

基于机器学习的军事装备知识分类方法*

陈 昇 谢俊杰 赵 梅 汤 杰

(中国电子科技集团公司第二十八研究所 南京 210007)

摘 要: 基于开源获取的军事百科知识提取知识中关键特征并赋予特征权重,分别以词频-逆文档频率(TF-IDF)法和词向量(Word2Vec)作为文本表征手段,采用K最近邻(KNN)、支持向量机(SVM)、神经网络及其他机器学习算法开展军事装备知识分类研究。提出了装备知识大类(装备、地点和部队等)、装备目录层级小类2级分类模式,取得了较好的分类结果;比较了各算法的优劣,有助于形成更高效、准确的军事装备知识模型,可支撑军事装备知识图谱的构建和应用。

关键词: 军事装备;机器学习;文本分类;神经网络

中图分类号: TP181 **文献标识码:** A **文章编号:** 1674-909X(2020)04-0034-06

Knowledge Classification Method for Military Equipment Based on Machine Learning

CHEN Ao XIE Junjie ZHAO Mei TANG Jie

(The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China)

Abstract: Based on the military encyclopedia knowledge acquired from open source, the key features of knowledge are extracted and the feature weights are given. Taken the term frequency-inverse document frequency (TF-IDF) algorithm and Word2Vec as the text representation means, using K-nearest neighbor (KNN), support vector machine (SVM), neural network and other machine learning algorithms, the knowledge classification research of military equipment is carried out. The two-level classification mode including the first type of the equipment knowledge and the second type of the equipment catalog level are presented, and good classification results are achieved. The advantages and disadvantages of each algorithm are compared. It's helpful to form a more efficient and more accurate military equipment knowledge model, and it can support the construction and application of the military equipment knowledge graph.

Key words: military equipment; machine learning; text classification; neural network

0 引 言

现实生活中很多实际应用问题可以转化为数据分类问题,如情感分析、疾病诊断、气象预报、商品推荐和网络检测等^[1-4]。近年来,随着信息化的蓬勃发

展及数据的爆炸式增长,如何有效提取数据,并进一步准确组织分类,是当前的研究热点,而K最近邻(KNN)、支持向量机(SVM)和神经网络等机器学习分类算法也得以持续、快速和长足发展^[5-6]。

目前,知识分类主要过程为:首先,对非结构化

* 基金项目:航天系统部装备部“十三五”预研课题资助项目。

收稿日期:2019-12-04

引用格式:陈昇,谢俊杰,赵梅,等.基于机器学习的军事装备知识分类方法[J].指挥信息系统与技术,2020,11(4):34-39.

CHEN Ao, XIE Junjie, ZHAO Mei, et al. Knowledge classification method for military equipment based on machine learning[J]. Command Information System and Technology, 2020, 11(4): 34-39.

数据进行预处理,以结构化形式表示数据,即特征表示;其次,进行特征选择,抽选出最能代表数据内容的特征项,以降低特征向量空间维数;然后,利用训练集构造并训练分类器,达到良好的训练精度和验证精度;最后,使用构造的分类器对新数据进行分类处理。

彭争等^[7]通过词频-逆文档频率(TF-IDF)法进行词频统计,基于主题权重的聚类 and KNN算法实现农业新闻文本的自动分类;Bai等^[8]在向量空间模型中,使用一个带有特定权值的情绪词典构建文本向量,然后通过训练SVM模型和朴素贝叶斯(NB)模型分别进行读者情绪预测;邵曦等^[9]提出了一种改进的SVM主动学习方法,即在选取样本时考虑选取那些距离分类超平面较近的样本,同时考虑确保样本的多样性;杜亚璞等^[10]改进了KNN的微博文本分类方法。

本文基于开源获取的军事百科知识,旨在对军事装备、军事地点和军事部队等9类军事知识进行分类识别;在此基础上,针对军事装备6级划分(14个一级目录和下设若干层级),逐层逐级识别定位出装备实体所属的装备目录层级。本文分别以TF-IDF法和词向量(Word2Vec)表征知识文本,通过KNN、SVM、线性回归(Logistic)和循环神经网络等多种机器学习算法进行计算和训练,并对比各算法的分类准确性,在军事装备知识领域取得了较好的分类效果。

