应的模型性能评价指标, 验证了传统机器学习模型在小规模语料上的整体性能, 证明增加了人工特征模型后的条件随机场模型的整体性能并未突出, 同时对条件随机场的性能进行个案分析.

关键词:机器学习; 条件随机场模型; 国家社科基金; 文本挖掘

中图分类号:G255.1

文献标志码:A

Research on Automatic Determination Model Construction of Subject Classification of National Social Science Foundation

SHEN Si[†], WENG Xiaoying¹, SUN Hao¹, WANG Dongbo²

(1. School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China;

2. College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095, China)

Abstract: The words of National Social Science Foundation (NSSF) titles are expressed into the train and test corpus. And then, the category determination model of the NSSF project by using the conditional random field model and the bidirectional short and long time memory model is verified from many angles and levels. The results are compared with the experimental results of the support vector machine model. Based on the corresponding model performance evaluation indexes, this paper not only verifies the overall performance of the traditional machine learning model on the small-scale corpus, but also proves that the overall performance of the conditional random field model with the artificial feature model is not certain to be outstanding, meanwhile, the performance of the conditional random field model is analyzed in a case.

Key words: machine learning; conditional random field; National Social Science Foundation; text mining

国家社科基金成立二十多年以来, 其学科类别逐步得到完善, 形成了一套相对完整的类别体系. 排

除数量相对较少的艺术学和军事学这两个学科类别, 目前, 国家社科基金主要由马列·科社、党史·党

* 收稿日期: 2019-10-30

基金项目: 国家自然科学基金资助项目 (71974094), National Natural Science Foundation of China (71974094); 国家社科基金资助项目 (19FTQB015), National Social Science Foundation of China (19FTQB015); 江苏省自然科学基金资助项目 (BK20190450), Jiangsu Natural Science Foundation (BK20190450)

作者简介: 沈思(1983—), 女, 江苏南京人, 南京理工大学讲师, 博士

† 通讯联系人, E-mail: shensi@njjust.edu.cn

建、哲学、理论经济、应用经济、政治学、社会学、法学、国际问题研究、中国历史、世界历史、考古学、民族学、宗教学、中国文学、管理学、教育学、外国文学、语言学、新闻学与传播学、人口学、统计学、图书馆、情报与文献学、体育学等24个学科类别构成。上述24个学科覆盖了重点项目、一般项目、青年项目、西部项目、后期资助、成果文库和中华学术外译等不同类别的国家社科项目。但有部分国家社科基金,特别是国家社科重大项目有些是缺乏类别的知识。如何对国家社科基金,特别是国家社科基金重大项目进行分类判断,不仅对于后续的项目申请者和研究者具有直接的指导意义和价值,还可以提高项目管理者对项目管理的精准度,从而提升对国家社科项目管理的科学性和高效性。

条件随机场(Conditional Random Field, CRF)、支持向量机(Support Vector Machine, SVM)和双向长短时记忆模型(Bidirectional Long-Short Term Memory Model, Bi-LSTM)均可以实现对短文本的自动类别判定,围绕这3个模型的代表性研究如下:Lannoy等^[1]提出一种在心电信号中自动分类心跳的方法,根据该应用的特定特征提出了一个特定的分类器,该分类器是条件随机场分类器的加权变体,实验结果优于以往的心跳分类方法,尤其在病理性心跳方面;Delays等^[2]提出了一种在不受约束的在线手写文档中,区分文本和非文本笔迹的新方法,该方法基于条件随机场联合多个信息源建模,实现了在笔画级别提高分类精度,同时也突出了上下文信息不同来源的贡献;Hu等^[3]提出了一种新的高光谱图像分类处理方法,即在CRF中加入低维表示,研究了LE(Laplacian Eigenmaps)、SSSE(Spatial-Spectral Schroedinger Eigenmaps)、LLE(Local Linear Embedding)3种基于图的降维算法对随后基于CRF分类的影响,有效地解决了CRF在高光谱分类应用中需要大量计算的问题;张春元^[4]提出了一种基于条件随机场的文本分类模型,利用特征选择将文本表示成为CRFs的观察序列和状态序列,提取序列之间的关联特征,用前向或后向算法评估出各状态序列的概率,有效提高了文本分类的效率;曾佳妮^[5]对基于条件随机场的短文本分类算法进行了改进,将条件随机场理论和序列标注的方法用于短文本分类领域,并利用类别作为标注使用条件随机场模型进行标注;汪光亚^[6]提出了一种基于CRF模型的多时相遥感影像分类方法,运用最大期望算法,结合空间以及时间上下文信息构造了条件随机场模型。条件随机场作为线性序列模型的代表,不仅在分词、词性和实体等识别上

