doi: 10.7690/bgzdh.2025.06.007

一种基于改进加权 LDA 模型的敏感词识别模型

曾　玲，林天余，何秋霞，陈　莹，胡娟娟
（中国南方电网有限责任公司海南电网有限责任公司，海口 570203）

摘要：针对目前互联网中主题识别时存在数据复杂、预测精度低的缺陷，提出一种基于改进加权潜在狄利克雷分配（latent Dirichlet allocation，LDA）模型的敏感词识别模型。建立特定领域敏感词语料库；为提高敏感信息主题的识别效率，对语料库进行粗粒度文本分类；通过加权模型，提高共现频率低但敏感特征明显的词的分布权重，从而可以发现更多具有低频隐式关系的词；以主流新闻网站爬取的数据为例，对所提模型进行验证。结果表明：该模型可识别和提取每个类别的文本更详细的敏感信息主题，该模型有效且准确。

关键词：主题识别；敏感词；自然语言处理；潜在狄利克雷分配
中图分类号：TP391.43　　**文献标志码：**A

Sensitive Word Recognition Model Based on Improved Weighted LDA Model

Zeng Ling, Lin Tianyu, He Qiuxia, Chen Ying, Hu Juanjuan
(*Hainan Power Grid Co*., *Ltd.*, *China Southern Power Grid Co*., *Ltd.*, *Haikou* 570203, *China*)

Abstract: In view of the defects of complex data and low prediction accuracy in the current Internet topic recognition, this paper proposes a sensitive word recognition model based on an improved weighted latent Dirichlet allocation (LDA) model. A corpus of sensitive words in a specific field is established; in order to improve the identification efficiency of sensitive information topics, a coarse-grained text classification is proposed for the corpus; a weighting model is proposed, and more words with low-frequency implicit relations can be found by increasing the distribution weight of words with low co-occurrence frequency but obvious sensitive characteristics; Taking the data crawled by mainstream news websites as an example, the proposed model is verified. The results show that the proposed model can identify and extract more detailed sensitive information topics from each text category, The simulation results further verify the effectiveness and accuracy of the proposed model.

Keywords: topic identification; sensitive words; natural language processing; latent Dirichlet distribution

0　引言

因为网络具有即时、隐蔽、虚拟等特征，越来越多的用户借助互联网传播思想、交流情感以及表达观点[1-2]。而且，越来越多的网络用户日益频繁地参与到最近发生的事件中，人们的社会责任意识和维权意识也逐渐增强。网络具有显著的便利性以及开放性，互联网用户在网络发布越来越多的信息，表达自己的观点及情绪，而这些情绪往往与一些敏感信息混杂在一起。此外，许多网络新闻发布平台允许用户对发布的新闻发表评论，可能不少评论涵盖了诸多敏感信息，这些信息都极大地表明了网络用户对各种特定事件的观点与立场[3-5]。敏感信息在网络舆论的传播以及形成过程中起到了非常关键的作用，也会在一定程度上对社会安全形成潜在的威胁。

基于上述分析，主题识别[6]将离散和孤立的新闻报道整合起来，并提取主题的关键词，如色情、暴力和其他敏感信息，从而把握舆论的发展趋势。这对非常规突发事件的应急管理具有重要意义。目前，许多研究对主题及敏感词识别进行了研究与分析，并提出了许多独到的见解。文献[7]提出了一种基于技术特征相似性的专利数据中主题识别模型，融合技术特征向量模型与聚类算法，可有效识别多维度下新兴技术主题；然而，互联网中数据样式复杂，与专利数据有很大区别，同时聚类算法分析能力有限，该模型无法直接引入互联网敏感词识别。随着深度学习技术的发展，许多学者将深度学习技术引入自然语言处理（natural language processing，NLP）领域。文献[8]提出了一种基于情感计算与深度学习的弹幕文本敏感词识别方法。文献[9]提出了一种基于 BERT 模型方法和语义分析方法相结合衡量新闻的敏感程度的模型，进而评估新闻的风险水平。深度学习模型可有效提取文本中词与词之间的

