Expert Systems with Applications 41 (2014) 3585–3595

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Global geometric similarity scheme for feature selection in fault diagnosis

Chao Liu *, Dongxiang Jiang, Wenguang Yang

State Key Laboratory of Control and Simulation of Power System and Generation Equipments, Department of Thermal Engineering, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Keywords:
Feature selection
Global geometric similarity scheme
Condition classification
Feature ranking
Feature wrapper

ABSTRACT

This work presents a global geometric similarity scheme (GGSS) for feature selection in fault diagnosis, which is composed of global geometric model and similarity metric. The global geometric model is formed to construct connections between disjoint clusters in fault diagnosis. The similarity metric of the global geometric model is applied to filter feature subsets. To evaluate the performance of GGSS, fault data from wind turbine test rig is collected, and condition classification is carried out with classifiers established by Support Vector Machine (SVM) and General Regression Neural Network (GRNN). The classification results are compared with feature ranking methods and feature wrapper approaches. GGSS achieves higher classification accuracy than the feature ranking methods, and better time efficiency than the feature wrapper approaches. The hybrid scheme, GGSS with wrapper, obtains optimal classification accuracy and time efficiency. The proposed scheme can be applied in feature selection to get better accuracy and efficiency in condition classification of fault diagnosis.

1. Introduction

As an important aspect of pattern recognition, feature selection is researched in many subjects including filter approach, wrapper approach and hybrid method (Liu, Zhao, Li, & Li, 2012; Oreski & Oreski, 2014). The filter approach selects features by their performance, which is evaluated with metrics including distance, information, dependency, and consistency (Vieira, Sousa, & Kaymak, 2012; Wu, Chen, Kechadi, & Sun, 2013). Feature selection with wrapper approach is realized by comparing the classification accuracy in feature subsets, where the feature subsets are formed by search algorithms and the classification accuracy is acquired by classifiers (Cadenas, Carmen Garrido, & Martínez, 2013). Therefore, wrapper approach can obtain more significant feature subset while it needs high computational cost. Compared with the high computational cost in the wrapper approach, the filter approach is more efficient, especially in the situation when the original feature space consists of high dimensions (Uncu & Türkşen, 2007). Also, hybrid scheme with filter and wrapper is analyzed to balance the accuracy and efficiency (Zhang, Li, Scarf, & Ball, 2011).

Feature ranking, a widely applied filter approach, selects features with metrics such as data variance, distance estimate, correlation estimate, relief algorithm, information entropy, and mutual information (Kappaganthu & Nataraj, 2011; Lei, He, Zi, & Chen, 2008; Sakar, Kursun, & Gurgen, 2012; Zhang et al., 2011). Decision tree is also applied to sort the features, which is based on information entropy of each feature (Saimurugan, Ramachandran, Sugumaran, & Sakthivel, 2011; Sugumaran, Muralidharan, & Ramachandran, 2007). Feature ranking selects features based on the performance of each single feature. However, a combination of individually good features does not necessarily lead to good classification performance (Jain, Duin, & Mao, 2000; Peng, Long, & Ding, 2005). Therefore, feature combinations are more applied to select optimal feature subset including linear algorithms and nonlinear algorithms.

K-nearest neighborhood based approach is a linear algorithm to evaluate the performance of feature subsets (Casimir, Boutleux, Clerc, & Yahoui, 2006; Wang, Neskovic, & Cooper, 2006). The shortcoming of K-nearest neighborhood based approach is that it cannot keep the geometric structure of the data in the originally high dimensional space, when representing the data in a lower dimensional space. For example, the shortest path of Swiss roll data with Euclidean distance cannot reflect the structure of the data set (Tenenbaum, De Silva, & Langford, 2000). Principal Component Analysis (PCA) is a category of dimensionality reduction approaches in discriminating directions in data set and finding out the sensitive directions of maximal variance (Yu, 2011; Zimroz & Bartkowiak, 2013). It is a linear transformation from a high dimensional space to a lower dimensional space, and less effective in nonlinear data with a certain type of topological manifold. As Euclidean structure acquired, PCA cannot reflect the true low-dimensional geometry of the manifold, such as Swiss roll data (Tsai, 2011; Yang, Nie, Xiang, Zhuang, & Wang, 2010).

* Corresponding author. Tel.: +86 1062794046; fax: +86 1062770209.
E-mail address: chaoliu13@tsinghua.edu.cn (C. Liu).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.11.037

Nonlinear algorithms in filter approaches are researched to find out the geometric structure of the data set, which can get better feature selection results. Multiple manifolds analysis, extended from manifold learning, is a nonlinear approach to obtain the geometric structure by embedding data set into a high-dimensional space and then getting neighbors to approximate its local geometry (Li, Xu, Yang, Yang, & Wang, 2009). Also, Isomap and Locally Linear Embedding (LLE) are nonlinear approaches to retain the geometric structure of the data set in dimensionality reduction. Based on classical Multidimensional Scaling (MDS), Isomap is proposed to preserve the intrinsic geometry of the data, which is captured in the geodesic manifold distances between all pairs of data points (Tenenbaum et al., 2000). In Isomap, Geodesic distances of faraway points are approximated by adding up a sequence of "short hops" between neighboring points. LLE is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs. LLE recovers global nonlinear structure from locally linear fits, and maps its inputs into a single global coordinate system of lower dimensionality (Roweis & Saul, 2000). With similar approach, Locality Preserving Projections (LPP) is proposed to discover the global structure of the Euclidean space (Niyogi, 2004), and applied in bearing performance degradation assessment (Yu, 2011). Among the nonlinear algorithms, the geometric structure of the data is represented by the neighborhood relationship and shortest paths with geodesic distances. If disjoint clusters are presented, Isomap and LLE will fail to detect the global geometric structure of the data (Varini, Degenhard, & Nattkemper, 2006).

