「水电机组」

基于数据挖掘的水电机组振动故障诊断研究

文/国网新源控股有限公司本部　孙　勇
安徽省响洪甸抽水蓄能发电有限公司　金守迁
华东琅琊山抽水蓄能有限责任公司　王　磊

提出一种利用聚类分析、离差标准分法、神经网络等方法建立机组综合评价模型，用来解决水电机组故障诊断问题。该算法应用到某些电站机组的故障诊断中，结果表明，该方法能够对故障做出准确判别，分离各种故障类型。由于需要存储的数据比较少，速度快，非常适于水电机组故障诊断中大规模图形序列挖掘。

数据挖掘技术是一门主要由统计学和人工智能学组成的交叉学科，它的主要任务是从海量的数据集中提取隐藏有用的信息。目前，数据挖掘技术已被广泛应用于电信、金融、网络安全、天气预报等领域。在水电机组运行过程中，振动信号包含了很多机组运行信息。据统计，水电机组约有80%的故障或事故都在振动信号中有所反映，而在机组在线监测过程中，机组振动信息数据量非常大，并且随着监测时间的延长而增长，信息数据成为大规模海量数据集，这就给信号的处理分析带来一定的难度。一般的方法是通过典型标准故障样本对分类器进行训练，但是个体机组之间则存在着差异，这会影响到分类的精度。利用聚类分析、离差标准分法、神经网络等方法建立机组综合评价模型，并利用可视化技术对评价结合进行多维立体展示，让工作人员快速定位问题，找出问题机组，从而对整个机组的运行状态充分把握，提升问题异动分析的精确度，通过实践证明本文所提内容切实可行，能够对机组状态进行立体展示，促进整体机组运维与规划。

相关算法的概念

1. 聚类算法

聚类算法又称群集分析，常用于将大量数据按照一定规则分为不同类别，其与分类算法的差异在于，聚类算法可在非监督模式下处理数据，不需要人为输入数据标签。非监督式，即是聚类分析与分类和回归分析的区别所在，聚类不需要人为的输入标签，尽管某些聚类算法需要设定初始划分方法或者根据输入参数确定聚类集合个数，但在分析过程中，算法无须人为输入分类标准。因此，聚类分析在很多时候用于大型数据库的分类预处理，当然聚类分析也常作为独立分析数据的工具。

2. 离差标准分法

离差－标准分是离差标准化（也称Min－max标准化，0－1标准化，归一化）的计算结果，是对原始数据的线性变换，使结果落到[0，1]区间上。其中max（1）为样本数据的最大值，min（0）为样本数据的最小值。在某些比较和评价的指标处理中经常会用到，去除数据的单位限制，将其转化为无量纲的纯数值，便于不同单位或量

级的指标能够进行比较和加权。

数据特征综合评分是使用各台机组的标准值对每个振摆测点历史5年数据进行综合评分。计算方法和过程如下：

中位数离差标准分是中位数与国标规定振摆最大值的比值。

中位数离差标准分＝中位数/国标规定振摆最大值

集中分布区间离差标准分是区间上限与区间下限的差与国标规定振摆最大值的比值。

集中分布区间离差标准分＝(区间上限－区间下限)/ 国标规定振摆最大值

机组部位综合评分为中位数离差标准分、50%集中分布区间离差标准分、95%集中分布区间离差标准分之和。机组综合评分为该机组各部位发电（额定负荷）工况和抽水工况综合评分的平均值。电站综合评分为该电站各机组综合评分的平均值。

3. 神经网络

BP网络是典型的多层网络，分为输入层、隐含层和输出层，层与层之间采用完全互连方式，同一层单元间不存在相互连接。BP网络的第一层即输入层，接收外界输入，最后一层即输出层，产生输出。在输入层和输出层之间，有若干个隐含层。在各层之间，只有相邻层的神经元之间存在联系。对输入X进行某种函数运算，即得到神经元的输出Y。