表现出了极强的性能,而且在序列的分类任务上也有较好的表现。

Ravi等^[7]提出了一种基于深度特征的SVM分类模型,针对裁判员的手势数据集进行分类,利用预训练卷积神经网络,使用线性支持向量机分类器,从而得出基于vgg19提取的特征训练的支持向量机实现网络具有最佳的分类性能;Maldonado等^[8]提出用SVM分类的嵌入式特征选择算法,来处理高维类不平衡数据集,所提出的嵌入式策略与支持向量数据描述(Support Vector Data Description, SVDD)和成本敏感型SVM(CS-SVM)一起使用,对12个高度不平衡的微阵列数据集进行实验,可实现最高平均预测性能。牛国成等^[9]通过层次分析方法和信息熵值,界定了影响变压器健康的主、客观权重,并基于支持向量机设计了判断变压器未来是否正常的算法。王峥等^[10]基于句法决策树、N-gram模型特征要素提取方法和SVM分类器,提出一种语境分类模型,解决字词在不同语境的多义性问题,有效解决文本挖掘中语境识别难题。冷强奎等^[11]提出了一种基于混合二叉树结构的多类支持向量机分类算法,该分类模型由提升分类速度的超平面和完成最终精确分类的支持向量机混合构成,既保证了分类精度,又提升了分类效率。林香亮等^[12]回顾了近30年来支持向量机的发展历史与基本理论,介绍了其改进算法,系统总结了支持向量机在分类与回归问题中的具体应用实例及优势,肯定了传统和改进的支持向量机在未来的发展潜力。Abidine等^[13]对支持向量机、条件随机场和线性判别分析模型进行比较,分别对智能家居活动进行自动识别,C-SVM能够纠正大多数的固有偏差,提高活动分类的类精度。证明向量机性能,支持向量机作为分类的代表不仅在图像、音频等数据上应用广泛,还在非结构化的文本自动分类任务上表现突出。

胡新辰^[14]提出一个基于LSTM的深度学习模型来解决语义关系分类问题。先提取文本特征,再抽取对应实体的特征组成实体类型特征,最后对这两种类型特征做特征融合并分类。该模型在标准评测集合上取得的成绩达到了目前最好水平。赵明等^[15]针对饮食领域文本分类,提出了一种基于word2vec和长短期记忆网络(Long-Short Term Memory, LSTM)的分类模型,利用word2vec构建饮食文本的文本向量作为LSTM模型的输入,训练LSTM模型进行分类,有效解决了数据表示稀疏及维度灾难问题,提高了分类准确率。近年来,深度学习在文本挖掘的各个研究任务上得到了迅速应用,在本文的社科基金项目类

别自动判定上引入深度学习模型,并与传统的机器学习模型的整体性能进行对比。

基于已有的研究,本文的整体研究框架如下。首先,确定本文所使用的类别判定的机器学习模型,并对模型的原理进行简单介绍。其次,给出了本文所使用的模型训练与测试的数据源,明确对模型判定所使用的评价指标体系。之后在所确定的数据源基础上,完成基于支持向量机、条件随机场和深度学习的类别模型训练,并对模型的整体性能进行判定和分析。最后,在分析模型整体性能的基础上,对类别判定模型所存在的错误和不足之处进行分析。

1 类别构建模型说明

从线性序列、整体文本和神经网络的角度,条件随机场模型、支持向量机模型和长短时记忆模型被用来验证国家社科基金项目的学科分类,各个模型的整体状况如下。

1.1 CRF

条件随机场^[16]在分词、词性、实体识别、句法器的开发等自然语言处理与文本挖掘的研究上表现出极强的性能。本文通过把社科项目标题学科类别的判定这一分类问题转化为序列标注的问题,为构建基于条件随机场的社科项目类别判定模型提供了条件。从而实现把分类问题转化为序列标注的问题。