In machine condition monitoring and fault diagnosis, features extracted from field measurements of sensors are in high dimensions, as the sensors installed in the condition monitoring system are numerous and redundant. The extracted features, forming high dimensional feature space, provide substantial information for condition classification and fault diagnosis. However, condition classification is more complicated and less reasonable in the high dimensional feature space than that in a lower one. Therefore, feature selection is necessary to reduce the dimensions of the feature space, improve the understandability of the condition classification, and get better accuracy in fault diagnosis.

Feature selection methods applied in fault diagnosis include feature ranking method, feature wrapper approach and hybrid method. Feature ranking method with distance estimate is applied in fault diagnosis of locomotive roller bearings (Lei et al., 2008). Mutual information and decision tree are applied in feature selection of fault diagnosis (Kappaganthu & Nataraj, 2011; Saimurugan et al., 2011; Sugumaran et al., 2007). Feature selection with nearest neighbors rule is analyzed in faults diagnosis of induction motors (Wan, Lee, Rajkumar, & Isa, 2012). PCA-based feature selection scheme is established for bearing fault diagnosis (Malhi & Gao, 2004). Hybrid feature selection is also applied in fault diagnosis (Zhang et al., 2011), where feature ranking with eight filter models are firstly applied to pre-select features using a weighted voting scheme, and then the wrapper approaches with binary search model and Sequential Backward Search model are formed for feature selection. The above methods in fault diagnosis are linear algorithms. Nonlinear algorithms are also researched in fault diagnosis such as Locally Linear Embedding (LLE), which is applied in gearbox multi-fault diagnosis (Li et al., 2013). However, the situation that the data set contains disjoint clusters in fault diagnosis is not discussed.

As the development of wind energy, fault diagnosis in wind turbines is more and more researched (Chen, Matthews, & Tavner, 2013; García Márquez, Tobias, Pinar Pérez, & Papaelias, 2012; Kusiak & Li, 2011). However, the application of feature selection in fault diagnosis of wind turbines is less concerned. Manifold learning is recently applied in dimensionality reduction for a wind turbine transmission system, where orthogonal neighborhood preserving embedding (ONPE) is used (Tang, Song, Li, & Deng, 2014). Although the feature selection in wind turbine fault diagnosis is similar to that in other machines, analysis of dominant features of wind turbine faults by feature selection is beneficial for optimal arrangement of sensors in condition monitoring.

For the application of feature selection approaches in condition classification of fault diagnosis, conditions are disjoint, which may be identified as outliers in traditional filter approaches with nonlinear algorithms. To adapt the situation with disjoint clusters in fault diagnosis, this work presents a global geometric similarity scheme (GGSS) including global geometric model and structure similarity metric. With global geometric framework, global geometric model is formed to form connections between disjoint clusters, and shortest path matrices of feature subsets are obtained. Then similarity metrics of global geometric model are formed to select the feature subsets. To evaluate the performance of the proposed approach, experiments are carried out as follows: (I) data with twelve conditions from a direct-drive wind turbine test rig is collected; (II) GGSS is applied in the data where global geometric similarities of feature subsets are obtained; (III) classification accuracy in each feature subset is acquired with classifiers established by Support Vector Machine (SVM) and General Regression Neural Network (GRNN), respectively; (IV) GGSS is validated by estimating the correlation between condition classification accuracy and global geometric similarity; (V) classification accuracy and time efficiency in features selected by GGSS are compared with feature ranking methods and feature wrapper approaches.

2. Global geometric similarity scheme for feature selection

As the combination of individually good features does not necessarily lead to good classification performance (Jain et al., 2000; Peng et al., 2005), comparisons of performance in feature combinations are necessary to find out the best one. The performance of the feature subsets is evaluated by classifiers in feature wrapper approaches. The feature wrapper approach can select optimal feature subset with exhaustive search strategy, but cost much more computation than other methods.

This work presents a global geometric similarity scheme (GGSS) for feature selection in fault diagnosis. GGSS consists of global geometric model and structure similarity metric. The global geometric model is applied to establish shortest path matrices in the feature subsets formed with exhaustive search strategy, and the structure similarity metric is applied for evaluating the performance of the feature subsets.

2.1. Global geometric model with disjoint clusters

Global geometric framework was proposed in 2000, and applied for nonlinear dimensionality reduction (Tenenbaum et al., 2000). Different from the classical techniques such as Principal Component Analysis (PCA) and Multidimensional Scaling (MDS), global geometric framework is trying to acquire a globally optimal solution, which can preserve the structure of the data during the projection from the high dimensional space to the low dimensional space. To form the global geometric framework, geodesic distance is applied to search shortest paths between data points. Both Dijkstra's algorithm and Floyd's algorithm can calculate the shortest paths to acquire the geodesic distances of the data set. The previous one calculates the shortest path from one point to the other points in the set. While the latter one can acquire the shortest paths in all pairs of the data set once.

When there's only one cluster of the set, the shortest path will connect most of the points except the outliers. If there're more