BP算法的具体步骤框图如图1所示。

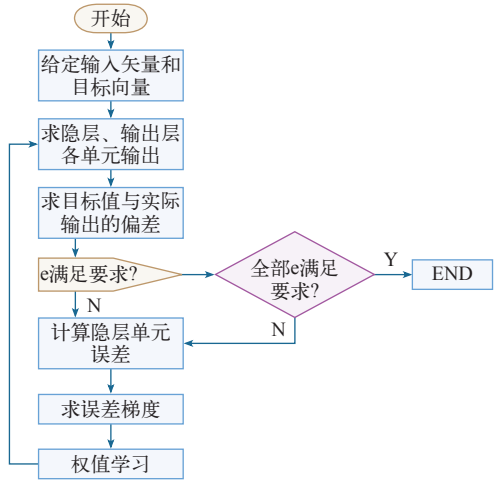

图1　神经网络算法流程图

故障诊断方案设计

结合抽水蓄能电站机组的管理与其在实际业务中的指导意义，可将机组的综合评价结果应用于运维管理、客户服务及外部因素等方面。在每个实际要应用的方面，提取相应的评价指标。比如，在运维管理方面，可提取油温、负荷、振动、摆度、电流、电压、水头及有功功率等八个指标，在此基础上，应用大数据的多指标综合评价方法合理分析实际各个机组的“过去”“现在”及“将来”。

1）根据机组的有功功率特征进行分类，将有功功率特征相似的机组分为一类。本文的研究对象包含28个抽水蓄能电站106个机组，包含的对象较多，本文将在数据清洗的基础上，尝试利用聚类算法将机组依据其实际使用功率的特征进行分类，从而实现对于机组差异化的管理。

2）基于分类后的每一类机组，提取相应的各类指标，建立一套完整的能充分覆盖抽水蓄能电站运维管理方面的指标体系。机组的评价指标应具有反应机组充足性、稳定性、安全性。基于以上认识，构建抽水蓄能电站的综合评价指标体系应注意以下三项原则：综合性、科学性、操作性。本文结合实际业务，提取的指标主要包含：油温、负荷、振动、摆度、电流、电压、水头及有功功率等八个指标等。

3）基于2）中构建的抽水蓄能电站机组的综合评价指标体系，分别从客观和主观两个角度计算指标体系中各指标的权重。客观方面，应用离差标准分法分别计算每类机组不同指标的客观权重，该方法适应系统采集的功率、电压、电流及功率等数据，能够完全从客观的角度反映数据的真实性；主观方面，基于专家评分法，计算各个指标的主观权重，该方法需要集合专家长期的业务经验，另外，该方法的使用也可以将定性问题转化为定量问题的计算，能够帮助客观问题的实现。

4）综合主客观权重，从而更加合理地反映上述建立的指标体系中各指标对综合评价模型的影响权重。本文采用最小二乘法，优化主客观权重，并基于优化之后的权重结合各指标的得分计算最终的综合评价值。在评价模型选择上，首先根据单个指标的数据分布特点，采用方差贡献度方法，形成客观的评价模型，再辅以专家调整，尽可能降低纯粹主观因素的影响。