若 $x = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ 代表输入被观察的国家社科项目的数据序列,如“庄子今注今译”,则 $y = \{y_1, y_2, \dots, y_{n-1}, y_n\}$ 表示有限状态的整体集合,其中每个状态对应一个国家社科基金项目待标注的类别标记。在给定输入的社科项目序列 x 的条件下,参数 $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_{n-1}, \lambda_n\}$ 的线性链 CRFs 的状态序列 y 的条件概率为:

$$P(y|x, \lambda) = \frac{1}{Z_x} \exp\left(\sum_{i=1}^n \sum_j \lambda_j f_j(y_{i-1}, y_i, x, i)\right) \quad (1)$$

$$Z_x = \sum_y \exp\left(\sum_{i=1}^n \sum_j \lambda_j f_j(y_{i-1}, y_i, x, i)\right) \quad (2)$$

式中: Z_x 为归一化因子,表示所有可能的国家社科项目文本的状态序列的得分,确保所有可能状态序列的条件概率之和为 1; $f_j(y_{i-1}, y_i, x, i)$ 为特征函数,表示二值表征函数; λ_j 是基于条件随机场模型对所训练语料中国家社科项目数据之和获得的相应特征函数的权重。

实验中,简单特征模板主要基于词这 1 列特征进行模型训练;复杂特征模板使用了词、词长、左边界词和右边界词这 4 列特征进行训练。

1.2 SVM

支持向量机^[17]的主旨思想是通过设计分割面将结构化、半结构化和非结构化数据进行两个及两个以上的分类。支持向量机具有添加特征便捷、操作简单、整体性能较为突出等特点,该模型广泛应用于图像、文本等数据媒介分类领域。

国家社科项目标题学科类别的判定,由该项目标题所组成的文本中的不同词频,作为特征输入到支持向量机模型中,结合所确定的 24 个学科类别知识,构建相应的分类模型。针对国家社科基金项目学科类别判定这一探究任务,按照该基金项目 24 个不同学科类别,对某一项目标题进行学科归属的判定,适合基于支持向量机进行构建模型,这一探究是典型的多分类任务。基于训练语料构建国家社科基金项目分类模型过程中确定的核函数为 linear,使用 word2vec 构建国家社科基金项目特征向量,实验中使用的维度数为 50,其中惩罚参数为 2.0,径向基核函数的系数 gamma 值为 0.5。

1.3 Bi-LSTM

结合社科基金项目的具体类别,长短时记忆模型公式中的 i_t, f_t, O_t, c_t 分别表示 t 时刻社科基金项目类别数据输入门(Input gate)、社科基金项目类别数据遗忘门(Forget gate)、输出门(Output gate)和社科基金项目类别数据细胞状态单元(Memory cell)的输出,有待训练的社科基金项目类别数据权重矩阵 W 、 V 和偏置向量 b 。 σ 表示用于控制社科基金项目类别数据输入和社科基金项目类别数据遗忘和记忆单元的记忆程度、遗忘程度以及社科基金项目类别数据输入门保留程度的激活函数 sigmoid。 x_t 表示项目标题的第 t 个词汇的 embedding 向量; h_{t-1} 表示 LSTM 中 $t-1$ 时刻更新门; U 表示上一个隐含状态。长短时记忆网络的训练过程为^[18]:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tanh(W_c x_t + V_c h_{t-1} + b_c) \quad (6)$$

$$h_t = O_t \otimes \tanh(c_t) \quad (7)$$

国家社科基金项目单向 LSTM 对于后续的信息不能有效利用,通过在单层国家社科基金项目 LSTM 上添加一层后向的 LSTM,实现国家社科基金项目双向长短时记忆网络(Bidirectional LSTM, Bi-LSTM)的构建,在一定程度上可以解决国家社科基金项目自动分类长距离依赖的问题,并有效地利用国家社科基金项目序列前后的特征信息。国家社科基金项目自