1 文本分类理论

1.1 机器学习分类

1.1.1 传统分类方法

基于传统分类模型的文本分类方法主要集中在特征提取和选择上,常用的有TF-IDF、词频、文档频次、互信息、KNN、NB和SVM等。

KNN算法指若一个样本在特征空间中 k 个最相似样本大多属于某一类别,则该样本也属于该类别^[3,11-13]。该算法主要依赖周围样本,优点是文本分类效果较好、支持样本动态增加而无需重复训练,缺点是计算开销大和花费时间长。

1.1.2 神经网络方法

近年来,文本分类的研究重点转移到基于人工神经网络^[14]的分类模型。依托强大的非线性映射能力,无需特征提取选择,将词的分布式表示作为特征输入至网络中,即可自动抽取对文本分类有价值的信息。

循环神经网络(RNN)能够共享权值 W ,根据具有相关性语义的输入预测输出结果。长短时记忆网络(LSTM)是一种特殊的深度循环神经网络,可以更好地捕捉文本的长距离信息。

1.2 文本特征表达

文本相似度计算是文本分类的重点,一般先进行文本特征表达,再结合特征权重计算不同文本特征表达间距离以获取相似度。基于TF-IDF词条空间^[15]和基于Word2Vec空间模型^[11]是2种不同的文本特征表达方式。

1.2.1 TF-IDF词频表征

词频(TF)描述了某个词语 t 对某篇文档的重要性,而逆文档频率(IDF)描述了词语相对整个文档集的重要性。基于TF-IDF的权重词频计算法根据字词在文件中出现的次数判断其重要性。如果某个词语或短句在一篇文章中出现的频率高(TF高),并且在其他文章中很少出现(IDF高),则认为该词语或短句具有较好的区分能力,适用于分类。

1.2.2 Word2Vec表征

Word2Vec利用深度学习的思想,将词表征为实数值的向量。通过训练将文本内容处理简化为 K 维向量空间中的向量运算,而向量空间上的相似度用来表示文本语义上的相似度。

1.3 文本相似度计算

文本相似度计算在信息检索、数据挖掘、机器翻译和文档复制检测等领域有着广泛应用。主要有编辑距离、汉明距离、余弦相似度、杰卡德系数、TF、TF-IDF和Word2Vec等计算方法。

2 基于机器学习的军事装备知识分类

本文基于机器学习的军事装备知识分类过程为:参考军事信息资源分类法^[12]和相关军事资源分类理论实践,提出军事知识的分类需求;在区分军事装备和非装备的基础上,针对军事装备,进一步细化提出军事装备的6级目录分类需求。通过提取词条中标题等要素,文本特征表征后赋予权重组合,作为词条实体输入;通过提取目录及定义描述中名称等要素,文本特征表征后赋予权重组合,作为目录特征输入;通过网站已有分类粗标以及人工核检方式细标,提供小批量的词条样本输出;通过传统机器学习或当前深度学习模型和算法,实现从输入到输出的训练学习,采用军事装备与非装备、装备目录2级粗、细分类模式,达到良好的分类效果,最终应用于军事装备知识全集的准确分类。

2.1 需求和知识获取

借鉴已有理论和具体实践,本文将军事知识划分为9大类,包括军事装备、军事地点、军事部队、军事人物、军工企业、军事活动、军事理论、军事技术和其他实体。军事知识分类如图1所示。

针对9大类中识别出的军事装备,根据网络公开的《军事装备详细分类》^[16],军事装备共分为轻武器、火炮、弹药、装甲车辆、工程装备、军用舰船和军用飞行器等共14类,并由第1级分类逐渐细化至第2~6级小类,各分类给出简短的定义和包含的装备项。军事装备层级分类如图2所示。

