

基于支持向量机的电力系统不良数据 在线检测辨识与修正

包永金

(宜宾电业局, 四川 宜宾 644000)

摘要: SVM 是数据挖掘中一种具有优良模式识别性能的新方法, 该方法具有学习速度快、全局最优和泛化能力强等优点。首先利用支持向量机回归(SVR)构建辨识遥测不良数据的模型, 在状态估计前通过比较预测值与实测值之间的差值来一次性辨识遥测不良数据。接着将状态估计后得到的标准残差作为支持向量机分类(SVC)的输入, 依靠拓扑错误的残差特性来分类辨识出拓扑错误。通过对 IEEE-30 母线的仿真分析证明了该方法的有效性, 现行状态估计器的效率及合格率可以得到很好的提高。

关键词: 不良数据; 电力系统状态估计; 检测辨识; 支持向量机

Abstract: SVM is a new method with excellent pattern recognition properties in data mining, which has the advantages of fast learning, global optimum and high generalization. Firstly, support vector machine regression is utilized to establish the identification model for the telemetric bad data, which compares the differences between the predicted values and the measured values before state estimation. Then the obtained standard residuals after state estimation are used as the input of SVC classification, and the topology error is identified based on the characteristics of these residuals. The efficiency of the proposed method is proven by the simulation analysis of IEEE-30 bus model, thus the efficiency and the percent of pass of the existing state estimators can be highly improved.

Key words: bad data; power system state estimation; detection and identification; support vector machine

中图分类号: TM744 文献标志码: A 文章编号: 1003-6954(2013)01-0059-05

0 引言

在能量管理系统(energy management system, EMS)中, 电力系统状态估计负责处理由 SCADA 采集到的包含噪声的冗余数据以便为诸如: 经济分配、安全分析等应用软件提供精确的实时数据。电力系统运行人员通过分析这些软件的结果来采取决策。如果所使用的实时数据库中包含错误数据, 这些应用软件的结果可能毫无意义, 根据这些结果采取的决策也很可能对电力系统的安全运行产生危险。故而不良数据辨识是状态估计中一项十分重要的任务。

现行电力系统中有很多状态估计的方法^[1]。在正常运行(噪声数据主要来自量测不准确)的条件下, 这些方法通常都会获得很好的结果。不过当出现较大的量测误差和拓扑结构错误的时候, 情况可能就不一样了。不良数据辨识的方法正是基于这些问题而提出的。大多数的方法都是在进行状态估

计计算后, 利用量测残差进行假设检验的统计理论方法^[2-4]。在很多情况下这些方法都表现得很好, 不过仍然存在着一些缺点: 难于辨识多相关不良数据, 对于关键量测点的不良数据和包含不相关支路时的拓扑结构错误无法辨识。而且这些方法很难同时处理遥测和遥信不良数据。针对以上这些问题, 又提出了基于数据挖掘的方法。主要有基于神经网络(ANN)^[5-7]、基于模糊理论和聚类分析^[8]及基于间歇统计(GSA)^[9]等一些方法。基于神经网络的方法大多采用典型工况的正确量测数据作为训练样本来构造了一个神经元网络, 用估计前滤波来辨识各种形式的不良数据。但由于神经网络算法的固有特性, 不能从根本上避免训练过饱和及陷入局部最小值的问题, 且随着时间的推移, 训练样本集应该动态更新, 但此类文献也没有提出对样本集更新的动态处理措施。

支持向量机(support vector machine, SVM)是数据挖掘中一种具有优良模式识别性能的新方法。利

用 SVM 回归和分类算法分别建立起了辨识遥测和遥信不良数据的模型: 回归模型采用 SVM 非线性回归算法对各种运行情况下的正常遥测数据进行曲线拟合(即训练), 使模型具有对遥测数据的一步预测能力; 针对状态估计中拓扑结构错误的特点(一条支路上的拓扑结构错误会在该支路周围的支路潮流和母线注入功率上产生较大的残差), 应用支持向量机分类算法构建起了辨识遥信不良数据的分类模型。值得一提的是上述两种模型都可进行在线训练, SVM 在线学习方法能对每次迭代过程中增加的样本进行学习, 利用前一次迭代的运算结果, 减少计算复杂度, 实现在较小时间代价下的新样本学习。模型构建好后首先运用回归模型对实际遥测数据进行预测, 通过比较预测值与实测值之间的残差来一次性辨识量测数据中的遥测不良数据。一旦检测出不良数据就可将其替换成合理的预测值以避免出现系统的不可观测性。遥信不良数据辨识时, 先对消除不良数据影响后的量测数据进行状态估计, 算出量测残差, 再将量测残差输入分类模型, 运用 SVM 分类的良好泛化能力辨识出遥信错误。通过对 IEEE-30 节点模型的仿真结果表明, 所提出的方法对电力系统不良数据检测的效率和准确性都有较大的提高, 说明了方法的有效性。

