

## 基于文本分类和 SVM 的雷达侦察装备故障诊断研究

仝奇<sup>1</sup>, 叶霞<sup>1</sup>, 李俊山<sup>1</sup>, 张仲敏<sup>2</sup>

(1. 第二炮兵工程大学信息工程系, 西安 710025; 2. 西安通信学院, 西安 710106)

**摘要:** 针对自然语言描述的雷达侦察装备故障诊断问题, 提出了一种基于文本分类技术和支持向量机的故障诊断方法。首先, 对获取的故障文本集进行分析, 提取故障特征建立故障特征词库; 然后, 采用布尔模型实现文本向量的表示, 构建故障向量库; 最后, 通过 SVM 多分类中的一对一算法建立故障诊断分类模型, 并采用网格搜索法进行参数优化, 实现了雷达侦察装备的故障诊断。实验分析验证了该方法的有效性和正确性, 并最终将故障诊断的最大识别精度提高到 90%。

**关键词:** 雷达侦察装备; 故障诊断; 文本分类; 支持向量机

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1671-637X(2016)02-0094-05

## Fault Diagnosis of Radar Reconnaissance Equipment Based on Text Classification and SVM

TONG Qi<sup>1</sup>, YE Xia<sup>1</sup>, LI Jun-shan<sup>1</sup>, ZHANG Zhong-min<sup>2</sup>

(1. Department of Information Engineering, The Second Artillery Engineering University, Xi'an 710025, China;  
2. Xi'an Communications Institute, Xi'an 710106, China)

**Abstract:** To the fault diagnosis of radar reconnaissance equipment described with natural language, a fault diagnosis approach based on text classification and support vector machine was proposed. First, a fault feature lexicon was built up by analyzing the fault text sets and extracting the fault features. Then, the text vectors were represented with the Boolean model to construct the fault vector library. Finally, the diagnosis classification model was established by One-Against-One method of the SVM classification, and the parameters were optimized by using grid search method. And thus the fault diagnosis was implemented. The experimental result shows the effectiveness and the validity of the approach, and the maximum recognition accuracy of fault diagnosis could be improved to 90% ultimately.

**Key words:** radar reconnaissance equipment; fault diagnosis; text classification; Support Vector Machine(SVM)

### 0 引言

雷达侦察装备作为电子战的重要组成部分, 在现代高技术战争中发挥着极其重要的作用。虽然雷达侦察装备对确保战争的胜利提供了可靠保障, 但对其故障诊断也提出了更高的要求。因此, 研究与装备相适应的故障诊断技术, 形成与装备发展相适应的维修保障能力, 已成为装备保障研究的重要内容。

许多研究者在雷达装备故障诊断方面进行了相关研究。文献[1]先利用小波变换对雷达装备数据进行处理和特征提取, 建立故障诊断模型, 然后利用最小二

乘支持向量机进行故障诊断, 提高了雷达故障诊断的正确性和效率; 文献[2]先对装备的各性能参数监测数据进行了标准化处理, 然后将提取出的特征向量输入到建好的 HMM 模型进行训练, 设计出了复杂设备分类器; 文献[3]采用粗糙集对雷达系统进行故障诊断规则提取, 约简了故障诊断所需的测试参数, 简化了整个测试过程; 文献[4]将支持向量机应用于雷达板级电路进行故障诊断, 建立了故障诊断样本训练和测试平台, 与普遍采用的专家系统、字典诊断等方法相比, 具有较好的诊断效果。

目前雷达侦察装备故障主要描述方法有自然语言描述、信号量描述、状态表述等。由于信号量描述不具有通用性, 实时状态表述不具有可操作性, 而采用自然语言描述既能保证准确性又便于理解, 故本文对雷达

侦察装备故障采用自然语言处理方式,结合文本分类和SVM进行故障诊断,从装备故障的成因、部位、现象及维修方法中提取出特征词,将其表示成特征向量形式后转入基于SVM的故障诊断模块,最后运用SVM算法对某雷达侦察装备故障进行训练和诊断,最终获得了较理想的试验结果。