- 1) 军事装备: 包括枪械、弹药、舰船、飞机、坦克等;
- 2) 军事地点: 包括地名、区域、行政区、战区等;
- 3) 军事部队: 包括海军、陆军、火箭军、特种部队、日军、德军、美军、民兵、军、师、旅等;
- 4) 军事人物: 包括战争人物、军事人物、军事家、将领、烈士、军事理论家、军事领域人物等;
- 5) 军工企业: 包括企业、工厂、研究院校等;
- 6) 军事活动: 包括战争、战役、演习、阅兵、军事事件等;
- 7) 军事理论: 包括军事书籍、军事术语、战争史等;
- 8) 军事技术: 军事科学、军事技术;
- 9) 其他实体。

图1 军事知识分类

- 第一类 轻武器
- 1.1 轻武器: 单兵或班组携行使用的武器。
- 1.1.1 枪械: 主要利用火药燃气等能量通过管件发射枪弹弹头,口径小于20毫米(0.78英寸)的身管武器。包括手枪、冲锋枪、步枪、机枪及其他特种用途枪械。
- 第二类 火炮及其他发射装置
- 第三类 弹药、地雷、水雷、炸弹、反坦克导弹及其他爆炸装置
- 第四类 坦克、装甲车辆及其他军用车辆
- 第五类 军事工程装备与设备
- 第六类 军用舰船及其专用装备与设备
- 第七类 军用航空飞行器及其专用装备与设备
- 第八类 火箭、导弹、军用卫星及其辅助设备
-
- 第十四类 其他产品

图2 军事装备层级分类

例如,针对“歼-20”实体词条,先从军事装备、军事地点和军事部队等9大类中识别为军事装备,再按照军事装备详细分类,依次识别定位到“第七类军用航空飞行器及其专用装备与设备”、“7.1 军用航空飞行器”和“7.1.2 歼击机”等层级。

百科网站知识领域具有覆盖全面,词条命名规范,内容准确、客观,扩展内容丰富及页面格式统一等特点。本文从数个较大百科网站获取军事装备领域知识词条,以获取高质量语料;通过爬虫爬取,经过初步清洗,获取百科实体35 937条,其中军事装备实体8 852条,非装备实体27 085条;基于关键词粗略标记实体3 640条,涉及装备实体906条,经人工辅助检查完善标记后用作训练和检验样本。

2.2 表征和特征选择

一个典型的词条实体,包括标题、标签、属性、属性值和详细内容等部分,词条特征抽取界面如图3所示。分别对各部分进行文本特征表征并作归一化处理,再根据各组成部分对实体概念的重要性赋予权重,最后按权重累计形成词条实体的特征表征。针对详细内容多为多段长文本内容,为提高该特征表征能力和训练学习计算速度,使用句法分析工具从详细内容中抽取实体的关键核心名词替代详细内容。

一个典型的装备目录层级包括目录名称、目录定义和描述等。使用句法分析工具从定义和描述中抽取核心名词和限定词,将目录名称、抽取的核心名

标题

歼-20

编辑

同义词

歼20一般指歼-20

标签

歼-20(英文: Chengdu J-20^[1], 代号: 威龙, 北约代号: Fire Fang, 译文: 火焰獠牙/焰齿)是中航工业成都飞机工业集团公司研制的一款具备高隐身性、高态势感知、高机动性等能力的隐形第五代制空战斗机。

歼-20采用了单座双发、主动双垂尾、DSI鼓包进气道、上反鸭翼带尖拱边条的鸭式气动布局。头部、机身呈菱形,垂直尾翼向外倾斜,起落架舱门为锯齿边设计,机身以高亮银灰色涂装(原型机为深绿墨色)。侧弹舱采用创新结构,可将导弹发射挂架预先封闭于弹仓外侧,同时配备新型的PL-15和PL-21空空导弹。^[2]