准确率、召回率和调和平均值。

$$\text{准确率 Precision} = \frac{\text{正确识别的数量}}{\text{正确识别的数量} + \text{错误识别的数量}} \times 100\% \quad (8)$$

$$\text{召回率 Recall} = \frac{\text{正确识别的数量}}{\text{正确识别的数量} + \text{未识别的数量}} \times 100\% \quad (9)$$

$$\text{调和平均值 } F = \frac{2PR}{P + R} \times 100\% \quad (10)$$

在所构建的国家社科基金项目学科分类模型中,只使用了上述3个指标,鉴于正确率不能准确地衡量国家社科基金项目学科分类模型的性能,为了更好地判定所构建国家社科基金项目的学科分类模型的整体性能,在条件随机场模型、双向长短记忆模型与支持向量机模型的性能判定上引入调和平均值。

3 国家社科基金项目的学科分类模型性能分析

在对国家社科基金项目进行分类训练和测试语料基础上,验证了3个模型的整体性能。在相应实验的设计过程中,通过十折交叉验证的方法,来测试所构建的3个模型的性能,将国家社科基金项目标题的数据集按照9:1拆分为训练和测试的语料。为了获得高性能的国家社科基金项目类别判定模型,设置了两组对比实验,一组是条件随机场、支持向量机和双向长短时记忆模型;另一组是基于简单特征基础的条件随机场分类模型和复杂特征基础上的条件随机场模型。

3.1 3个模型整体性能的对比如

基于国家社科基金标题的词汇,通过双向长短时记忆、支持向量机和条件随机场模型,验证了所构建的国家社科基金项目学科分类的整体性能,如图2~图4所示。

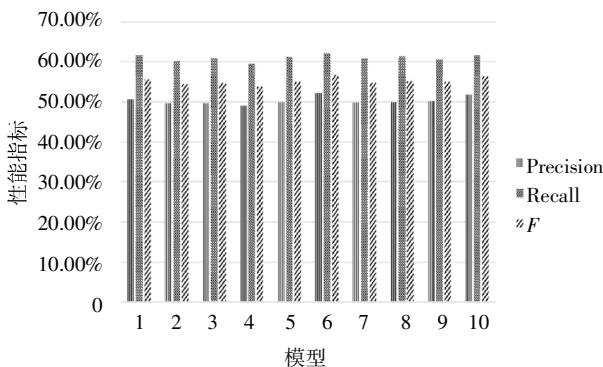


图2 基于双向长短时记忆模型的分类模型的性能图
Fig.2 Performance of the classification model based on the Bi-LSTM

由图2可知,双向长短时记忆模型的召回率相对精准率而言,整体性能较为突出,最高召回率为61.67%。调和平均值整体上低于60%,通过计算,双向长短时记忆模型的平均调和平均值为55.14%。双向长短时记忆模型的性能之所以这么低,根本原因是待分类标题整体上较短并且数量分布不充分,导致神经网络在序列化分类模型的构建上整体性能较差。

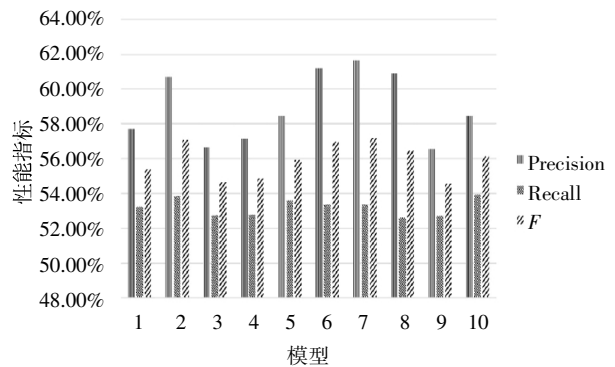


图3 基于支持向量机的分类模型的性能图
Fig.3 Performance of the classification model based on the SVM

由图3可知,相对于双向长短时记忆模型而言,基于支持向量机的分类模型精准率较高,调和平均值较低,最高精准率为61.64%。调和平均值整体上低于58%,平均调和平均值为55.92%,最低的调和平均值仅为54.58%,相较于双向长短时记忆模型,整体调和平均值有所提升,虽然在短文本和多分类的任务上该模型整体性能不佳,但该模型对于数据量的依赖性并不强。