1 支持向量机算法介绍

1.1 支持向量机回归算法

支持向量机(SVM)是由 Vapnik^[10]最早提出的一种统计学习方法。目前国内外所研究的支持向量机回归算法主要是多输入、单输出, 即每次只能对一个特征量进行回归预测, 对于多输出支持向量机的研究并不多^[11-12], 且算法的效果也不太理想。鉴于此, 采用多个单输出支持向量机来实现多输出问题。

将支持向量机用于解决回归问题即支持向量回归(support vector regression, SVR)。SVR 用来解决回归预测的基本思想是: 通过一个非线性映射 φ 将输入空间中的数据映射到高位特征空间 G 中, 并在 G 空间中进行线性回归。假设有这样的训练数据 $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset x \times R$, 这里 x 表示输入样本空间(如: $x = R^d$)。SVR 通过引入损失函数来解决回归问题, 采用式(1)来计算函数。

$$y = f(x) = (\omega \cdot \varphi(x)) + b \quad \varphi: R^m \rightarrow \omega \in G \quad (1)$$

对优化目标取极值

$$\min Q = \frac{1}{2} \|W\|^2 + c \sum_{i=1}^n (\xi_i^* + \xi_i) \quad (2)$$

约束条件为

$$\begin{cases} y_i - (\omega \cdot \varphi(x_i)) - b \leq \varepsilon + \xi_i^* \\ (\omega \cdot \varphi(x_i)) + b - y_i = \varepsilon + \xi_i \\ \xi_i^*, \xi_i \geq 0 \quad i = 1, 2, \dots, n \end{cases} \quad (3)$$

式中 c 为惩罚因子; ξ_i^* 、 ξ_i 为松弛因子; ε 为损失函数。引入拉格朗日因子 a_i 和 a_i^* , 把凸优化问题转化为最大二次型, 用核函数来代替内积运算, 则优化目标形式如下。

$$\max w(a, a^*) = \sum_{i=1}^n y_i (a_i - a_i^*) - \frac{1}{2} \sum_{i,j=1}^n (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) \quad (4)$$

约束条件为

$$\begin{cases} \sum_{i=1}^n a_i = \sum_{i=1}^n a_i^* \\ 0 \leq a_i \leq c \quad i = 1, 2, \dots, n \\ 0 \leq a_i^* \leq c \quad i = 1, 2, \dots, n \end{cases} \quad (5)$$

对优化目标取极值和引入拉格朗日因子 a_i 和 a_i^* 函数可表示为

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x_j) + b \quad (6)$$

高维特征空间中线性问题的内积运算可以用核函数来代替, 即

$$K(x_i, x_j) = \phi(x_i) \phi(x_j) \quad (7)$$

常用的核函数有: 径向基函数、多项式函数、Sigmoid 函数、线性函数等。

目前, SVR 的训练方式有两种: 批量训练与在线训练。其问题表述和最优化求解的原理相同, 不同点在于: 在线支持向量机不断将新样本添加到训练集, 通过在线调整模型参数, 实现预测模型的不断变化, 而不是批量式的一次训练所有样本。这样不但能提高其训练时间, 而且能提高模型的适应能力, 使系统的运行情况发生变化时, 模型能相应地发生变化。增量式 SVR 的基本思想如图 1 所示。

这里采用增量学习的 SVR 来训练数据, 具体算法可以查看文献[13]。

1.2 支持向量机分类算法

支持向量机分类(support vector classification, SVC)与 SVR 算法很相似, 具体算法可查看相关文献[14], 在此不再赘述。采用在状态估计前后分别进行不良数据辨识的方式, 在状态估计后采用 SVC

