KPCA Feature Extraction Based on Bacterial Foraging Algorithm

X.J.Li\textsuperscript{a}, D.L.Yang\textsuperscript{b}, Z.Q.Deng\textsuperscript{c} and L.L.Jiang\textsuperscript{d}

Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan, 411201, China

\textsuperscript{a}hnkjdxlxj@163.com, \textsuperscript{b}hyydl216@163.com, \textsuperscript{c}dengzongqun@163.com \textsuperscript{d}linlyjiang@163.com

Keywords: Bacterial foraging algorithm (BFA); Kernel principal component Analysis (KPCA); Feature extraction; Parameter optimization

Abstract. KPCA is a commonly used method for feature extraction, for the problems of kernel function and its parameters have a great influence on performance of KPCA feature extraction but the optimal parameters are difficult to select. This paper applied the bacterial foraging algorithm on KPCA feature extraction and the method of KPCA feature extraction based on bacterial foraging algorithm was proposed. The experiment of bearing feature extraction shows that the method which proposed in this paper is effective.

Introduction

Rotating machinery is the core equipment in modern industrial production, carrying out rotating machinery fault diagnosis technology research has important significance for economic and social to ensure such equipment running safe and efficient and to avoid the huge economic loss and catastrophic accidents, meanwhile, it can enrich and develop the technology of mechanical equipment condition monitoring and fault diagnosis.

Feature extraction is an important part of fault diagnosis, the feature which is extracted from time-domain signal is effective only it can represent the different fault information very good, however, because of interference, the signal contains a large amount of noise, and the fault feature are often nonlinear. Kernel method maps the original space to feature space implicitly by kernel function, find the linear relationship in the feature space, so that it can solve the nonlinear problem, linear search in the feature space can be realized efficiently solve nonlinear problems. Kernel principal component is such a method which mainly implied in feature extraction, however, its performance are affected by the kernel function type and parameters very large but the optimal parameters are difficult to select. Currently, the method implied in parameter optimization include grid search method, cross validation (CV) [1], genetic algorithm (GA)[2], ant colony optimization (ACO)[3], Particle Swarm Optimization,(PSO)[4],but all of them are not ideal for optimal efficiency. Bacterial foraging algorithm (BFA)[5] is a new bionic algorithm, which has many advantages such as group intelligence, capable of parallel search, easy to jump out of local minima and so on, and which obtain good result implied in power systems[6], shop scheduling[7] and other fields[8,9,10,11], but less involved in fault diagnosis.

Principles of bacterial foraging algorithm

Bacteria foraging algorithm which abstract from biological behavior of E. coli is a bionic intelligent optimization algorithm, the whole process of this algorithm is a process of iteration, including three basic operations: chemotaxis, reproduction and elimination-dispersal.

Chemotaxis. This process mainly simulates “tumble” and “swim”. Suppose \( P^i(j,k,l) \) represents the bacterium’s position at \( jth \) chemotactic, \( kth \) reproductive, and \( lth \) elimination-dispersal step, the next position defined as:

\[
P^i(j+1,k,l) = P^i(j,k,l) + C(i)V(j)
\]

(1)
Where \( C(i) \) is the step length vector of the \( i \)th bacterium, \( V(j) \) is the direction vector of the \( j \)th chemotactic; where is randomly generated;

Reproduction. This process mainly simulates the selection process of bacterium survival of the fittest individuals. Suppose the populations’ size of bacterium is \( N \), \( F'(j,k,l) \) is the fitness of the \( i \)th bacterium, sort by descending, then use the front \( N/2 \) bacterium replace the behind \( N/2 \) bacterium.

Elimination-dispersal. This process use new individual to replace the original individual, which different from the reproduction, this process occur according probability \( p \), when an individual meets the conditions of migration, this individual will be eliminated, to maintain population balance, randomly generated an new individual instead of the individual which is eliminated.

KPCA feature optimization extraction based on bacterial foraging algorithm

When use KPCA to carry out feature extraction, kernel parameters and different kernel functions have a great impact on it, if the kernel parameter is unsuitable, the number of kernel principal opponent which cumulative contribution rate more than 0.85 larger than the dimensions of original data, and not achieve the role of dimensionality reduction at all. Take the Gaussian kernel for example, suppose there are \( n \) kinds of sample set, the step of KPCA feature extraction based on bacterial foraging algorithm are as follows:

1. Reading the \( k \)th sample set \( S_k \), Set the number of bacteria population as \( N \), and randomly generated \( N \) groups \( \sigma \) as the initial position of the location of \( N \) bacteria;
2. Selecting the fitness evaluation function \( \text{fitness} = f(\sigma) \), carrying out the KPCA with \( \sigma \) of each bacterium respectively, here, \( \text{fitness} \) is the number of number of principal components which the contribution rate more than 0.85;
3. Completing the iterations times of chemotaxis, reproduction and elimination-dispersal, then obtain the optimal parameter \( \sigma^*_k \) and the number of kernel principal component \( \text{fitness}^*_k \);
4. Selecting \( \text{fitness}^* = \max(\text{fitness}^*_k, k = 1, 2, \cdots, n) \) as the number of principal components of for all categories.

