

# 改进的贝叶斯算法在智能变电站 网络故障诊断系统中的研究<sup>\*</sup>

周凤丽 李 聪 伍永豪

(武汉科技大学城市学院信息工程学部 武汉 430083)

**摘 要** 提高智能变电站网络故障诊断能力对于确保电力系统的稳定运行和供电可靠性具有重要意义,故障的分类是目前智能变电站网络故障诊断系统所面临的一个主要问题。常用分类算法存在着训练数据多样化,特征的选择标准具有不确定性,学习潜力匮乏等问题,文章在原有贝叶斯算法的基础上加入了特征的选择标准及学习过程,实验结果表明改进的贝叶斯算法在很大程度上能有效解决故障分类问题,从而提高智能变电站的网络故障诊断能力。

**关键词** 贝叶斯算法;智能变电站;故障诊断;故障分类

**中图分类号** TP391 **DOI**:10.3969/j.issn1672-9722.2014.02.001

## An Improved Bayesian Algorithm in the Intelligent Substation Network Fault Diagnosis System

ZHOU Fengli LI Cong WU Yonghao

(Faculty of Information Engineering, City College of Wuhan University of Science & Technology, Wuhan 430083)

**Abstract** To improve the fault diagnosis ability of the intelligent substation is very important for stable operation of power system and power supply reliability. One of the major problems that the power network fault diagnosis system faced is the fault classification problem. The existing classification algorithm has some shortcomings, such as the training samples imbalance, lack of consistent characteristics, weakness learning ability. The feature selection and learning strategies are added to Bayesian algorithm to propose an improved Bayesian fault classification algorithm. The experimental results show that the improved Bayesian algorithm can solve the fault classification problem effectively, and enhance the fault diagnosis capability of intelligent substation network.

**Key Words** Bayesian algorithm, intelligent substation, fault diagnosis, fault classification

**Class Number** TP391

### 1 引言

随着网络的广泛应用,电力系统中各种监控设备的规模也在不断扩大,智能变电站网络故障的诊断则更为重要。因为一旦智能变电站网络发生故障,现场电脑保护、安全自动装置、故障录波器会记录到大量的故障数据,并在很短一段时间内送到调度端,以供调度和继电保护专业技术人员进行分析<sup>[1]</sup>。但是,这些故障数据可能存在复杂的关系,

从而导致同一种类型的网络故障表现出了不同的症状,几种不同类型的网络故障却通过相互作用引发了相同的表现,因此必须要有一个准确高效的故障分类算法来获得引发智能变电站网络异常的最有可能的故障集合,从而提高智能变电站网络故障诊断能力。

### 2 智能变电站网络故障分类处理过程及存在问题

<sup>\*</sup> 收稿日期:2013 年 8 月 13 日,修回日期:2013 年 9 月 27 日

基金项目:湖北省教育厅科学技术研究计划指导性项目(编号:B2013257)资助。

作者简介:周凤丽,女,硕士,讲师,研究方向:计算机网络与通信,物联网,智能计算。李聪,女,硕士,讲师,研究方向:人工智能。伍永豪,男,硕士,讲师,研究方向:软件工程。

## 2.1 智能变电站网络故障分类处理过程

智能变电站网络故障诊断系统通过不同的途径获取到了大量的原始观测数据,在这些数据中得到的故障特征信息越准确,系统的诊断能力就越强,因此应该充分利用这些信息来完成故障诊断;另外,智能变电站网络故障诊断系统需要监测的对象和需要处理的故障较多,需处理的信息量也非常大,因此我们不能迅速获得所有相关的智能变电站网络故障信息<sup>[2]</sup>。而事实证明,即便能够将全部信息都收集起来,也会存在如何对这些相互关联的信息进行分类的问题。

分类处理就是将大量的待分类数据归到不同的类别中,其处理过程如图 1 所示<sup>[3]</sup>。



图 1 分类处理一般过程

如图 1 所示,分类处理由训练阶段及分类阶段构成。训练过程中先对已分类数据进行学习,提取出相应特征后可得到一个分类器;然后在分类过程中使用分类器去计算和分析未被分类数据,得到分类的最终结果。

## 2.2 分类处理过程中存在的问题

由于智能变电站网络系统非常复杂,如果网络突发故障,在对网络故障做分类操作时可能会牵涉到多个方面的不同知识,而要想使得分类效果更好,提高智能变电站网络故障的诊断能力,可综合利用不同的策略<sup>[4]</sup>。目前用于对智能变电站网络故障进行分类的方法非常多,经过研究后发现有如下不足之处:

1) 训练数据多样化:一般情况下,在网络中最容易捕获的训练数据主要有 UDP、ICMP 和 TCP 报文等,但这些数据的质量有着很大差异,甚至有一些错误的信息,如果不经过数据过滤而直接去学习,分类的效果会大打折扣。