收稿日期：2024-08-10；**修回日期：**2024-09-17
第一作者：曾　玲（1989—），女，海南人。

关键特征；然而，监督深度学习模型在训练样本时需要大量标签，这将消耗大量人力成本。LDA 是一种概率生成模型，其可以有效提取文本的隐含主题。文献[10]提出了基于种子约束 LDA（隐含 Dirichlet 分布）的产品属性提取方法。考虑到 LDA 模型得到的主题特征分布倾向于高频词，这将导致大多数可以代表文本的词被一些高频单词淹没；然而，敏感信息词经常以较低的频率出现在文本中。

笔者通过构建特定领域的敏感词词汇，将敏感词嵌入主题模型中，从而提高主题词的质量。

1　模型设计

基于改进加权 LDA 模型的敏感词识别模型可分为 2 个步骤：根据领域对语料库进行粗粒度分类；利用所提出的敏感词加权 LDA 模型来识别和提取每个类别的文本更详细的敏感信息主题。

1.1　粗粒度分类

为提高敏感信息主题的识别效率，需要对语料库进行粗粒度文本分类。先获取文本的矢量表示。

一般情况下，只需几个关键字就可以表征文本描述。所以，文本矢量代表着能够转化为以关键词为基础的向量，将不同维度中最显著的特征提取出来，即可取关键字向量的每个维度的最大值来表示文本 $\boldsymbol{d}$：

$$\left.\begin{aligned}\boldsymbol{d}&=\max(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_L)\\ \boldsymbol{v}_i&=\{w_{i1},w_{i2},\cdots,w_{iR}\}\end{aligned}\right\}。\tag{1}$$

式中：max(•)为最大值函数；$\boldsymbol{v}_i$ 为文本 $\boldsymbol{d}$ 的第 i 个关键字的矢量表示；R 为词组嵌入的维度；w_{ij}($j\in\{1, 2, \cdots, R\}$)为 $\boldsymbol{v}_i$ 的第 j 个元素；L 为关键字的数量。笔者将文本 $\boldsymbol{d}$ 的矢量表示通过最大池化运算获得，即 $\boldsymbol{d}$ 的第 j 个分量代表着集合$\{w_{i1}, w_{i2}, \cdots, w_{iR}\}$中最大的元素。以最大池化操作为基础，能够将若干非核心的词组过滤掉，仅保留最核心的关键字。

当获取文本的向量表示后，向全部文本的向量表示实施相应的平均池化操作，以得到文本类别 C 领域的中心向量：

$$\boldsymbol{y}_C=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{d}_i。\tag{2}$$

式中：$\boldsymbol{y}_C$ 为类别的中心向量；N 为类别 C 中包含的文本数量。

同理，对于任意未分类文本，可通过计算类别 $\boldsymbol{C}_j$ 和未分类文本 $\boldsymbol{d}_i$ 之间的相似度，并将具有最大相似度的类指定给未分类文本。相似度计算公式如下：

$$\mathrm{sim}(\boldsymbol{d}_i,\boldsymbol{y}_i)=\frac{\sum_{s=1}^{R}(w_{is},w_{js})}{\sqrt{\sum_{s=1}^{R}(w_{is})^2}\sqrt{\sum_{s=1}^{R}(w_{js})^2}}。\tag{3}$$

式中：$\boldsymbol{d}_i$ 为待分类的文本向量；$\boldsymbol{y}_i$ 为文本类别 $\boldsymbol{C}_j$ 的中心向量。