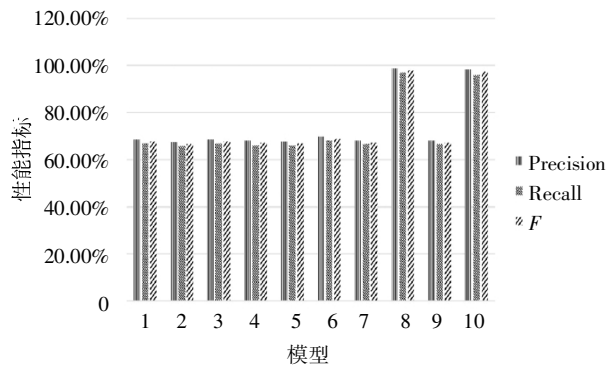


图4 基于条件随机场的分类模型的性能图
Fig.4 Performance of the classification model based on the CRF

由图4可知,相对双向长短时记忆模型和支持向量机模型,条件随机场模型整体性能较为突出,在平均调和平均值上,条件随机场模型整体达到90.70%,比双向长短时记忆模型和支持向量机模型分别高出35.36%和34.78%。通过十折交叉验证方法所获取的10个分类模型中,后8个模型的整体调和

平均值均在 96%以上, 而前 2 个模型的调和平均值刚突破 60%, 说明相对支持向量机, 通过序列化的思想对国家社科基金的标题进行分类整体性能较为突出, 相较双向长短时记忆模型而言, 条件随机场模型更适应数据量较小的短文本。

3.2 不同特征下的条件随机场模型的对比

通过调研, 在构建条件随机场模型的过程中, 相应特征会被添加到模型的构建中, 以验证不同特征对所构建模型性能的影响. 在此基础上, 本研究选择了标题的左边界词、右边界词和词长来探究模型的整体性能. 所界定的左边词主要由“中国、我国、基于、新、当代、西部、社会和现代”等词汇构成; 右边界词主要由“研究、分析、发展、建设、影响、实践、建设、对策和视角”等词汇构成. 标题词汇的长度分布情况, 对于整个类别的判定也有一定的影响, 尤其是基于序列化思想进行类别判定的情况下. 根据对所有标题词汇分布情况的统计可知, 国家社科基金项目标题词汇的长度主要是集中在 2、3、4 这 3 个长度上. 在这 3 个特征的基础上, 所构建的 10 个国家社科基金项目学科类别自动分类模型的性能如图 5 所示。

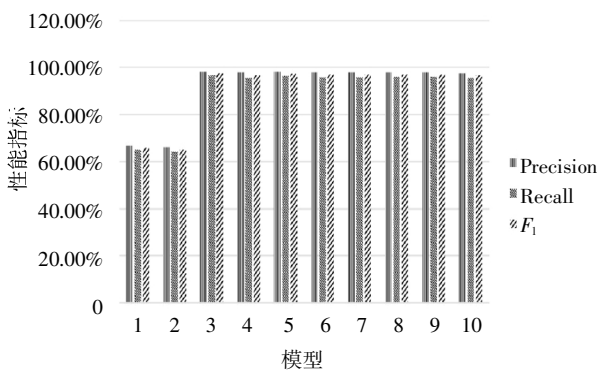


图 5 基于条件随机场的多特征分类模型的性能图
Fig.5 Performance of the multi-feature classification model based on the CRF

由图 5 可知, 相对所构建的单特征的条件随机场模型而言, 多特征分类模型的整体性能并未得到有效提高, 所构建的十折交叉模型的平均调和平均值仅为 73.50%, 其中 8 个模型的调和平均值均低于 70%, 仅有 2 个模型的调和平均值超过了 90%. 说明在规模相对较小的数据上, 通过增加特征并不能确保所构建条件随机场模型的性能得到提升, 且在均分小规模语料进行测试时, 导致模型效果出现较大的偏差。

3.3 条件随机场模型分类结果的个案分析

为了从微观角度分析基于条件随机场模型所构

建分类模型的性能, 从调和平均值达到 90% 的构建模型中任意选择一定量的测试结果, 从测试结果中任意选择 10 个识别结果进行分析, 如表 2 所示。

表 2 国家社科基金项目学科类别判定分析样例

Tab.2 Examples of subject classification determination and analysis of National Social Science Foundation