5）基于4）得到的综合评价，可以积累一部分历史数据，当积累的数据达到一定量时，就可

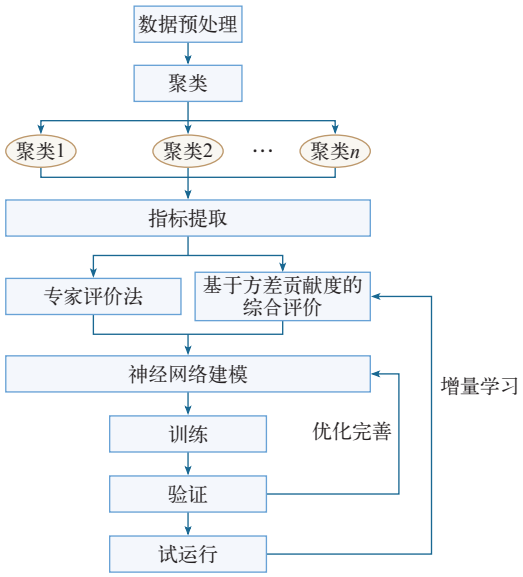

图2　综合评价流程图

以进一步借助神经网络模型进行优化，利用神经网络的算法对机组的各指标及综合评价值进行训练并预测。

故障诊断方案在机组实际应用与效果分析

抽水蓄能电站综合评价分析，主要包括综合聚类分析、评价模型和神经网络三部分，基于以下的技术路线进行分析。

1. 归一化处理

从底层信息采集系统采集到的数据源中对数据进行预处理，然后做聚类操作。从数据库中提取的有功功率数据是来自所有抽水蓄能电站机组的，不同机组之间功率值数据差异很大，直接聚类会导致数据量级的差异掩盖实际机组特征的差异，一般在聚类操作之前会采用公式1将原功率数据进行归一化处理。

功率数据归一化

$$\bar{p}_{ij}=\frac{p_{ij}-\min\limits_{j}(p_{ij})}{\max\limits_{j}(p_{ij})-\min\limits_{j}(p_{ij})} \quad (1)$$

2. 指标提取

根据聚类结果，提取其中一类机组的运维数据，在数据清洗的基础上，分别提取对应的各类指标，如表1所示。

表1　指标提取

运维管理							
油温	负荷	振动	摆度	电流	电压	水头	有功功率

表2　指标获取图

油温	负荷	振动	…	有功功率
0.242 44	0.675 88	0.710 96	…	0.922 85
0.563 86	0.463 78	0.638 19	…	0.943 25
0.945 53	0.734 60	0.941 53	…	0.927 80
0.145 46	0.321 20	0.845 24	…	0.997 6

3.模型构建

（1）要实现综合评价在抽水蓄能电站中的应用

首先根据评价结果应用的方向，如：运维管理方面，那么先提取油温、负荷、振动、摆度、电流、电压、水头及有功功率等指标，并将所有指标转换为向好型指标值。

对应每个指标，分别计算每个指标的方差，记为：$v=(v_1,v_2,\cdots,v_8)$。进一步，将计算得到的方差进行归一化：$\lambda_i=v_i\Big/\sum_{j=1}^{8}v_j(i=1,2,\cdots,8)$，得到每个指标的贡献度即权重。借助基于方差贡献度的综合评价方法，对各类机组进行评价，$\mathrm{Value}=\sum_{i=1}^{8}\lambda_i p_{ji}$。

其中，p_{ji}代表第j个机组的第i个指标值。

基于上述的评价结果，可进一步通过神经网络来学习，因为神经网络算法在拟合和预测方面具有较高的准确性。借助神经网络的深度学习，就可以实现模型的不断更新，也能保证预测模型拟合误差不断减小，同时使得预测更加准确。

（2）基于神经网络算法的综合评价模型

使用神经网络算法是构建综合评价模型的核心及创新点，神经网络算法的学习需要收集来自实际业务的输入及输出，输入即上述（1）中提取到的各类指标，输出是对于机组的历史评价数据，对应最后评价分值越高则该机组在某一方面就越好，比如机组的运维管理得分最高，那么该机组在运维管理方面的各方面指标均是最优，这就要保证所有的指标都要转换为向好型指标，即指标值越高，对应状态越好，这和实际业务中向小型指标相反，但只要做一个变换即可。

对于用神经网络来对机组进行综合评价，就需要考虑影响神经网络的输入与输出，输入即各个指标的得分，输出为综合评价 得分。而当前对于机组的综合评价，纯属创新，历史上并没有相应的综合评价值，这就需要利用实际数据对综合评价值进行一些积累，在此基础上才可以利用神经网络算法对机组进行综合评价。