## 1 理论基础

### 1.1 文本分类

文本分类<sup>[5]</sup>是对所要进行分类的文本,给出每一类所预先定义类别标号,对文本进行准确、高效的分类。

文本分类也可以用以下数学形式描述。

文本分类就是一个二元组  $(d_i, c_j) \in D \times C$  映射到一个布尔值的任务。该映射可表示为

$$f: D \times C \rightarrow A \quad (1)$$

式中:函数  $f$  表示一个分类器或者分类模型;  $D = (d_1, d_2, \dots, d_n)$  是一个无限集合,表示待分类的文本的集合;  $C = (c_1, c_2, \dots, c_m)$  必须是有限集合,表示给定分类体系下所有预先定义的类别的集合;  $A = (a_{ij})_{n \times m}, a_{ij} = \{0, 1\}$ ,  $A$  表示分类结果的集合,  $a_{ij} = \{0, 1\}$  表示文本  $d_i$  按照其内容是否属于类别  $c_j$ , 属于,用“1”表示,不属于,则用“0”表示。

一般通过以下步骤实现文本分类:

- 1) 获取进行分类的文本集;
- 2) 选择文本分类模型;
- 3) 根据所选的分类模型,构建文本的特征向量;
- 4) 通过训练样本建立文本分类器;
- 5) 通过测试样本预测分类,并根据预测结果进行参数的选择以优化分类器。

融合文本分类应用于故障诊断需解决4个问题,即获取训练故障库、文档特征选择和故障库表示、选择分类算法、性能评估<sup>[6]</sup>。具体过程如图1所示。

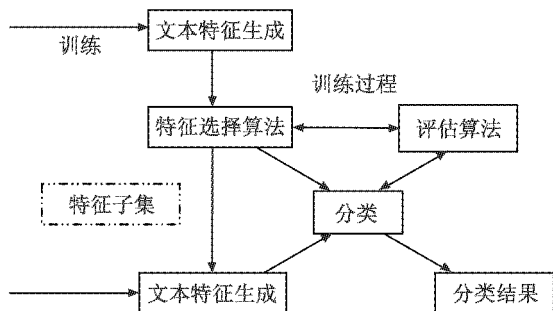


图1 文本分类过程

Fig.1 Text classification process

### 1.2 支持向量机

支持向量机 (Support Vector Machine, SVM) 是 Vapnik<sup>[7]</sup> 等人首先提出的,主要思想是建立一个超平面作为决策曲面,使得正例和反例之间的隔离边缘最大化,目前在故障诊断领域的应用越来越广<sup>[8-9]</sup>。

设有两类可分的样本集合:  $(x_i, y_i), i = 1, \dots, n, x_i \in R^d, y_i \in \{+1, -1\}$ , 存在分类面方程为

$$w \cdot x + b = 0 \quad (2)$$

满足

$$y_i [(w \cdot x_i) + b] - 1 \geq 0 \quad i = 1, \dots, n \quad (3)$$

则称训练样本集是线性可分的,其中,  $w$  是超平面的法向量,  $b$  是超平面的常数项,寻找最优的分类超平面就是找最优的  $w$  和  $b$ , 即最优分类超平面等价于求最大间隔,有了超平面就可以推出判别函数。

SVM 在解决线性不可分问题时,利用核函数将数据样本映射到高维空间,通过引入一个非负的松弛因子  $\xi_i \in R$  实现问题的求解,即

$$\min \left( \frac{1}{2} \|w\|^2 \right) + C \sum_{i=1}^m \xi_i \quad (4)$$

$$\text{s. t. } y_i (w \cdot x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, m$$

式中,  $C$  表示惩罚系数,其值越大表示对错误分类的惩罚越大。式(4)的求解可通过如下的二次规划来实现。

$$\max L(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=0}^m \alpha_i \alpha_j y_i y_j x_i x_j \quad (5)$$

$$\text{s. t. } \sum_{i=1}^m y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, m$$

式中,  $\alpha_i$  表示式(2)第  $i$  个约束的拉格朗日乘子。经非线性映射函数  $\varphi$  将输入数据映射到高维特征空间,为了避免高维空间点积  $\varphi(x_i) \cdot \varphi(x_j)$  的复杂运算,引入满足 Mercer<sup>[10]</sup> 理论的核函数  $K$ , 即