编号	国家社科基金项目名称	学科类别
1	现代性问题的马克思哲学革命	PHY
2	中国文化资源产权交易法律保障机制研究	LAW
3	农村人口转移背景下惠农政策效果的跟踪、评价与保障研究	DEY
4	新形势下提高舆论引导能力研究	JAC
5	新常态下我国商业银行流动性风险的防控及监管效果研究	AE
6	自然灾害与清代新疆社会	CH
7	反垄断法实施中的承诺制度执行难题研究	LAW
8	我国体育制度分析与设计研究	PE
9	玄言诗派研究	CL
10	中国大学核心价值体系教育模式研究	PEY

由表 2 可知, 在所选取的 10 条国家社科标题类别的判定结果样例中, 对有明显特征词汇的标题进行了正确的类别判定, 如“中国文化资源产权交易法律保障机制研究”、“现代性问题的马克思哲学革命”、“玄言诗派研究”等. 但也出现了类别判定不当的情况, 如“农村人口转移背景下惠农政策效果的跟踪、评价与保障研究”这一标题, 虽然有“人口”等与“人口学”这一学科相关的特征词汇, 但此项目明显是研究经济学的, 同样, 对于“中国大学核心价值体系教育模式研究”这一标题, 由于有“教育”这一特征词汇, 类别判定模型自动把此项目划分到教育学, 实际上, 应该归属为马列·社科这一学科。

通过分析这 10 个例子可以看出, 目前所构建的国家社科基金项目学科类别判定模型, 对于语义不是太复杂的标题可以相对精准地给出具体的学科类别, 但对于语义复杂、语义内容指向多样性的标题, 所构建的模型在自动标注过程中相对较差。

4 结论

以对未有学科类别的国家社科基金项目标题进行学科类别的自动判定为研究切入点, 基于机器学

习的系列模型,构建了系列国家社科基金项目学科类别判定模型,完成了对条件随机场模型整体性能的各种判定。

1)基于国家社科基金项目数据库,通过开发相应的网络爬虫,获取了24个学科的国家社科基金项目标题数据,并对所获取的数据进行清洗、整理与加工,为构建国家社科基金项目类别判定模型奠定了数据基础。