歼-20是用于接替歼-10、歼-11等机型的中国第五代重型战斗机,于1997年美国第五代战斗机F-22猛禽首飞的同年正式立项,首架技术工程验证机于2009年制造成功,并于2011年1月11日在成都黄田坝军用机场实现首飞。2016年11月1日,歼-20参加珠海航展并首次对外进行双机飞行展示^[3]。2017年3月9日,中央电视台报道第五代战斗机歼-20已正式进入空军序列^[4]。2018年2月9日,歼-20开始列装空军作战部队。同时也意味着中国是继美国之后世界上第二个走完第五代战斗机论证评价、设计、研发、原型机测试、定型生产、最终服役全部阶段的国家。2018年11月6日,歼-20战机在第12届珠海航展上以新涂装、新编队、新姿态进行飞行展示。^[5-6]

中文名称	歼-20	制造单位	中航工业成都飞机工业集团公司
英文名称	Chengdu J-20	飞机类型	单座双发隐形战斗机
前型/级	歼-10	世代划分	第五代战斗机
次型/级	歼-20B(预计2020年左右)	首飞时间	2011年1月11日
研制时间	1997年	北约代号	Fire Fang/“火牙”“苍龙”(译文)
服役时间	2017年	总设计师	杨伟
定型时间	2014年	中国代号	威龙
国属性	中华/属性值	列装时间	2018年2月9日
设计单位	中航工业成都飞机设计研究所	隶属部队	中国人民解放军空军

歼-20图册

TA说 解读词条背后的知识

科罗廖夫 军事科普作者

反制欧洲最强飞行队: 歼20和歼16玩大象漫步霸气冲天!

日前,网络上曝光了3架歼-20超音速隐形战斗机与3架歼-16重型多用途战斗机在空军某... 张景龙

V百科 往期回顾

图3 词条特征抽取界面

词和限定词等按权重累计形成目录层级的特征表征。

目录层级特征抽取示例如图 4 所示,针对手枪(目录名称),可分析获取短枪管武器(核心名词)、瞄准射击(限定词)和本能射击(限定词)等词汇集,按权重特征表征。

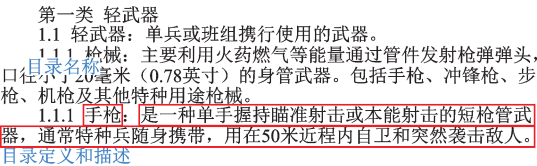


图 4 目录层级特征抽取示例

2.2.1 基于 TF-IDF 的词频向量化

针对词条实体标签、属性、属性值和详细内容等部分,经过中文分词后,统计全部分词后字典大小,计算各个具体实体包含的字典及其在所有词条实体的分布数量,分别计算各组成部分文本的 TF 值和 IDF 值,获得相应组成部分的 TF-IDF 值,并对各部分进行归一化处理,使其服从均值为 0、方差为 1 的分布。在此基础上,默认平均分配各组成部分权重,并根据该部分重要性分析及学习训练情况优化权重设置,累加获取基于 TF-IDF 的词频向量化。

本文经过多次参数试验,针对词条实体,实体名称、标签、属性、属性值和详细内容(核心名词)各部分特征权重为[0.30, 0.15, 0.20, 0.05, 0.30];针对目录,名称、定义(核心名词)和限定词各部分特征权重为[0.40, 0.40, 0.20];能够取得更优的分类效果。

2.2.2 基于 Word2Vec 词向量的文本表征

基于 TF-IDF 的词频向量化,需要对词条实体全集进行统计计算,计算开销大;对于小数据样本的语料库,对语料的完整性和比例均衡性更加敏感,受语料影响较大;而基于已预先训练的词向量库,包含各领域语料训练的词向量,语料充分丰富,各部分分类数量契合实际比例,计算处理速度更快。

本文选取人民大学等机构的研究者开源的中文词向量语料库^[13],包含百度百科、维基百科、人民日报、知乎、微博、文学、金融和古汉语等数十种各领域语料,且包含多种训练设置。选取与军事装备知识领域关联性较大的语料(300 维),如百度百科、维基百科和人民日报等;针对词条实体标签、属性、属性值和详细内容等部分,经过中文分词后在语料库中查询读取分词的词向量,各部分分别进行向量求和平均,根据权重加权求和平均求取词条实体的词向