1.2　敏感词加权 LDA 模型

LDA 模型是一种典型的概率生成模型。根据 LDA，文档可理解为一个词包，即文档是一组忽略了语法或词汇的顺序关系构成的多个词。此外对 LDA 模型而言，即便增加训练文档的数量，其参数空间也不再跟着增加。就 LDA 模型而言，其本质是一个 3 层分层贝叶斯模型，包括文档层、主题层和词语层，可以在潜在的低维语义空间映射出高维文本集。无论是文档映射到主题的分布，还是主题映射到单词的分布，都符合多项式分布，而且隐含主题还把其当作词特征方面的软聚类。LDA 模型在更抽象的层次上使文本信息形成了泛化能力。图 1 为 LDA 模型的层次结构。

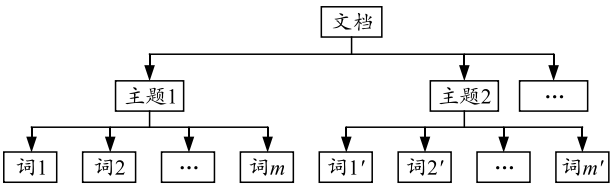

图 1　LDA 模型的层次结构

以下是主题生成的过程：

$$\left.\begin{aligned}&p(\theta_m)=\mathrm{D_{ir}}(\theta_m|\alpha),\ \ m=1,2,\cdots,M\\ &p(z_n^m|\theta_m)=M_{\mathrm{ult}}(z_n^m|\theta_m),\ \ n=1,2,\cdots,N_m\\ &p(w_n^m|z_n^m,\beta)=M_{\mathrm{ult}}(w_n^m|\beta_{z_n^m}),\ \ n=1,2,\cdots,N_m\end{aligned}\right\}。\tag{4}$$

式中：$p(\bullet)$为概率分布；$D_{ir}(\bullet|\bullet)$为狄利克雷分布函数；$M_{\mathrm{ult}}(\bullet|\bullet)$为多项式分布函数；$M$ 为文本总数；N_m 为第 m 个文档里面的总字数；α、β 分别为各文档下主题多重分布的狄利克雷先验参数以及各主题中词的多项式分布的狄利克雷先验参数；θ_m、w_n^m、z_n^m 分别为第 m 个文件下的主题分布、第 m 个文档中的第 n 个词以及第 m 个文档中第 n 个词的主题。

改进的敏感词加权 LDA 模型结构如图 2 所示。

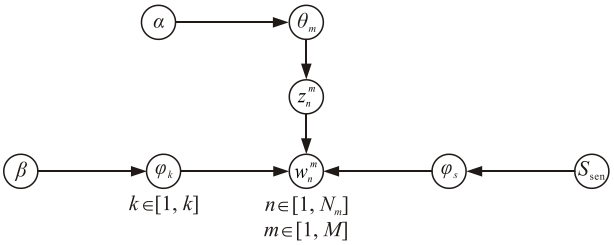

图 2　改进的敏感词加权 LDA 模型结构

图 2 中：K 为主题的总数量；θ_m 为第 m 个文档下的主题分布；φ_s 为敏感主题词分布；s_{cen} 为敏感词集。需注意，敏感主词分布可以有效改善基础 LDA 模型在识别词共现低频关系方面的不足，提高共现频率低但敏感特征明显的词的分布权重，从而可以发现更多具有低频隐式关系的词。

敏感词加权 LDA 模型以标准 LDA 模型为基础，并将约束变量 δ 加入其中。模型参数的相关计算涵盖主题词分布 φ 以及文档主题分布 θ。笔者侧重采取以蒙特卡洛马尔可夫链方法为基础的 Gibbis 抽样法推导模型：

$$p(z_i=k|\ \vec{z}_{\neg i},\vec{w},\alpha,\beta,\delta)=\frac{p(\vec{w},\vec{z},\alpha,\beta,\delta)}{p(\vec{w},\vec{z}_{\neg i},\alpha,\beta,\delta)}=\frac{p(\vec{w},\vec{z},\alpha,\beta,\delta)}{p(w_i=t,\vec{w}_{\neg i},\vec{z}_{\neg i},\alpha,\beta,\delta)}=\frac{p(\vec{w},\vec{z},\alpha,\beta,\delta)}{p(\vec{w}_{\neg i},\vec{z}_{\neg i}|w_i=t,\alpha,\beta,\delta)\cdot p(w_i=t)}\propto\frac{p(\vec{w},\vec{z},\alpha,\beta,\delta)}{p(\vec{w}_{\neg i},\vec{z}_{\neg i},\alpha,\beta,\delta)}。\tag{5}$$