本小节分三个部分：

1）指标转换，将向小型指标、中间型指标均转换为向好型指标。

2）构建综合评价模型，构建综合评价的根据提取的指标，构建综合评价的历史数据。

3）神经网络综合评价，利用上述构建的综合评价模型，训练神经网络并进行预测。

4）指标转换。

根据实际业务，在运维管理方面，八个指标的所得值均属于[0，1]之间，这里就不再需要进行归一化的转换，直接进行向好型的转换即可。

根据基于方差贡献度的综合评价算法时提到的指标同质化方法，分别将上述三个指标进行转换：

峰谷差率：该指标的计算实际和负荷，计算求和为1，所以在这里需要进行特殊转换，以避免转换后的指标与最小负荷率重叠。

$$fgcl = 1 - \arctan(fgcl) \tag{2}$$

振动和摆度：一般来说，振动和摆度值越小越好，这里需要按以下公式进行转换

$$sxbphl = 1 - sxbphl \tag{3}$$

配变最大负载率：根据实际业务，机组的负荷在0%～30%之间属于轻载，30%～80%之间是最好的，80%～100%属于重载，100%～以上属于过载，所以需要按照以下的公式进行转换

$$pbzdfzl = \begin{cases} 1 - \dfrac{0.3 - pbzdfzl}{0.3} & pbzdfzl < 0.3 \\ 1 & 0.3 \leqslant pbzdfzl \leqslant 0.8 \\ 1 - \dfrac{pbzdfzl - 0.8}{0.3} & pbzdfzl > 0.8 \end{cases} \tag{4}$$

至此，运维管理的八个指标均属于向好型指标。

（3）综合评价模型

本文的综合评价应用方向主要包括运维管理方面，下面就机组的运维管理方面说明综合评价模型的应用。基于数据处理小节提取某一类别的评价指标，将所有指标按照前述的转化为向好型指标后，如表3所示。

1）基于方差贡献度的综合评价方法计算指标权重：根据得到的各个指标值，在日期上面和聚类结果进行关联，在此基础上，分别计算每一类机组各个指标的方差，如表4所示。

在此基础上，根据公式：$b_j = \dfrac{v_j}{\sum_{i=1}^{m} v_i}$（$v_j$是第$j$个指标的方差），计算不同类别下每个指标的权重，如表5所示。

2）基于专家评分法的方法计算各指标权重，

表3　向好型指标值

日期	油温	负荷	…	振动
2015－01－01	0.893 78	0.106 22	…	0.786 34
2015－01－02	0.783 62	0.216 38	…	0.713 62
⋮	⋮	⋮	⋮	⋮
2015－12－31	0.489 47	0.510 53		0.373 61
2015－01－01	0.337 36	0.662 64		0.746 44
2015－01－02	0.333 84	0.666 16		0.738 64
⋮	⋮	⋮	⋮	⋮
2015－12－31	0.383 75	0.616 25		0.728 64
⋮	⋮	⋮	⋮	⋮
2015－12－31	0.267 22	0.732 78		0.937 36
Var()	0.018 662	0.024 12		0.012 148
$Var_i / \sum_i Var_i$	0.059 51	0.125 67		0.280 44

表4　各类指标的方差

类别	油温	负荷	振动	摆度	电流	电压	水头	有功功率
1	0.018 662	0.024 12	0.039 226	0.123 744	0.075 503	0.079 022	0.099 886	0.012 148
2	0.008 985	0.009 941	0.019 878	0.072 063	0.057 498	0.074 5	0.101 8	0.004 934
3	0.020 561	0.013 125	0.026 203	0.031 681	0.161 727	0.122 965	0.114 635	0.001 783
4	0.027 43	0.037 561	0.047 268	0.115 132	0.101 502	0.153 864	0.197 877	0.014 846

表5　各类指标的权重

类别	油温	负荷	振动	摆度	电流	电压	水头	有功功率
1	0.039 511	0.051 068	0.083 051	0.261 997	0.159 859	0.167 309	0.211 485	0.025 721
2	0.025 701	0.028 435	0.056 858	0.206 131	0.164 469	0.213 101	0.291 191	0.014 114
3	0.041 733	0.026 641	0.053 184	0.064 304	0.328 259	0.249 585	0.232 675	0.003 62
4	0.039 441	0.054 008	0.067 964	0.165 543	0.145 945	0.221 234	0.284 518	0.021 347