量;对目录作同样处理。

2.2.3 基于 LSTM 的循环神经网络的矩阵表示

上述 2 种文本处理方法,丢失了输入的文本序列中分词顺序,利用 LSTM 能够很好地处理文本数据长度不一且有序的输入序列,更好地模拟文本阅读和处理的过程,具有更显著的文本理解能力。

本文基于 Word2Vec 词向量表征的基础上,尝试不再对分词向量求和平均,而是顺序组织各部分、各分词向量,构建矩阵表示。

2.3 分类算法

对于军事装备知识分类算法,主要包括装备与非装备大类分类和装备目录层级小类分类 2 个部分,主要步骤为:

1) 对已完成特征表达的词条样本,拆分为训练集和测试集,按照军事装备、军事地点和军事部队等 9 大目标分类,使用 KNN、SVM 和神经网络等分类算法,对训练样本进行训练,从而获得分类模型;使用分类模型对测试集进行验证以检验分类效果;

2) 使用分类模型对词条进行分类,获得装备以及部队与机构等非装备分类结果;

3) 针对分类为装备的词条,计算装备词条与装备目录之间的特征(语义)近似度,并根据近似度排名逐级定位所属装备目录层级;

4) 考虑到装备目录层级间的继承性,在计算当前目录层级近似度时,还需加权累加计算向下层级的特征近似度,依据该结果进行近似度排名。

2.3.1 装备与非装备分类

在完成文本特征表征后,拆分已标记样本集分别进行训练和验证。以基于 TF-IDF 词频的 KNN 算法,对军事装备、军事地点和军事部队等 9 大类知识分类为例,9 大类知识训练验证对比如表 1 所示。整体分类准确率 88.85%,其中装备分类准确率 97.35%。

表 1 9 大类知识训练验证对比

序号	类别	数据集 (训练/验证)	验证结果/% (训练/验证)
1	军事装备	815/91	97.18/98.90
2	军事地点	464/52	67.24/63.46
3	军事部队	77/9	98.70/100.00
4	军事人物	333/37	99.10/91.89
5	军工企业	473/53	99.15/94.34
6	军事活动	553/61	80.83/81.97
7	军事理论	60/7	100.00/100.00
8	军事技术	341/38	93.84/92.11
9	其他实体	158/18	67.72/72.22

2.3.2 装备目录层级分类

针对已识别的装备实体,与标准的装备目录层级进行特征相似度计算,可采用 Word2Vec 词向量表征计算和 TF-IDF 词频向量化表征计算等。

1) 计算装备实体与目录的特征相似度

由于2个实体向量表征时,向量间的距离关系也表明实体间近似程度。通过计算实体与目录间的特征相似度来定位实体与当前目录的近似关系。

2) 定位装备实体所属目录层级

由于装备目录具有继承性,在计算装备实体与当前目录相似度时,还需加权累计计算下设层级的相似度^[17-19]。

$$\text{Sim}'_{\text{cur}} = \text{Sim}_{\text{cur}} + \frac{1}{n} \sum_n \text{Sim}'_{\text{cur}-1}$$

如区分第1层级目录(轻武器和火炮等)时,需加权累加计算的与第2层级和第3层级目录的相似度,归一处理后当前层级与装备实体具有最大相似度的为装备实体第1层级分类结果。按此方式,逐步计算第2~6级语义近似度并识别最大值,逐层定位装备实体所在目录层级。

以 Word2Vec 词向量表征计算,使用百度百科语料库经训练验证后,装备目录层级的分类验证情况如表2所示,整体定位准确率在75%以上。

表2 装备目录层级的分类验证情况

序号	目录层级匹配	验证结果/%
1	第1层	92.42
2	第2~3层	82.46
3	第4~6层	67.42

2.3.3 不同算法分类效果

针对不同的文本特征表征方式,试验加载不同预训练语料库,分别选用 KNN、SVM、Logistic 和 RNN 等算法进行训练学习^[10,14-15],调节优化各算法参数设置,拆分已标记样本集,分别进行训练和验证。以装备实体与非装备实体(军事部队、军事人物、军工企业和军事活动等)分类为例,主要进行以下3个方面的试验:

1) 加载百度百科、维基百科、人民日报和文学等语料库,基于词向量的 SVM 算法,比较语料库对分类效果的影响;

2) 开展基于词向量的 KNN、SVM 和 Logistic 算法的训练验证,比较机器学习中不同算法的优劣;

3) 开展基于 TF-IDF 词频的 KNN 算法、基于词

向量的 SVM 算法及基于矩阵表示的循环神经网络模型算法的训练验证,比较 TF-IDF 词频、Word2Vec 词向量和矩阵等文本表征方式及传统算法、机器学习与深度学习算法的优劣。

为确保试验结果稳定性,每个算法模型学习训练均测算20次,训练精确度和验证精确度均取20次的平均值。针对不同方式的训练和验证结果对比如表3~表5所示。

表3 基于词向量的 SVM 算法(不同语料)训练验证对比

序号	语料库	数据量/MB	训练情况/%	验证结果/%
1	百度百科	386	91.14	89.51
2	人民日报	412	90.92	89.47
3	文学	312	87.95	86.07

表4 基于词向量的机器学习算法(KNN、SVM 和 Logistic 等)训练验证对比

序号	机器学习算法	训练情况/%	验证结果/%
1	KNN 算法	89.41	85.57
2	SVM 算法	91.14	89.51
3	Logistic 算法	92.46	88.98
4	NativeBeyas 算法	84.57	83.85
5	AdaBoost 算法	99.84	87.39

表5 基于不同表征的机器学习算法(KNN、SVM 和 Logistic 等)训练验证对比

序号	机器学习算法	训练情况/%	验证结果/%	备注
1	基于 TF-IDF 词频的 KNN 算法	100	96.41	单次 1.4 h
2	基于词向量的 SVM 算法	91.14	89.51	—
3	基于矩阵表示的 RNN 模型算法	96.64	88.11	—

2.4 试验分析

试验数据分析后可得如下结论:

1) 专业的语料库对分类效果存在一定影响,本领域在选择百度百科、维基百科、人民日报、微博和文学等方向时,百度百科优于文学和微博等其他语料;

2) 小语料库的 TF-IDF 算法在语料集中且领域特性强时,其精确度相对宽域的大语料库更加精确,针对性更强;但 TF-IDF 算法需加载全部语料库并计算向量表征,而预训练语料库已完成该步骤,故 TF-IDF 算法性能一般,耗时较长;

3) 深度学习神经网络在图像识别方面有较强效果,但在文本分类领域与一般的机器学习算法相比并无显著优势。越复杂的模型需要越多的训练数据,而对于训练数据缺乏(尚未达到大数据要求)的情况,简单的模型会得到更好的性能;

4) 选取更多的特征,调整完善特征权重,能够有效提升分类效果,但换来的代价是计算时间更长;

5) 依托目录层级的继承性特点,对实体与当前和下设目录的相似度计算,能够更准确定位和识别装备目录层级。

2.5 实践应用

从百科知识中抽取军事装备知识的实体、属性和关系,并依据分类算法对实体分类分层分级,采用图数据库建立军事装备知识图谱,能够实现装备实体与非装备实体的自动检测和分类,以及装备目录层级的自动定位,可实现基于装备知识图谱的语义检索和智能问答,有利于形成更高效明确的军事装备知识体系。