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) \quad (6)$$

则相应的判别函数为

$$y(x) = \text{sgn} \left( \sum_{i=1}^m \alpha_i^* y_i K(x_i, x) + b^* \right). \quad (7)$$

## 2 设计思路

本文基于文本分类技术和支持向量机进行雷达侦察装备故障诊断的设计思路如下所述。

一是利用 SQL Server 建立相应的故障特征词库。词库的建立一方面是从系统结构、工作原理、故障机理和长期积累的故障知识等的分析中产生;另一方面是从装备重要部位的温度、气味、面板的异常显示、测试信号值等状态信息中产生。

二是利用 C++ Builder 软件通过布尔模型实现文本向量的表示,如图2所示。

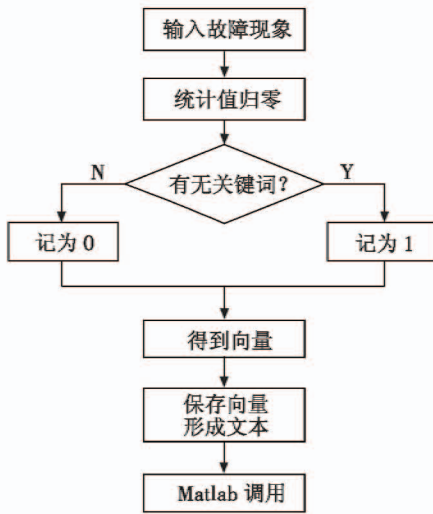


图 2 向量表示流程

Fig. 2 Flow chart of vector construction

通过查询串  $q$  和文本的相似度来判断它们是否相关,相似度算式为

$$Sim(d, q) = \begin{cases} 1 & q \in d \\ 0 & q \notin d \end{cases} \quad (8)$$

由式(8)可知,如果值为 1,则表示它们相关,否则表示它们不相关。本文通过遍历数据,集中保存的方法,得到雷达装备的故障向量库。

三是通过 Matlab,利用 SVM 算法对装备故障进行训练和测试。本文采用 SVM 多分类算法中较经典的一对一(One-Against-One)<sup>[11]</sup>算法进行诊断,将某雷达侦察装备故障分为 8 类,也就是 8 种故障模式,对每一种模式来说,可根据两类 SVM 分类机求得实值函数  $g^{i-j}(x)$  和判定样本  $x$  是属于第  $i$  类或  $j$  类的分类机,即

$$f^{i-j}(x) = \begin{cases} i & g^{i-j}(x) > 0 \\ j & \text{其他} \end{cases} \quad (9)$$

然后就可以对测试样本  $x$  进行类别测试,结果属于  $i$  类表示  $i$  类加一票,属于  $j$  类表示  $j$  类加一票,哪一类票数最多  $x$  就属于哪一类。因此,可以把雷达装备中获取的故障样本分为训练样本和测试样本,训练、测试后就可用于实际诊断。

整个设计的主要思路如图 3 所示。

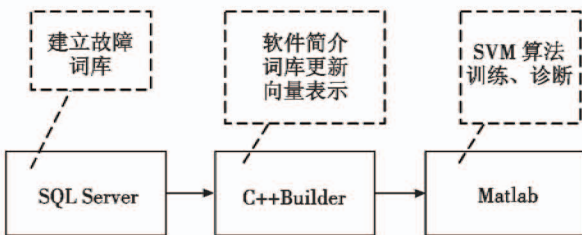


图 3 设计思路图

Fig. 3 Design idea

### 3 基于“一对一”算法的故障诊断方法

#### 3.1 文本挖掘

文本挖掘包括故障特征词库的建立和文本向量的表示,这个过程是 SVM 算法进行诊断的前提和基础。由于文本都是非结构化的数据,因此需要将其表示成计算机能够处理的形式,也就是用简单而准确的数学模型将其表示成计算机能够处理的结构化数据。

特征词库的建立是从雷达侦察装备的结构、功能和状态信息中梳理出能区别各类别故障的典型故障文本。而实现词库的建立需要对装备进行调研、资料收集、整理、分析和测试以及与相关领域专家进行深入交流。现采用 SQL Server 实现故障特征词库的建立。