2) 特征的选择标准具有不确定性:电网如果突发故障,外在表现差异很大,而且不同类型的故障有着不同的特征,故障特征的选取会直接影响到分类的准确度<sup>[5]</sup>。

3) 学习潜力匮乏:如图 1 所示,分类过程完成之后不能再获得实际的分类处理更新信息,一旦实际情况有所变化,分类模型也不会改变,因此是缺乏学习能力的,当系统的运行时间越来越长,数据量越来越大,分类性能会越来越弱<sup>[6]</sup>。

4) 多模型处理技术:目前分类方法很多,如神

经网络、Rule-based、Naïve Bayesian、Decision Tree 和 SVM 等,不同方法从不同方面实现了对分类问题的描述。Naïve Bayesian 可允许假设做出不确定性的预测,相比其它方法具备一定的先进性<sup>[7]</sup>。

## 3 改进的贝叶斯故障分类算法

基于概率知识表达的贝叶斯理论具有强大的不确定性问题处理能力,而 Bayesian 网络能够有效地描述与网络故障相关的不同信息以及它们之间的关系。

由于 Naïve Bayesian 计算简单并且结果精度高,已广泛用于故障分类等问题中<sup>[8]</sup>。文章对原有 Bayesian 算法进行了改进,通过引入特征选取和学习过程来更好地解决智能变电站故障诊断系统中的故障分类问题。

### 3.1 改进贝叶斯算法的相关策略

先给出改进的贝叶斯算法的相关定义(变量用大写字母表示,其取值用小写字母)。

定义 1  $X = \langle X_1, X_2, \dots, X_n \rangle$  表示某一类型的智能变电站网络故障所对应特征向量,特征分量分别为  $X_1, X_2, \dots, X_n$ 。

定义 2  $DFS = \{DF_i | i = 1, \dots, m\}$  是已知的故障类型。

定义 3  $p(DF_i/x)$  是  $p(DF = DF_i/X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$  的缩写,代表在出现特征  $x$  的情况下,发生故障  $DF_i$  的概率。

为了计算  $p(DF_i/x)$ ,根据 Bayes 概率公式有:

$$p(DF_i/x) = p(x/DF_i)p(DF_i)/p(x) \quad (1)$$

式(1)中  $p(x/DF_i)$  表示某类故障  $DF_i$  发生时,该类故障的特征向量显示为特征  $x$  的概率,  $p(DF_i)$  表示发生  $DF_i$  类故障的概率,也是先验概率,  $p(x)$  则代表某类故障的特征向量等于  $x$  发生的概率。

#### 3.1.1 特征选取

为了将数据处理进行简化,假设特征分量都是独立的。

假设 1 假设包含在特征向量里面的所有特征分量之间各自独立。

根据假设 1,  $p(x/DF_i)$  可使用式(2)计算:

$$p(x/DF_i) = \prod_{j=1}^n p(x_j/DF_i) \quad (2)$$

$p(x)$  可使用式(3)计算:

$$p(x) = \prod_{j=1}^n p(x_j) \quad (3)$$

其中  $x_j$  为  $x$  第  $j$  个分量值,  $p(x_j/DF_i)$  表示如果  $DF_i$  类的故障发生,特征  $x_j$  出现的概率。

构建一个 Bayesian 分类器要解决的关键问题是特征向量的选取标准,因为特征向量的构成结构以及构成长度都将极大地影响了故障诊断的准确性<sup>[8]</sup>。

一旦发生一个故障,智能变电站的网络会表现出不同的异常,因此不同故障的特征向量是不完全相同的。实际特征向量选取是根据故障类型来完成的,而且构成该特征向量的特征分量也要最大限度地反映故障的特征并尽可能的符合特征独立性假设<sup>[9]</sup>。下面用  $x_i$  表示特征向量中包含的特征矢量取值。

定义 4  $O(DF_i) = P(DF_i) / P(\neg DF_i)$  为  $DF_i$  类故障的先验概率。计算先验概率时可使用式(4), Sample 表示训练数据集:

$$O(DF_i) = \frac{|\{d | d \in DF_i \cap d \in \text{Sample}\}|}{|\{d | d \notin DF_i \cap d \in \text{Sample}\}|} \quad (4)$$

定义 5 相对于特征  $x_j$  和  $DF_i$  类故障,其似然比被定义为

$$L(x_j | DF_i) = p(x_j | DF_i) / p(x_j | \neg DF_i) \quad (5)$$

定理 1 后验几率

$$O(DF_i | x_j) = \frac{p(DF_i | x_j)}{p(\neg DF_i | x_j)} = L(x_j | DF_i) \cdot O(DF_i)$$