首先需要收集各个专家针对个类别不同指标的打分，如果专家对各类别的打分情况一致，那么就可以采用同一个专家的打分列表，如表6所示。

根据上述各专家对不同指标的打分情况，根据公式 $P(X_i)=\sum_{j=1}^{k}P_{ji}/k$，（$i=1,2,\cdots,n$），计算每个指标的平均打分（9，9，9，10，9，9，10，9）；再结合公式：$a_i=\frac{P(X_i)}{\sum_{j=1}^{n}P(X_j)}$，（$i=1,2,\cdots,n$），计算可得每个指标的主观权重为：（9/74，9/74，9/74，10/74，9/74，9/74，10/74，9/74）。

3）结合专家评分法和基于方差贡献度的综合评价方法，利用最小二乘法重新计算每类机组各个指标的权重。

$$\begin{cases}\min f(c)=\sum_{i=1}^{m}\sum_{j=1}^{n}\{[(a_j-c_j)]^2+[(b_j-c_j)]^2\}\\ \text{s.t.}\quad \sum_{j=1}^{n}c_j=1\\ c_j\geqslant 0,\quad j=1,2,\cdots,n\end{cases}\tag{5}$$

计算出综合权重，式中，a_j为求得的主观权重；b_j为求得的客观权重，如表7所示。

表6　专家打分表

专家	油温	负荷	振动	摆度	电流	电压	水头	有功功率
专家1	9	9	9	10	9	9	10	9
专家2	9	9	9	10	9	9	10	10
专家3	9	9	9	10	9	9	10	10
专家4	9	9	9	10	9	9	10	9
专家5	9	9	9	10	9	9	10	7
平均得分	9	9	9	10	9	9	10	9
权重	9/74	9/74	9/74	10/74	9/74	9/74	10/74	9/74

表7　综合权重值表

类别	油温	负荷	振动	摆度	电流	电压	水头	有功功率
1	0.049 511	0.061 068	0.093 051	0.251 997	0.169 859	0.157 309	0.201 485	0.015 721
2	0.035 701	0.038 435	0.066 858	0.196 131	0.154 469	0.203 101	0.281 191	0.024 114
3	0.041 733	0.036 641	0.063 184	0.064 304	0.318 259	0.239 585	0.222 675	0.013 62
4	0.039 441	0.064 008	0.067 964	0.155 543	0.145 945	0.211 234	0.274 518	0.041 347

根据综合权重值即可计算每个机组的综合得分。

需要注意的是：这里计算的所有权重均是根据当前采集日期范围的数据进行计算的，当采集到新的日期数据时，可以重新根据上述综合评价模型构建的方法进行计算，求得新的权重。

（4）评价模型优化

人工神经网络以突出的优点一直被广泛应用于各类工程当中，其优势主要包含：①可以充分逼近任意复杂的非线性关系。②所有定量或定性的信息都等势分布储存于网络内的各神经元，故有很强的鲁棒性和容错性。③采用并行分布处理方法，使得快速进行大量运算成为可能。④可学习和自适应不知道或不确定的系统。⑤能够同时处理定量、定性知识。本文研究就基于神经网络做模型的拟合。

本节的前面步骤，已基本完成神经网络所需集合的构造，下面就结合上述操作步骤，来实践对于神经网络模型的构建。主要采用经典的BP神经网络进行配网状态的学习，根据要形成的神经网络模型，有两个参数需要在实际操作中进行确