主要对遥信不良数据进行分类辨识。

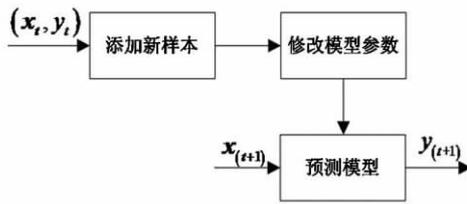


图1 增量式SVR的基本思想

一般来说,当系统某一元素(输电线或母线)发生拓扑错误或量测量突变时,不良数据只会出现在围绕该元素的量测点上,故而对每一个系统元素都采用一个SVC分类器来具体辨识出不良数据的类别。当出现不良数据时,只会调用该不良数据所对应的SVC分类器,以减少计算量。对任意支路 $s-r$ 来说,SVC的输入是该支路潮流和支路两端母线上注入功率的标准残差;当SVC用于辨识母线结构错误时,其输入为母线电压和母线上各支路潮流的标准残差。

状态估计后的不良数据分为支路拓扑错误、母线拓扑错误以及遥测错误。遥测错误是由于状态估计前辨识遥测不良数据时的漏判或误判造成,此外系统中某一元素发生的拓扑错误也可能在与之相邻的元素上产生较大的残差,此类异常数据点称作非不良数据点。在训练阶段,支路拓扑错误所对应的SVC输出选为-1,遥测错误的输出选为1,而非不良数据点的输出选为0。至于母线拓扑错误的输出则要根据具体母线可能发生的拓扑错误而定,有几种拓扑错误就选几个输出值。

2 基于SVM的不良数据检测与辨识

2.1 检测与辨识的模型

应用支持向量机来辨识电力系统不良数据的模型如图2所示。下面分别对模型各部分加以说明。

第1步:计算遥测数据的估计值和平方误差

由经过良好训练的SVR模型可得到原始遥测数据的估计值。初始SVR模型可通过离线训练获得,训练集可由各种运行条件下的历史正常遥测数据或对网络进行离线仿真获取。进行在线预测时,每采到一次遥测数据就可由回归模型获得该组遥测数据的估计值,然后再将该组遥测数据添加到训练数据中进行SVR增量学习。由于SVR增量学习不

需要从头进行训练,所以训练的时间很短,满足在线辨识的需要。获得估计值之后,就可由量测值和估计值计算平方误差。

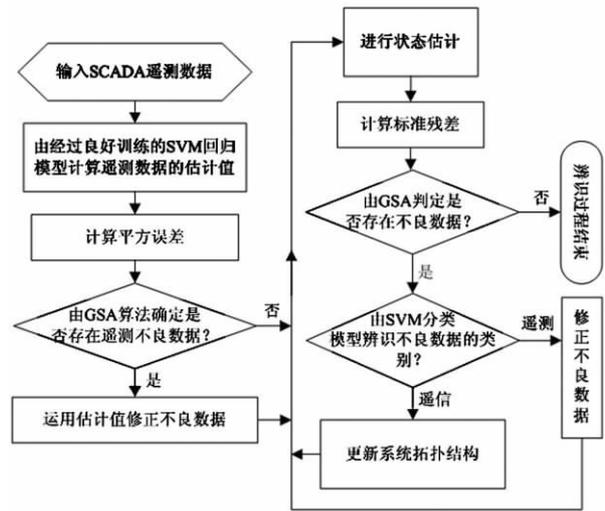


图2 不良数据辨识模型

第2步:辨识和修正遥测不良数据

为避免阈值选取主观性对辨识结果带来的影响,这里采用GSA算法来辨识遥测数据中的不良数据。GSA方法是一种强化聚类效果的数据挖掘算法,它可以估计数据集最佳的聚类个数。在电力系统不良数据辨识中,可以将良好数据和不良数据所在的聚类准确地区分进而检测和辨识不良数据^[9,13]。据此将上一步获得的平方误差进行聚类,如果聚类个数大于1,则表明有不良数据,接着辨识出不良数据并采用估计值进行修正。

第3步:进行状态估计

在消除遥测不良数据的影响后就可利用状态估计器对系统进行状态估计,这里采用加权最小二乘法状态估计。进行状态估计的目的是为了获取系统最可能的运行状态和为辨识遥信不良数据做准备。

第4步:计算标准残差并判别是否存在不良数据

由上一步计算得到的状态估计值和SCADA量测值就可以计算出标准残差,判别是否存在不良数据仍然采用GSA算法,判别的原则同辨识遥测不良数据时一样。如果不存在不良数据则状态估计结束,输出估计结果。依然存在不良数据时,此时的不良数据有两种情况:一是由于遥信数据错误产生,此类不良数据即遥信不良数据;此外在辨识遥测不良数据时漏判和相邻支路的遥信错误也会导致本支路残差过大而出现不良数据,此类的不良数据称为量