式中：$\vec{z}_{\neg i}$ 为除第 i 个词组的主题；$\vec{w}_{\neg i}$ 为除第 i 个词的其他词组；t 为执行时间步长。同时，根据蒙特卡洛过程，有：

$$\frac{p(\vec{w},\vec{z},\alpha,\beta,\delta)}{p(\vec{w}_{\neg i},\vec{z}_{\neg i},\alpha,\beta,\delta)}=\frac{p(\vec{w},\vec{z}|\alpha,\beta,\delta)}{p(\vec{w}_{\neg i},\vec{z}_{\neg i}|\alpha,\beta,\delta)}。\tag{6}$$

通过式(5)和(6)可推导出：

$$p(z_i=k|\ \vec{z}_{\neg i},\vec{w},\alpha,\beta,\delta)\propto\frac{p(\vec{w},\vec{z}|\alpha,\beta,\delta)}{p(\vec{w}_{\neg i},\vec{z}_{\neg i}|\alpha,\beta,\delta)}=p(z_i=k,\vec{w}|\alpha,\beta,\delta)。\tag{7}$$

进一步，式(7)可扩展为：

$$p(z_i=k,w_i|\alpha,\beta,\delta)=p(z_i=k|\alpha,\beta,\delta)p(w_i|z_i=k,\alpha,\beta,\delta)=p(z_i=k|\alpha)p(w_i|z_i=k,\beta)p(w_i|z_i=k,\delta)。\tag{8}$$

根据式(7)和(8)，可推导出下式：

$$p(z_i=k|\vec{z}_{\neg i},\vec{w},\alpha,\beta,\delta)\propto p(w_i|z_i,\delta)\frac{n_{k,\neg i}^{(t)}+\beta}{\sum_{t=1}^{V}n_{k,\neg i}^{(t)}+V_\beta}\cdot\frac{n_{\text{d},\neg i}^{(t)}+\alpha}{\sum_{k=1}^{K}n_{\text{d},\neg i}^{(k)}+K_\alpha}。\tag{9}$$

式中：$n_{k,\neg i}^{(t)}$ 为词 w_i 属于主题 k 的次数(除采样 i 外)；$\sum_{t=1}^{V}n_{k,\neg i}^{(t)}$ 为除采样 i 外，所有词属于主题 k 的次数；V 为语料库中的词的总数；$n_{\text{d},\neg i}^{(t)}$ 为除采样 i 外，文档 d 中词 w_i 属于主题 k 的次数；$\sum_{k=1}^{K}n_{\text{d},\neg i}^{(k)}$ 为文档 d 中除采样 i 外，属于所有主题的次数。

根据以上公式可计算分布参数 θ 和 φ：

$$\theta_{\text{d},k}=(n_{\text{d},k}+\alpha)\Big/\Big(\sum_{k=1}^{K}n_{\text{d},k}+\alpha\Big);\tag{10}$$

$$\varphi_{k,w}=p\big(w_i|z_i=k,\delta\big)(n_{k,t}+\beta)\Big/\Big(\sum_{t=1}^{V}n_{k,t}+\beta\Big)=\big(1+f^k(w)\big)\big(n_{k,t}+\beta\big)\Big/\Big(\sum_{t=1}^{V}n_{k,t}+\beta\Big)。\tag{11}$$