故障文本向量的表示是将故障文本进行分词后,由这些特征词作为特征向量的维来表示故障文本。现采用 C++ Builder 软件通过布尔模型设计实现了装备的基于人机交互的故障特征词库建立的程序,实现了特征向量的表示。

#### 3.2 One-Against-One 算法的故障诊断流程

One-Against-One 算法的主要思想是组合多个两类分类器以构成多类分类器来对样本进行训练诊断。当有  $k$  个类别时,它需要构造  $k(k-1)/2$  个两类分类器,假设将样本  $x$  输入到由  $k$  类中第  $m$  类和  $n$  类构造的两类分类器,其分类函数为

$$f^{mn}(x) = \text{sgn} \left\{ \sum_{\text{支持向量}} \alpha_i y_i^{mn} K(x_i, x) + b^{mn} \right\} \quad (10)$$

例如将 One-Against-One 算法应用于装备的 4 类问题时需要构建 6 个两类分类器,分解结果见表 1。

表 1 分解策略示意图

Table 1 Indication of the decomposition strategy

模式	类别 1	类别 2	类别 3	类别 4
1	+1	-1	0	0
2	+1	0	-1	0
3	+1	0	0	-1
4	0	+1	-1	0
5	0	+1	0	-1
6	0	0	+1	-1

当将算法应用于 8 类问题时,须构建 28 个两类分类器,其实现的思想过程跟上述是一样的。现故障词库中词的数量为 629,将其中 240 个故障都表示成向量形式,人工设定故障所属类别。按照雷达侦察装备的组成和诊断方案把故障目标类型分为 8 类:控制车配电部分、控制车控制部分、侦察车接收部分、侦察车干扰部分、侦察车伺服部分、侦察车电源自检、电源车油机、电源车输出显示,将每类选取出的样本分为训练样本和测试样本,经训练和测试后便可将其应用于实际的装备故障诊断。SVM 算法整个训练诊断过程如图 4 所示。

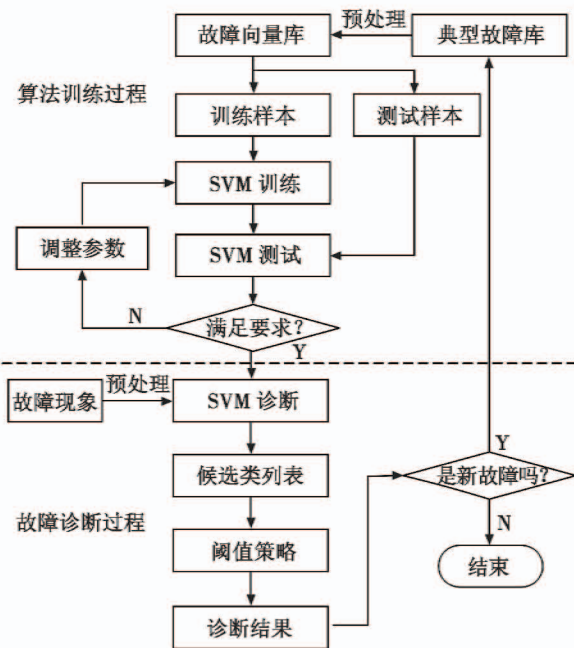


图 4 SVM 算法训练诊断流程图

Fig. 4 Flow chart of SVM training algorithm

### 3.3 SVM 训练、测试及诊断过程

SVM 训练主要是为了求出  $w$  和  $b$  的最优解,得到判别函数来建立故障诊断训练模型。一对一算法训练是将两种状态各训练样本作为分类器的两类输入,分别标识为 +1 和 -1,对应 8 类问题,共建立 28 个两类分类器。通过式(5)的二次规划寻优得到  $\alpha_i$  的最优解  $\alpha_i^*$ ,再根据样本的输入,求解出  $w$  的最优解为

$$w^* = \sum_{i=1}^N \alpha_i^* y_i x_i \quad (11)$$

$b^*$  可以由  $w^*$  和式(3)求出,最终得到判别函数即式(7),从而建立雷达侦察装备的故障诊断训练模型。

SVM 测试是将测试样本  $x$  代入式(7)中进行预测。各输入测试样本集通过训练模型输出相应的类别,并把输出的类别与已知类别做对比,进而可以得到预测的准确率,通过预测出的准确率来判断是否进行参数的调整,以获取更高的准确率。