测量突变。这里采用 SVM 分类算法来辨识上述不良数据的类别。

第 5 步: 由 SVM 分类模型辨识不良数据的类别

由于发生拓扑错误时的残差特性与发生遥测不良数据时的残差特性有很明显的区别,所以在获取训练数据集时不需要模拟系统所有运行状态。进行在线辨识时,对上一步获得的每一个不良数据调用它所对应的分类器,由分类器就可获得该不良数据的类型。对于遥测不良数据可采用其估计值进行修正或者直接丢弃;对于拓扑结构错误则更新系统拓扑结构。接着转入下一次状态估计,重复第 3 步到第 5 步直到不再出现不良数据为止。

2.2 算例分析

通过 IEEE-30 节点系统为例来验证方法的可行性。在缺乏真实量测数据的情况下,假设系统的运行情况如下:将原系统中给定的负荷作为最大负荷需求,最小负荷假定为最大负荷的 20%,用一个负荷取值区域模型来模拟母线上的负荷波动。这样的一个例子如图 3 所示,图 3 中一条母线上的最大和最小负荷通过一个 6 个时间段的取值区域来表示。母线上所有可能的负荷波动都包含在阴影区域以内。对母线 2 上所接的发电机采用同样的处理方法,母线 1 作为平衡节点,变压器的变比选取在 0.95 ~ 1.05 之间。

根据上述运行情况,通过仿真计算获取了 288 组运行数据,其中三分之二用于训练 SVM 模型,其余用于测试。模型将添加噪声后的数据作为输入数据,将潮流计算值作为目标值来进行训练。对于大型系统不良数据辨识实时性的要求,在训练和测试阶段可将量测值分为几组分别计算,即每组数据都可以有它相应的模型。这样就可将每次采集到的量测数据送入其对应的分组模型进行不良数据的辨识。这里采用文献 [7] 所述将数据按电压等级分为了两组。

情况 1: 遥测不良数据

假设第 35 组测试数据中出现了 6 个不良数据,分别是 P_{6-9} 、 P_{6-10} 、 Q_{19-20} 、 Q_{8-28} 、 Q_2 和 V_{15} 。不良数据与正常量测值的偏差在 $\pm(20-100)$ 标准差之间。此种情况下,GSA 算法正确地区分出了上述不良数据。不良数据的情况如表 1 所示。

表 1 遥测不良数据辨识结果

不良数据	真实值	量测值	估计值	估计误差
P_{6-9}	0.221 7	0.530 2	0.226 5	0.004 8
P_{6-10}	0.127 0	0.301 5	0.129 8	0.002 8
Q_{19-20}	-0.023 8	0.149 2	-0.023 0	0.000 8
Q_{8-28}	-0.028 9	0.028 1	-0.027 9	0.001 0
Q_2	0.396 7	-0.085 1	0.400 8	0.004 1
V_{15}	1.053 0	1.398 7	1.060 0	0.007 0

情况 2: 支路遥信变位错误

假设在第 72 组数据的时刻收到支路 4-12 错误的遥信变位信息。支路 4-12 开关实际为闭合状态,但收到的遥信信息为开关断开。根据前面所提方法,在进行状态估计后,只有当出现不良数据时才会调用不良数据所对应的 SVM 分类模型对不良数据进行辨识,且每个系统元素都对应着一个 SVM 分类模型。由 GSA 算法得出的一组可疑数据如表 2 所示。从分类结果可以得出 SVM 分类模型成功辨识出了支路 4-12 的遥信变位错误。