式中：$f^k(w)$为词 w 分配给主题 k 的加权值，该值计算如下：

$$f^k(w)=\mu n_{ck}。\tag{12}$$

式中：n_{ck} 为主题 k 中出现敏感词类别 c 的数量；μ 为调节因子。

2 实验与分析

实验主要包括 3 部分：介绍数据集及对数据的采集、预处理操作；介绍词嵌入训练、敏感词扩展；对比所提敏感信息加权 LDA 模型的敏感信息主题识别性能。

2.1 数据集

实验构建的语料库主要基于 Web 爬虫收集的中国主流新闻网站，共有 461 067 个新闻网页被抓取，其中 120 292 条新闻包含评论，评论数量为 11 674 233。评论中一般会涵盖不少敏感信息，此类敏感信息通常都会表明网民对具体事件的观点与立场。收集完数据后，对数据实施预处理并进行数据清理，从而建立了一定规模的半结构化新闻语料库。语料库包含基础语料库、敏感词语料库、符号语料库 3 部分。

基础语料库主要来自基于 NLPIR 系统生成的开源词典，共包含 298 032 个词。NLPIR 系统可以实现中文分词、词性标注、命名实体识别、新词识别和关键词提取。然而，基础语料库涉及范围太过广泛，会在一定程度上削弱一些需要关注的敏感词汇的信息。

同时，笔者又添加了一个敏感词语料库。敏感词语料库通过收集和过滤来自网络的词汇来构建基本敏感词汇，涵盖热点新闻、宗教、自然灾害、政治、社保、社会生计、教育、环境保护。特殊语料库共包含有 6 845 个基本敏感词。表 1 所示为敏感词语料库中基本敏感词汇统计信息。

此外，还构建了一个包含 2 792 个词和字符的符号语料库，包括一些常用的中文和英文单词停止词(符)。

表 1　基本敏感词汇统计信息

类别	数量	类别	数量
宗教	454	社会生计	1 867
政治	2 170	社会保障	71
热点新闻	1 022	教育	258
自然灾害	274	环境保护	636

2.2　词嵌入训练与敏感词扩展

当构建完语料库后，利用谷歌的 Word2vec 模型进行词嵌入训执行练。Word2vec 的主要参数设置如下：词嵌入维度为 400、上下文窗口大小为 5、词频率最小阈值为 5、负样本数为 5、迭代次数为 10 000。同时，笔者结合跳字模型(skip-gram)和负采样模型对词嵌入进行处理。最终，将基本敏感词扩展并删除重复后，扩展后敏感词汇的数量变化如表 2 所示。

表 2　扩展后敏感词汇的数量变化

类别	扩展前	扩展后	类别	扩展前	扩展后
宗教	454	1 129	社会生计	1 867	2 783
政治	2 170	3 186	社会保障	71	491
热点新闻	1 022	1 768	教育	258	1 107
自然灾害	274	831	环境保护	636	899

2.3　敏感信息主题识别

在收集到的 461 067 篇新闻文本中，426 306 篇新闻文本可以分类为：金融、新闻、体育、政治、社会、科技、娱乐、房地产、地区(香港、澳门和台湾)以及教育 10 个类别。对剩余的待分类新闻以及评论，基于本文中提出的粗粒度分类方法来识别类别。表 3 为待分类文本的最终分类结果。

表 3　待分类文本的最终分类结果

类别	已分类	待分类(含评论)	类别	已分类	待分类(含评论)
金融	92 775	17 270	娱乐	23 499	5 484
新闻	91 654	27 160	房地产	12 413	1 833
体育	82 724	49 952	地区	11 535	0
政治	51 600	605	教育	10 247	3 666
社会	49 856	724	总数	426 303	106 694

对于主题识别模型，参数设置如下：α=50/K(K 为主题的数量)，β=0.01，迭代次数为 100。n_{top} 的值为 15，即各类主题的概率值前 15 的那些词汇当作主题词。笔者借助困惑度指标对主题模型进行衡量。困惑度本身系信息理论的测量手段之一，一般能够对概率模型的泛化能力和优缺点进行评估。可以这样定义概率模型的困惑度：以概率模型为基础的熵的能量。困惑度本身越小，意味着概率模型本身的泛化能力反而越强，代表着生成主题的性能越高。困惑度的计算公式如下：

$$\mathrm{p_{el}} = 2^{-\sum_{i=1}^{N}\frac{1}{N}\log_2 q(x_i)} \quad 。 \tag{13}$$