经过训练和测试,建立了准确率相对较高的故障诊断模型,先将雷达侦察装备故障数据进行预处理,进而通过模型诊断出相对应的故障类别。

## 4 实验分析

通过 Matlab 实现算法的训练及诊断。

1) 清零,调用故障样本文档。

通过 Matlab 调用故障文本所对应的故障特征向量,首先对所有类别的值进行清零,然后将这些故障样本进行分类。分类时按故障样本所对应的故障测试点进行类别标注,并在  $m$  文件中标注在矩阵向量的最后

一列,将 240 个故障按检测点及所属的典型故障类别定义为 8 类,即无故障为 0,控制车配电部分为 1,控制车控制部分为 2,侦察车接收部分为 3,侦察车干扰部分为 4,侦察车伺服部分为 5,侦察车电源自检为 6,电源车油机为 7,电源车输出显示为 8,如表 2 所示。

表 2 故障类别代码

Table 2 Fault category code

故障检测点	代码	故障检测点	代码
正常	0	侦察车伺服部分	5
控制车配电部分	1	侦察车电源自检	6
控制车控制部分	2	电源车油机	7
侦察车接收部分	3	电源车输出显示	8
侦察车干扰部分	4		

在完成样本的分类形成文档后,通过 Matlab 调用,就可显示出雷达侦察装备的 8 个类别故障。

2) 进行数据的预处理,完成函数定义。

数据的预处理对后续故障的诊断效果起着非常重要的作用,因此在进行故障的训练和诊断之前,需要对故障样本进行预处理,即进行归一化处理:  $sa = [saa1; saa2; saa3; saa4; saa5; saa6; saa7; saa8]$ 。在完成数据预处理后,需要把训练和测试函数,即  $x\_train, y\_train, x\_test, y\_test$  定义出来,但是需要注意一点,由于分类好的特征向量最后一列标注的是故障类别,需要记空,以免干扰。

3) 故障的训练和诊断。

在故障的训练和诊断中,采用了径向基核函数将 240 个故障一半用作训练,一半用作预测,考虑到整个训练和测试中样本相对较少,而函数中的  $C$  和  $\gamma$  两个参数又对试验效果有重要影响,因此可通过最优参数的选择来达到较高的故障识别率。

4) 采用网格搜索法进行最优参数的选择,以优化装备的诊断效果。

如果单凭经验给参数赋值往往达不到较高的分类识别率,因此可通过网格搜索法进行最优参数的选择,以此来表示出取不同值时对最大识别率的影响效果,得到识别率定义为  $ClassRate; cor(k+6, l+3) = ClassRate$ ; 参数表示为  $a\_best = \max(cor)$ ; 经调试,列出了选取不同的参数值时对最大识别率的影响效果如表 3 所示。

表 3 统计结果

Table 3 Statistical results

参数			参数		
$\gamma$	$C$	最大识别率	$\gamma$	$C$	最大识别率
0.031 25	0.25	0.400 0	1	8	0.700 0
0.062 5	0.5	0.500 0	2	16	0.800 0
0.125	1	0.700 0	4	32	0.800 0
0.25	2	0.700 0	8	64	0.900 0
0.5	4	0.700 0	16	128	0.900 0