根据定义 4 和定义 5,定理 1 成立。

实际使用时,可先通过  $DF_i$  类故障样本集中出现特征  $x_j$  的频率来计算  $p(x_j | DF_i)$ ,通过非  $DF_i$  类故障样本集中出现  $x_j$  的频率来计算  $p(x_j | \neg DF_i)$ ,然后再依据式(4)、(5)和定理 1 可求出特征  $x_j$  出现时发生  $DF_i$  类故障的后验几率,该结果可直接反映选取特征  $x_j$  时,对  $DF_i$  类故障进行分类的效果。

定义 6 故障类型的不同,特征也会不同,因此要给不同的故障类型分别定义所对应的特征矢量,定义  $DF_i$  类故障的二进制特征矢量为  $X_i$ ,  $X_i = \langle X_{i1}, X_{i2}, \dots \rangle$ 。

定义 7 针对 DFS 中每一个类型的故障  $DF_i$  和  $DF_i$  所对应的特征向量,从中选出最大的  $n$  个  $O(DF_i | x_{ij})$ ,构成针对每一个故障类型  $DF_i$  的特征向量  $X_i = \langle X_{i1}, X_{i2}, \dots, X_{in} \rangle$ 。当某个特征  $X_{ij} = x_{ij}$  时,相应的分量  $x_{ij} = 1$ ,否则  $x_{ij} = 0$ 。

### 3.1.2 学习策略

经过一次训练后的分类系统,将不再从分类过程中获得更新的信息,在实际的使用过程中也不会根据实际情况调整分类模型,是缺乏学习能力的<sup>[10]</sup>。因此,我们使用了学习策略,在原有的系统对原始模型进行修正,使系统性能得到了优化。如图 2 所示,在原有的分类系统中引入了学习策

略,形成了一个闭环结构。

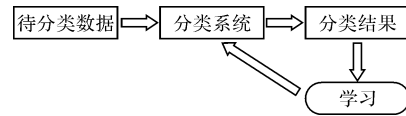


图 2 引入了学习策略的分类系统

在实际的学习过程中,学习重点是分类错误的的数据,经过不断地学习可对分类模型进行修正,从而使得分类性能大大提升<sup>[11]</sup>。

### 3.2 改进的贝叶斯算法

文章在原有 Bayesian 算法的基础上加入了前文描述的特征选取以及学习策略。描述如下:

#### • 构造和训练分类阶段

1) 从智能变电站网络系统中随机选取构成训练样本集 Sample 的  $N$  个样本,根据专家知识和经验,对这  $N$  个样本进行人为故障分类。

2) 针对 Sample 中包含的  $DF_i$  类故障及其所对应的全部特征,计算  $p(x_{ij} | DF_i)$ ,  $p(x_{ij} | \neg DF_i)$ ,  $p(DF_i)$ ,  $p(\neg DF_i)$  和  $p(x_{ij})$ ,其中  $p(x_{ij} | DF_i) = N_{DF_i x_{ij}} / N_{DF_i}$ ,  $N_{DF_i}$  表示 Sample 中共有多少个  $DF_i$  类的故障,  $N_{DF_i x_{ij}}$  为在  $DF_i$  类的故障中出现特征  $x_{ij}$  的次数。  $p(x_{ij} | \neg DF_i) = N_{\neg DF_i x_{ij}} / N_{\neg DF_i}$ , 其中  $N_{\neg DF_i}$  表示 Sample 中共有多少个非  $DF_i$  类的故障,  $N_{\neg DF_i x_{ij}}$  为非  $DF_i$  类的故障中出现特征  $x_{ij}$  的次数。  $p(DF_i) = N_{DF_i} / N$ ,  $p(x_{ij}) = N_{x_{ij}} / N$ , 其中  $N_{x_{ij}}$  表示特征  $x_{ij}$  在样本集 Sample 中的出现次数。

3) 根据定理 1 首先对每一类故障  $DF_i$  的后验几率  $O(DF_i | x_{ij})$  进行计算,从中选取  $m$  个最大的  $O(DF_i | x_{ij})$ ,构成每个故障类型  $DF_i$  的对应特征向量  $X_i = \langle X_{i1}, X_{i2}, \dots, X_{im} \rangle$ 。

#### • 分类阶段

4) 如果智能变电站的网络系统出现故障,先获取故障症状的数据信息及其特征向量,然后利用式(1)计算发生  $DF_i$  类故障的后验概率  $p(DF_i | x_i) = \prod_{j=1}^m p(x_{ij} | DF_i) p(DF_i) / \prod_{j=1}^m p(x_{ij})$ 。选择  $p(DF_i | x_i)$  中最大的  $DF_m$  当作目前网络系统中最可能出现的故障。如果  $DF_m$  是真正引发网络系统故障的原因,结束分类,否则步入学习阶段。