式中：$x_1, x_2, \cdots, x_N$ 为样本观测值；N 为样本总数；$q(\bullet)$为估计概率模型；$\mathrm{p_{el}}$ 为困惑度。

为验证所提模型性能，将所提模型与基础 LDA 模型进行比较，表 4 和图 3 所提为不同模型困惑度比较结果统计和不同模型困惑度对比图(主题数为 30)。

表 4　不同模型困惑度比较结果统计

类别	LDA	所提模型	类别	LDA	所提模型
金融	1 262.32	1 096.39	科技	1 308.23	1 198.45
新闻	1 288.02	1 006.58	娱乐	1 690.52	1 671.33
体育	1 325.57	1 155.95	房地产	1 353.33	1 304.05
政治	1 469.29	1 300.66	地区	1 394.81	1 213.57
社会	1 352.10	1 212.26	教育	1 178.42	1 158.39

从表 4 中可以看出：每个类别中，所提模型的困惑度明显小于基础 LDA 模型。这表明所提模型在敏感信息主题挖掘方面的性能优于 LDA。

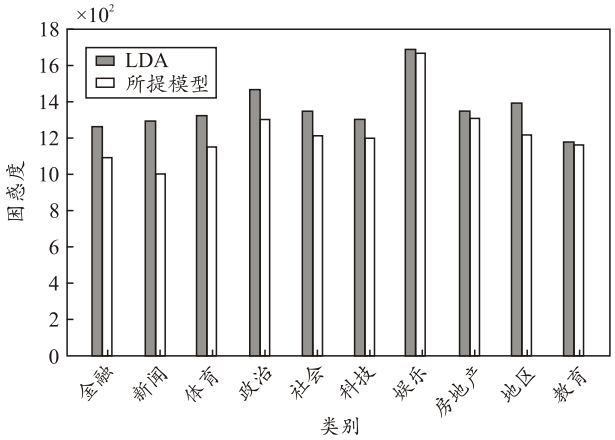

图 3　不同模型困惑度对比

3　结论

笔者对自然语言处理中敏感词识别进行了分析，建立了一种基于改进加权 LDA 模型的敏感词识别模型。根据领域对语料库进行粗粒度分类；利用所提敏感词加权 LDA 模型来识别和提取每个类别的文本更详细的敏感信息主题。结果表明，该模型为敏感词识别的发展提供了一定借鉴作用。

参考文献：

[1] 王烁. 移动互联网时代主流媒体短视频舆论引导进路浅析[J]. 教育传媒研究, 2022(6): 85-86.

[2] 张嘉琛, 杨一平. 基于元胞自动机的互联网舆论调控模型研究[J]. 信息系统工程, 2021(3): 149-150.

[3] 岳晗. 网络新闻评论融媒体传播的创新思考[J]. 新闻研究导刊, 2022, 13(16): 129-131.

[4] 马翔, 沈曙, 丁旭东. “数智传媒”融媒云的网络安全规划[J]. 网络安全技术与应用, 2022(5): 80-81.
[5] 卢腾, 李静, 魏家辉, 等. 基于语义分析的保密安全敏感事件可视化研究与应用[J]. 网络安全技术与应用, 2020(11): 43-44.
[6] 武帅, 施奕, 杨秀璋, 等. 基于社交网络分析和LDA主题挖掘的短文本挖掘研究[J]. 现代电子技术, 2022, 45(20): 124-128.
[7] 宋博文, 栾春娟, 梁丹妮. 机器学习视域下新兴技术主题识别研究——基于技术特征相似性[J]. 现代情报, 2022, 42(9): 49-57.
[8] 叶海燕. 基于情感计算与深度学习的弹幕文本敏感词识别方法[J]. 常州工学院学报, 2022, 35(3): 29-33.
[9] 李瀛, 王冠楠. 网络新闻敏感信息识别与风险分级方法研究[J]. 情报理论与实践, 2022, 45(4): 105-112.
[10] 陈可嘉, 郑晶晶. 基于种子约束 LDA 的产品属性提取方法[J]. 华南理工大学学报(自然科学版), 2022, 50(6): 37-48, 70.