Experiment

Data acquisition. The experiment of rolling bearing conducted on United States Spectra Quest’s comprehensive fault simulation test rig, this platform can simulate a variety of bearing faults. In this experiment, 8 accelerometers are configured (s1~s8), as shown in Fig.1. Where, s1 ~ s4 installed on base, s5 ~ s8 installed on the bearing seat, the sampling frequency is 10KHz, experiments collected several groups vibration signal about normal ring bearing, inner fault bearing, outer fault bearing and ball fault bearing in the speed 1800 r / min, 2100 r / min and 2400 r/min respectively.

Feature extraction. Rolling bearing primitive feature vector generation method as follows:

Suppose \( X_i = \{x_{i1}, x_{i2}, \cdots, x_{i8}\} \) as sequence collected by the \( i \)th sensor \( (1 \leq i \leq 8) \), Taking the cumulative energy \( E \) as the primitive feature,
\[ E_i = \sum_{j=1}^{n} x_{ij}^2 \]  

Then, obtaining an 8 dimension feature vector from 8 sensors:

\[ P = [E_1, E_2, \cdots, E_8]^T \]  

Experiment extract energy feature of different faults bearing vibration signal for follow-up analysis. In this experiment, for each type of bearing states, 150 groups of samples are collected, totally 600 groups, composed of \(600 \times 8\) sample data set, and set up primitive feature libraries 1~3.

**Analyze.** Selected feature library 1, feature library 2, feature library 3 as test data set, Set the initial population of bacteria is 50, Convergence Threshold is \(\varepsilon = 10^{-6}\), optimization space: \(2^{-5} < \sigma < 2^{15}\), Fig. 2, Fig.3 and Fig.4 shows all stages of spatial distribution of population in the process of optimization by bacterial foraging algorithm (Not units).

Seen from Fig.2, Fig.3 and Fig.4, obtaining the optimal parameter only carry out reproduction once, as the number of optimal parameter is not unique, just select one of them, actually, often select the optimal parameter which nearest from the center of population.

---

**Fig.2** The process of KPCA parameter optimization (1800 r/m rolling bearing)

**Fig.3** The process of KPCA parameter optimization (2100 r/m rolling bearing)

**Fig.4** The process of KPCA parameter optimization (2400 r/m rolling bearing)
In order to verify the validity of features, using Gaussian kernel, carry out KPCA feature extraction with the not optimal parameter $\sigma = 0$. Fig.5, Fig.6 and Fig.7 both are the projections of original features, not optimized features and optimized features in 2D space. In the following figure, the marker meaning show behind (*: Normal sample, △: Inner sample, o: Outer sample, *: Ball sample).

![Fig.5 The 2D projections of 1800 r/m rolling bearing](image)

![Fig.6 The 2D projections of 2100 r/m rolling bearing](image)

![Fig.7 The 2D projections of 2400 r/m rolling bearing](image)

Seen from Fig.5, Fig.6 and Fig.7, the original feature and the no optimized KPCA feature both have different levels of aliasing between different fault samples, however, the optimized feature completely separate between the different fault samples and the clustering effect is very good. That is to say, after optimization, the KPCA features represent the difference of different types of faults better.
Conclusion

Bacterial foraging algorithm is a novel bionic optimization algorithm, f or the problems of kernel function and its parameters have a great influence on performance of KPCA feature extraction but the optimal parameters are difficult to select, this paper applied the bacterial foraging algorithm in KPCA kernel parameter optimization and designed KPCA feature extraction algorithm based on bacterial foraging algorithm. The results show that bacterial foraging algorithm can solve difficult problem of KPCA kernel parameters selection, and that can find the optimal parameter quickly. Compared with the original feature and the no optimized feature, the optimized feature can represent the difference of different types of faults better; improve the effectiveness of the feature.

Acknowledgement

Financial support from Aid program for Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province, the CEEUSRO special plan of Hunan province (2010XK6066), The Industrial Cultivation Program of Scientific and Technological Achievements in Higher Educational Institutions of Hunan Province (10CY008), New Century Excellent Talents (NCET-08-0677), Natural Science Foundation of Hunan Province Key Project (09JJ8005), National natural science foundation of China (51105138) are gratefully acknowledged.

References