#### • 学习阶段

5) 如果分类结果中存在错误,将其修正后再次送回至训练集 Sample 中,重新完成训练和分类过程。

## 4 测试分析

为了比较文中所提出的改进 Bayesian 算法和朴

素 Bayesian 算法,我们先从智能变电站网络系统中选取 1000 个已经发生过的网络事件用于构造症状集,并将其划分为 10 类网络故障,然后从症状集中选择  $x$  个事件用于构建训练集,验证集则由症状集中剩下的事件来构成,最终得到如图 3 所示的结果。

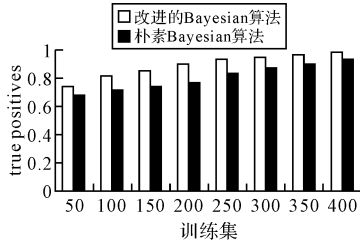


图 3 改进的 Bayesian 算法和朴素 Bayesian 算法的分类精度

如图 3 所示,当训练集增大时,这两个算法的分类精度都有所上升,由于改进的 Bayesian 算法对特征值的选取进行了优化,其分类精度大概高出了 4%~9%。

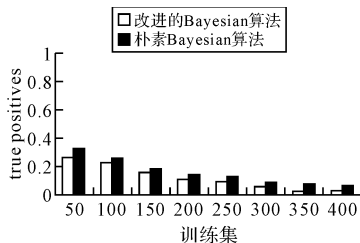


图 4 改进的 Bayesian 算法和朴素 Bayesian 算法的误诊率

从图 4 结果可见当训练集增大时,这两个算法的分类误诊率都有所下降,而由于改进的 Bayesian 算法加入了学习过程,其分类误诊率更低。

## 5 结语

随着电力系统自动化技术的迅速发展,加快对智能变电站网络故障的诊断是电力系统亟待解决的问题,文章对智能变电站网络故障的分类处理过程进行了分析,并在此基础上通过加入学习过程和特征选择标准对原有 Bayesian 分类器进行了修正,实验证明改进的 Bayesian 算法在对智能变电站网络故障进行分类时可提高分类精度并降低误诊率,达到优化分类系统性能的目的,在一定程度上提高了智能变电站的网络故障诊断能力。

### 参考文献

- [1] 丁道齐. 现代电网的发展与安全[M]. 北京:清华大学出版社,2012.  
DING Daoqi. The Development and Security of Modern Power Grid[M]. Beijing: Tsinghua University Press,

2012.

- [2] 王琨,马志欣. 网络故障诊断[M]. 西安:西安电子科技大学出版社,2011.  
WANG Kun, Ma Zhixin. Network Fault Diagnosis [M]. Xi'an: Xi'an University of Electronic Science and Technology University Press,2011.
- [3] 王之猛,孙庆彬,刘奇,等. 电网故障诊断的统一信息模型设计[J]. 电力信息化,2012(11):43-46.  
WANG Zhimeng, SUN Qingbin, LIU Qi, et al. A Unified Information Model Design of Fault Diagnosis[J]. Power Information,2012(11):43-46.
- [4] 张钧. 配电网智能故障诊断与谐波源定位研究[D]. 成都:西南交通大学,2012.  
ZHANG Jun. Distribution Network Intelligent Fault Diagnosis and Harmonic Source Localization [D]. Chengdu: Southwest Jiaotong University,2012.
- [5] Huajie Zhang, Charles X. Ling, Zhiduo Zhao. The Learnability of Naive Bayes [C]//Proceedings of the 13<sup>th</sup> Biennial Conference of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence. Lecture Notes in Computer Science,2000:432-441.
- [6] Xin-Yuan Song, Sik-Yum Lee. A tutorial on the Bayesian approach for analyzing structural equation models [J]. Journal of Mathematical Psychology,2012,56(3): 135-148.
- [7] J. Vandenplas, N. Gengler. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations[J]. Journal of Dairy Science,2012,95(3):1513-1526.
- [8] Lori A. Dalton, Edward R. Dougherty. Optimal classifiers with minimum expected error within a Bayesian framework[J]. Pattern Recognition,2012.
- [9] Ken McNaught, Andy Chan. Bayesian networks in manufacturing[J]. Journal of Manufacturing Technology Management,2011,22(6):734-747.
- [10] 杨彩虹,黄本雄. 基于贝叶斯信念网的网络流量分类与识别研究[J]. 计算机应用与软件,2011(1):216-219.  
YANG Caihong, HUANG Benxiong. The Research of Network Traffic Classification and Identification Based on Bayesian Belief Network[J]. Computer Applications and Software,2011(1):216-219.
- [11] 张增伟,吴萍. 基于朴素贝叶斯算法的改进遗传算法分类研究[J]. 计算机工程与设计,2012(2):750-753.  
ZHANG Zengwei, WU Ping. The Research of Improved Genetic Algorithm Classification Based on Naive Bayes Algorithm[J]. Computer Engineering and Design,2012(2):750-753.