（上接第 9 页）

4 结论

本文中机电引信采用主动端发射时的极低过载环境实现第一级保险的解保方案，该方案中拔销器直接锁定隔爆件，低过载开关仅作为信号使用，平时处于悬浮状态。由于机电引信在勤务处理、跌落等环境时引信未上电，即使低过载开关闭合持续时间达到电路触发的要求，保险仍不会解除。只有在作战时，引信上电，且低过载开关感受主动段过载闭合，闭合持续时间达到解保电路的触发时间时，引信才可以解除第一级保险，保证引弹药在发射解除之前的任何阶段的安全性，满足 GJB 373B—2019 引信安全性设计准则中“除非出现预定的发射情况，不应启动解除隔离流程”的要求[11]。

笔者根据火箭弹、巡飞弹、反坦克导弹及新型自杀式无人作战平台等弹药对极低过载引信保险的需求，以实现引信的第一级保险方案为研究对象，开展了极低过载惯性开关设计、低过载环境识别等关键技术研究，解决极低发射过载环境下引信安全系统的环境利用和保险问题，实现引信在极低过载环境下可靠解除第一级保险的功能，为引信安全系统的工程应用设计和解保策略的制定提供参考[12]。由研究分析结果可知：该方案原理可行，满足引信安全性设计准则要求，但后续需要进行模拟解保试验验证。

参考文献：

[1] SAM R, DANNY C, HOPPER C, et al. Low G MEMSinertia switches for fuzing applications[C]. Proceedings of the NDIA 61th Annual Fuze Conference. San Diego USA: CA, 2018: 15–17.
[2] 艾志远, 李世中, 杨超, 等. 极低过载引信无返回力矩擒纵后坐保险机构[J]. 探测与控制学报, 2023, 45(1): 44-49.
[3] 王海龙. 引信低过载后坐保险机构技术研究[D]. 南京: 南京理工大学, 2020.
[4] 李波, 李世中. 引信环境及其应用[M]. 北京: 国防工业出版社, 2016.
[5] 魏峥嵘, 姚宝珍, 冉雪磊, 等. 引信对极薄弱目标的动态发火测试方法[J]. 兵器装备工程学报, 2020, 41(7): 64-68.
[6] Livermore Software technology corporation. LS_DYNA Keyword users’manual[C]. California: Livermore, 1992: 0712.
[7] 宫雪峰, 李豪杰, 陈志鹏, 等. 基于能量域的引信环境识别[J]. 探测与控制学报, 2023, 45(1): 11-16.
[8] DU L Q, JIA S F, NIE W R, et al. Fabrication of Fuze Microelectro-mechanical System Safety Device[J].中国机械工程学报, 2011, 24(5): 836-841.
[9] 黄戈, 姚宝珍, 冉雪磊. 引信对极薄弱目标发火的作用方法[J]. 探测与控制学报, 2019, 41(4): 36-41.
[10] 李雪丽, 陈伟, 李运生, 等. 基于 Proteus 的双直流稳压电源的设计与仿真[J]. 制造技术与机床, 2010(12): 53-57.
[11] 引信安全性设计准则 GJB 373B—2019[S/OL]. https://max.book118.com/html/2023/1228/5312310144011032.shtm
[12] WEN Q, WANG Y S. Research on Simulation of Langlie Method Test of Fuze Arming Distance[J]. 兵工学报, 2009(3): 221-227.