3 结束语

本文围绕军事装备知识分类研究,从文本分类的特征选取、向量表达和分类算法等方面进行理论分析,并根据获取的军事装备知识样本,结合试验结果分析,重点从TF-IDF、词向量和矩阵3种文本表征方式,比较了KNN、SVM、Logistic和神经网络等多种机器学习和深度学习算法的优劣,实践了军事装备领域知识分类算法,提出了装备与非装备、装备目录的2级粗细结合的分类模式,取得了较好的分类效果。后续将进一步扩大语料来源和语料库规模,完善优化词汇抽取脚本,调整算法模型结构和参数配置,以获取更加优良、准确的军事装备知识领域分类效果,为开展军事领域知识图谱构建提供良好的数据来源和分类结果。

参考文献(References):

- [1] 冯成刚,田大纲. 基于机器学习的微博情感分类研究[J]. 软件导刊,2018,17(6):58-61.
- [2] 王正成,李丹丹. 基于词向量和情感本体的短文本情感分类[J]. 浙江理工大学学报(社会科学版),2018,40(1):33-38.
- [3] 严倩,王礼敏,李寿山,等. 结合新闻和评论文本的读者情绪分类方法[J]. 山东大学学报(理学版),2018,53(9):35-39.
- [4] 刘欢欢. 面向新闻的读者情绪自动分析研究方法[D]. 苏州:苏州大学,2015.
- [5] 马建斌,李滢,滕桂法,等. KNN和SVM算法在中文文本自动分类技术上的比较研究[J]. 河北农业大学学报,2008,31(3):120-123.
- [6] 付娟妮. 基于信息用户的新闻推荐系统特点及构建[J]. 企业科技与发展,2013(15):39-40.
- [7] 彭争,唐东明. 基于文本分类的农业种植信息集成推荐方法研究[J]. 西南民族大学学报(自然科学版),2018,44(1):69-74.
- [8] BAI S T, NING Y, YUAN S, et al. Predicting reader's emotion on Chinese Web news articles [C]//Proceedings of Joint International Conference on Pervasive Computing and the Networked World. Istanbul: Springer, 2012:16-27.
- [9] 邵曦,姚磊. 基于SVM主动学习的音乐分类[J]. 计算机工程与应用,2016,52(6):127-132.
- [10] 杜亚璞. 一种基于改进ML-KNN的微博文本分类方法[J]. 信息与电脑(理论版),2018(7):42-44.
- [11] 张敬谊,张亚红,李静. 基于词向量特征的文本分类模型研究[J]. 信息技术与标准化,2017(5):71-75.
- [12] 傅亮. 基于《军事信息资源分类法》标引经验的自动分类模式构想[J]. 现代图书情报技术,2007(11):76-79.
- [13] LI S, ZHAO Z, HU R F, et al. Analogical reasoning on Chinese morphological and semantic relations [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Volume 2: Short Papers. Melbourne: ResearchGate, 2018: 138-143.
- [14] 刘婷婷,朱文东,刘广一. 基于深度学习的文本分类研究进展[J]. 电力信息与通信技术,2018,16(3):1-7.
- [15] 甘秋云. 基于TF-IDF向量空间模型文本相似度算法的分析[J]. 池州学院学报,2018,32(3):41-43.
- [16] 武器装备详细分类[EB/OL]. (2018-06-27)[2019-12-03]. <https://wenku.baidu.com/view/3b03b1dd6f1aff00bed51e12.html>.
- [17] 戴璞微,潘斌,王玉铭,等. 一种基于层次分析法的改进KNN算法[J]. 辽宁石油化工大学学报,2018,38(4):87-92.
- [18] 肖辉辉,段艳明. 基于属性值相关距离的KNN算法的改进研究[J]. 计算机科学,2013,40(Z2):157-159.
- [19] 严晓明. 基于类别平均距离的加权分类KNN算法[J]. 计算机系统应用,2014,23(2):128-132.

作者简介:

陈 昇,女(1984—),高级工程师,研究方向为指挥信息系统软件总体和装备总体论证。

谢俊杰,男(1984—),工程师,研究方向为装备总体论证与装备管理系统开发。

赵 梅,女(1987—),工程师,研究方向为装备总体论证与装备大数据研究。

汤 杰,男(1990—),工程师,研究方向为装备总体论证与J2EE软件开发。

(本文编辑:李素华)