认，一个是隐藏层神经元的个数，另一个是训练神经网络需要迭代的次数。按照这个要求，我们将以如下思路进行模拟。

1）根据7∶3的比例，提取70%的历史数据构成训练集来构建并训练神经网络，这里面包含两个步骤：一是确认隐藏层包含神经元的个数，二是训练过程中的迭代次数。

2）取最后30%的数据，利用上一步构建的神经网络进行预测。

a）隐层最佳神经元个数的选取

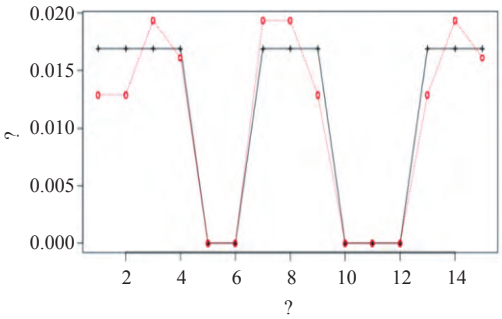

图3　隐层神经元个数选择图

b）最佳迭代次数的选取

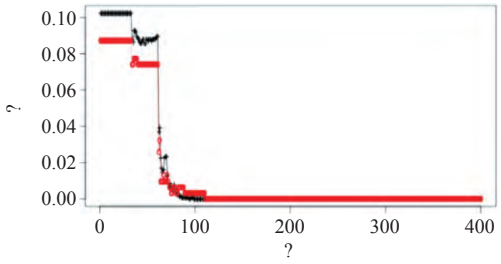

图4　神经网络最佳迭代次数选择图

经过隐层神经元个数及最佳迭代次数的选取，能够得到如图5所示的人工神经网络模型。

结束语

抽蓄电站的综合评价分析模型的输出结果即为各抽蓄电站机组的综合评价值，对于评价值较低的机组进一步进行原因分析，指出导致机组状态较差的具体原因及改进措施。

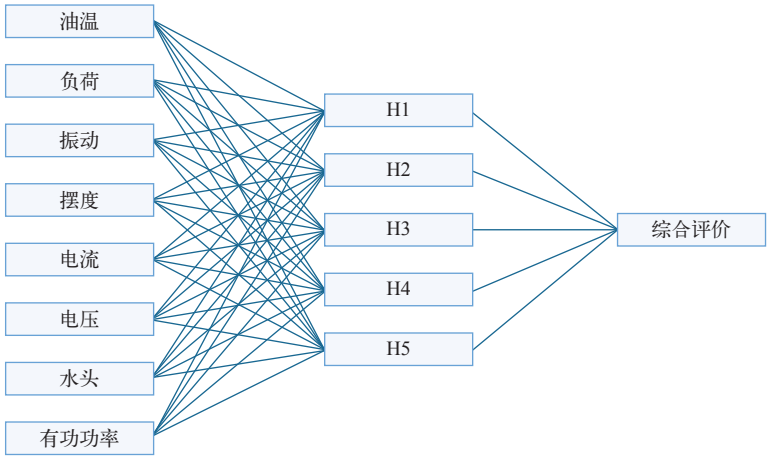

图5　综合评价模型神经网络结构图

按照最初的聚类结果，将不同类别的机组进行展示，结合基于方差贡献度和专家评分法的综合评价模型，避免了单一从基于方差贡献度的方法获取权重的客观评价方法，也避免了单一从专家评分法获取权重的主观评价方法，而是基于最小二乘法得出综合主客观方面的因素计算各指标的权重，这样的方法在结果上更加贴合实际。

另一方面，本次综合评价完全是对机组综合评价管理方面的一个创新，之前并没有形成相关的历史数据，那么在此，就可以借助综合评价方面对机组的评价构建一部分历史数据，数据量达到一定程度时，就可以借助神经网络进行学习和优化原有的模型，从而能够更加准确地判断抽水蓄能机组，还能从各个输入指标详细分析造成机组评价值高低的原因。由于神经网络对于学习的准确性，其不仅能够做到对当前机组的评价，更重要的是，还能对将来的机组状态做预测，这就扩展了关于机组综合评价的应用。**EA**

（收稿日期：2017.06.18）
（编辑：白文亭）