2)按照24个学科类别,以字为表示单位,把47352个国家社科基金项目标题的数据集分成了训练和测试语料,并对模型所使用的参数情况进行了说明。

3)通过系列实验,完成对双向长短时记忆模型、条件随机场模型和支持向量机模型的性能比较,并得出条件随机场模型在整体性能上要优于其他模型。

4)从单一特征、多特征和个案的角度对条件随机场模型在国家社科基金项目学科类别判定上进行了多个角度的验证和对比。

参考文献

- [1] LANNON G D, FRANCOIS D, DELBEKE J, *et al.* Weighted conditional random fields for supervised interpatient heartbeat classification [J]. *IEEE Transactions on Bio-Medical Engineering*, 2012, 59(1): 241—247.
- [2] DELAYE A, LIU C L. Text/non-text classification in online handwritten documents with conditional random fields [C]// *Chinese Conference on Pattern Recognition*. Berlin, Heidelberg: Springer, 2012: 514—521.
- [3] HU Y, CAHILL N D, MONTEIRO S T, *et al.* Low-dimensional representations of hyperspectral data for use in CRF-based classification [C]// *Image & Signal Processing for Remote Sensing XXI*. Toulouse, France: SPIE, 2015: 96430L.
- [4] 张春元. 基于条件随机场的文本分类模型 [J]. *计算机技术与发展*, 2011, 21(7): 77—80.
ZHANG C Y. Text categorization model based on conditional random fields [J]. *Computer Technology and Development*, 2011, 21(7): 77—80. (In Chinese)
- [5] 曾佳妮. 基于条件随机场的中文短文本分类算法研究 [D]. 上海: 上海交通大学信息安全工程学院, 2013: 1—63.
ZENG J N. Research on short text classification algorithm based on CRFs [D]. Shanghai: School of Information Security Engineering, Shanghai Jiao Tong University, 2013: 1—63. (In Chinese)
- [6] 汪光亚. 基于条件随机场的多时相遥感影像分类 [D]. 南京: 南京理工大学计算机学院, 2018: 1—59.
WANG G Y. Multi-temporal remote sensing image classification based on conditional random field [D]. Nanjing: Department of Computer Science, Nanjing University of Science and Technology, 2018: 1—59. (In Chinese)
- [7] RAVI A, VENUGOPAL H, PAUL S, *et al.* A Dataset and preliminary results for umpire pose detection using SVM classification of deep features [C]// *2018 IEEE Symposium Series on Computational Intelligence*. Bangalore, India: SSCI, 2018: 1396—1402.
- [8] MALDONADO S, LOPEZ J. Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification [J]. *Applied Soft Computing*, 2018, 67: 94—105.
- [9] 牛国成, 胡贞, 胡冬梅. 基于SVM与物元信息熵的变压器健康度分析与预测 [J]. *湖南大学学报(自然科学版)*, 2019, 133(8): 91—97.
NIU G C, HU Z, HU D M. Analysis and prediction of transformer health index based on SVM and matter element information entropy [J]. *Journal of Hunan University (Natural Sciences)*, 2019, 133(8): 91—97. (In Chinese)
- [10] 王峥, 刘师培, 彭艳兵, 等. 基于句法决策树和SVM的短文本语境识别模型 [J]. *计算机与现代化*, 2017(3): 13—17.
WANG Z, LIU S P, PENG Y B. An essay context recognition model based on syntax decision tree and SVM algorithm [J]. *Computer and Modernization*, 2017(3): 13—17. (In Chinese)
- [11] 冷强奎, 刘福德, 秦玉平. 一种基于混合二叉树结构的多类支持向量机分类算法 [J]. *计算机科学*, 2018, 45(5): 227—230, 244.
LENG Q K, LIU F D, QIN Y P. Multi-class classification algorithm for SVM based on hybrid binary tree structure [J]. *Computer Science*, 2018, 45(5): 227—230, 244. (In Chinese)
- [12] 林香亮, 袁瑞, 孙玉秋. 支持向量机的基本理论和研究进展 [J]. *长江大学学报(自科版)*, 2018, 15(17): 6, 54—59.
LIN X L, YUAN R, SUN Y Q. Basic theory and research progress of support vector machine [J]. *Journal of Yangtze University (Natural Science Edition)*, 2018, 15(17): 6, 54—59. (In Chinese)
- [13] ABIDINE M B, FERGANI B. Evaluating C-SVM, CRF and LDA classification for daily activity recognition [C]// *International Conference on Multimedia Computing & Systems*. Tangier, Morocco: ICMCS, 2012: 272—277.
- [14] 胡新辰. 基于LSTM的语义关系分类研究 [D]. 哈尔滨: 哈尔滨工业大学, 2015: 1—71.
HU X C. Research on semantic relation classification based on LSTM [D]. Harbin: School of Computer Science and Technology, Harbin Institute of Technology, 2015: 1—71. (In Chinese)
- [15] 赵明, 社会芳, 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): 207—213.
ZHAO M, DU H F, DONG C C, *et al.* Diet health text classification based on word2vec and LSTM [J]. *Transactions of the Chinese Society of Agricultural Machinery*, 2017, 48(10): 207—213. (In Chinese)
- [16] LAFFERTY J, MCCALLUM A, PRREIRA F. Conditional random fields: probabilistic models for segmenting and labeling sequence data [J]. *International Conference on Machine Learning (ICML)*, 2001(37): 282—289.
- [17] HSU C W, LIN C J. A comparison of methods for multi-class support vector machines [J]. *IEEE Transactions on Neural Networks*, 2002, 13(23): 415—425.
- [18] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers [J]. *Neural Processing Letters*, 1999, 9(3): 293—300.
- [19] 国家社科基金项目数据库 [EB/OL]. <http://fz.people.com.cn/skygb/sk/>, 2019-08-25.
Database of National Social Science Foundation [EB/OL]. <http://fz.people.com.cn/skygb/sk/>, 2019-08-25. (In Chinese)