表 2 支路遥信错误辨识

不良数据集{4-12;9-11;6-10}	
SVM 分类模型	分类输出
SVM_{4-12}	-1
SVM_{9-11}	0
SVM_{6-10}	0

值得注意的是当辨识出遥信错误后就可直接更新系统的拓扑结构,无需再进行其他的辨识。

情况 3: 母线分裂型错误

假定在 90 组数据的时刻母线 15 发生分裂,但并没有收到相应的遥信信息。分裂前后的情况如图 3 所示。

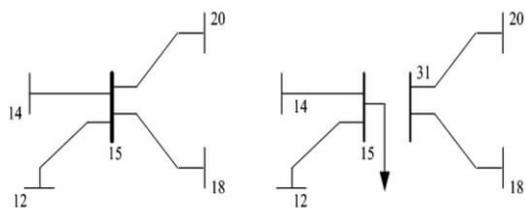


图 3 母线 15 分裂前后模型

表 3 列出了可疑数据集,在训练阶段将此类型拓扑错误对应的 SVC 输出选为 2。由分类结果可知 SVM 模型成功辨识出了母线 15 的分裂错误。

表 3 母线分裂型错误辨识

不良数据集{ 15 - 18; 23 - 24; 23 }	
SVM 分类模型	分类输出
SVM ₁₅₋₁₈	2
SVM ₂₃₋₂₄	0
SVM ₂₃	0

3 结 论

以上提出了一种在电力系统实时状态估计中辨识不良数据的新方法。与传统状态估计中的假设检验方法相比,该方法可以很好地避免残差污染和残差淹没,且能克服基于神经网络的方法中神经网络训练过饱和和容易陷入局部最小值的问题。仿真分析证明了本方法的有效性,将此方法与现行电力系统状态估计结合可有效实现不良数据的辨识与修正。

参考文献

[1] ALi Abur ,Antonin Gómez Expósito. Power System State Estimation: Theory and Implementation [M]. New York: Marcel Dekker ,2004.

[2] L. Mili ,Th Van Cutsem and M. Ribbens Pavella. Hypothesis Testing Identification: A New Method for Bad Data Analysis in Power System State Estimation [J]. IEEE Trans. on Power Apparatus and Systems ,1984 ,103 (11) : 3239 - 3252.

[3] L. Mili ,Th. Van Cutsem. Implementation of HTI Method in Power System State Estimation [J]. IEEE Trans. on Power Systems ,1988 ,3(3) : 887 - 889.

[4] Salehfar H ,Zhao R. A Neural Network Preestimation Filter for Bad Data Detection and Identification in Power System State Estimation [J]. Electric Power System Research. 1995 ,34(9) : 127 - 134.

[5] J. C. S. Souza , A. M. Leite da Silva , A. P. Alves da Silva. Online Topology Determination and Bad Data Suppression in Power System Operation Using Artificial Neural Networks [J]. IEEE Transactions on Power Sys-

tems ,1998 ,13(3) : 796 - 803.

[6] S. P. Teeuwesen. Neural Network Based Multi - dimensional Feature Forecasting for Bad Data Detection and Feature Restoration in Power System [C]. IEEE Power Engineering Society General Meeting ,2006: 18 - 22.

[7] D. Singh ,J. P. Pandey ,D. S. Chauhan. Topology Identification ,Bad Data Processing ,and State Estimation Using Fuzzy Pattern Matching [J]. IEEE Transactions on Power Delivery ,2005 ,20(3) .

[8] Huang SJ ,Lin JM. Enhancement of Power System Data Debugging Using GSA - based Data Mining Technique. [J]. IEEE Transactions on Power Systems 2002 ,17(4) : 1022 - 1029.

[9] Vapnik V. The Nature of Statistical Learning Theory [M]. New York: Springer ,1995: 138 - 145.

[10] F. Pérez - Cruz ,G. Camps ,E. Soria ,J. Pérez ,A. R. Figueiras Vidal ,and A. Artés - Rodríguez. Multi - dimensional Function Approximation and Regression Estimation [C]. in Proc. ICANN ,Madrid ,Spain 2002.

[11] M. P. Sánchez - Fernández ,M. de Prado - Cumplido ,J. Arenas - García ,and F. Pérez - Cruz. SVM Multiregression for Nonlinear Channel Estimation in Multiple - input Multiple - output Systems [J]. IEEE Trans. Signal Process 2004 ,52(8) : 2298 - 2307.

[12] Parrella F. Online Support Vector Regression [D]. Master's Thesis ,University of Genoa ,2007.

[13] Chih - Wei Hsu , Chih - Chung Chang , and Chih - Jen Lin. A Practical Guide to Support Vector Classification. Technical Report , Department of Computer Science and Information Engineering , National Taiwan University , Taipei , 2003. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

[14] 吴军基 杨伟 葛成 等. 基于 GSA 的肘形判别用于电力系统不良数据辨识 [J]. 中国电机工程学报 , 2006 ,26(22) : 23 - 28.

作者简介:

包永金(1975) ,男,宜宾电业局工程师,主要从事输配电运行工作。

(收稿日期:2012 - 08 - 10)

发展风电 保护环境