表 3 的结果显示可以取得最大识别率时的最优参数值。

结合整个雷达侦察装备的故障诊断过程,采用 One-Against-One 进行分类的算法流程如图 5 所示。

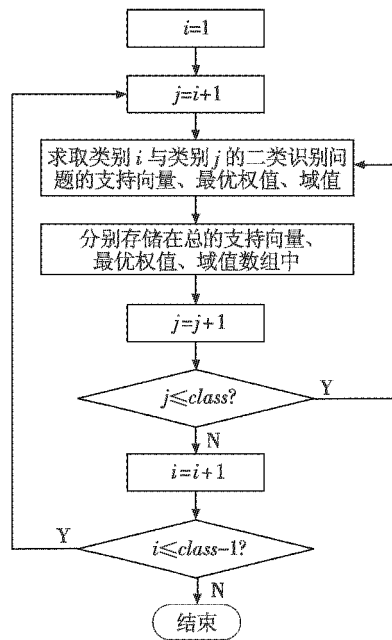


图 5 One-Against-One 分类训练算法流程图

Fig. 5 Flow chart of OAO classification training algorithm

## 5 结论

本文提出了一种基于文本分类技术和支持向量机的雷达侦察装备故障诊断方法。该方法利用文本分类技术建立了相应的故障特征词库,并通过布尔模型实现了文本向量的表示,最后又运用 SVM 算法实现了雷达侦察装备的训练和诊断,识别率可达 90%,取得了较好的试验效果。通过这种方法进行故障诊断,可使装备维修人员根据已有的故障库快速准确地定位故障,节约了检测和维修时间,为部队作战提供了可靠的保障。

## 参考文献

[1] 涂望明,宋执环,陈运涛,等. 基于小波变换和 LS-SVM 的雷达故障诊断[J]. 控制工程,2013,20(2):309-312. (TU W M, SONG Z H, CHEN Y T, et al. Radar fault diagnosis based on wavelet transformation and LS-SVM[J]. Control Engineering of China, 2013, 20(2):309-312.)

[2] 崔展宁,白云,徐宇亮. 基于 HMM 的某型雷达接收机电电子设备故障诊断研究[J]. 计算机测量与控制,2013,21

(11):3011-3013. (CUI Z N, BAI Y, XU Y L. Method of fault diagnosis for radar receiver based on HMM[J]. Computer Measurement and Control, 2013, 21(11):3011-3013.)

[3] 罗晓军,文莹,方甲永. 复杂航空电子装备故障诊断规则提取[J]. 计算机测量与控制,2012,20(2):294-296. (LUO X J, WEN Y, FANG J Y. Extracting diagnosis rules of complicated avionics under condition of small-sample[J]. Computer Measurement and Control, 2012, 20(2):294-296.)

[4] 张金泽,段修生,李和平,等. 支持向量机技术的雷达板级电路快速故障诊断[J]. 火力与指挥控制,2012,20(2):294-296. (ZHANG J Z, DUAN X S, LI H P, et al. Fast fault diagnosis of board-level circuit for radar based on SVM[J]. Fire Control and Command Control, 2012, 20(2):294-296.)

[5] CHOI H, VARIAN H. Predicting the present with google trends[J]. Economic Record, 2012, 88(s1):2-9.

[6] 段修生. 基于支持向量机的某火控系统故障诊断技术研究[D]. 石家庄:军械工程学院,2009. (DUAN X S. Research on fault diagnosis for fire control system based on support vector machine[D]. Shijiazhuang: Ordnance Engineer College, 2009.)

[7] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995:205-208.

[8] 王维刚,刘占生. 多目标粒子群优化的支持向量机及其在齿轮故障诊断中的应用[J]. 振动工程学报,2013,26(5):743-750. (WANG W G, LIU Z S. Support vector machine optimized by multi-objective particle swarm and application in gear fault diagnosis[J]. Journal of Vibration Engineering, 2013, 26(5):743-750.)

[9] 刘任洋,吴文全,李超,等. HMM-SVM 混合模型在模拟电路故障诊断中的应用[J]. 电光与控制,2013,20(2):89-92. (LIU R Y, WU W Q, LI C, et al. Application of HMM-SVM in fault diagnosis of analog circuits[J]. Electronics Optics & Control, 2013, 20(2):89-92.)

[10] INOUE T, ABE S. Fuzzy support vector machine for pattern classification [C]//Proceedings of International Joint Conference on Neural Networks, Washington, USA, 2001:1449-1454.

[11] DEBNATH R, TAKAHIDE N, TAKAHASHI H. A decision based on one-against-one method for multi-class support vector machine[J]. Pattern Analysis and Applications, 2004, 7(2):164-175.