**Effective Classification Based Feature Extraction Module For Big Data**

**Abstract**

High-dimensional data analysis is a challenge for researchers and engineers in the fields of machine learning and data mining. Feature selection provides an effective way to solve this problem by removing irrelevant and redundant data, which can reduce computation time, improve learning accuracy, and facilitate a better understanding for the learning model or data.

Feature selection involves identifying a subset of the most useful features that produces compatible results as the original entire set of features. A feature selection algorithm may be evaluated from both the efficiency and effectiveness points of view. While the efficiency concerns the time required to find a subset of features, the effectiveness is related to the quality of the subset of features.

We combine multi-strategy feature selection and grouped feature extraction and propose a novel fast hybrid dimension reduction method, incorporating their advantages of removing irrelevant and redundant information.

Firstly, the intrinsic dimensionality of the data set is estimated by the maximum likelihood estimation method. Fisher Score and Information Gain based feature selection are used as multi-strategy methods to remove irrelevant features. With the redundancy among the selected features as clustering criterion, they are grouped into a certain amount of clusters. In every cluster, Principal Component Analysis (PCA) based feature extraction is carried out to remove redundant information.

We aim to demonstrate that standard FS methods can be parallelized in big data platforms like Apache Spark so as to boost both performance and accuracy. We propose a distributed implementation of a generic FS framework that includes a broad group of well-known information theory-based methods.

**Introduction**

In this paper, we combine multi-strategy feature selection and grouped feature extraction and propose a novel fast hybrid dimension reduction method, incorporating their advantages of removing irrelevant and redundant information. Firstly, the intrinsic dimensionality of the data set is estimated by the maximum likelihood estimation method. Fisher Score and Information Gain based feature selection are used as multi-strategy methods to remove irrelevant features. With the redundancy among the selected features as clustering criterion, they are grouped into a certain amount of clusters. In every cluster, Principal Component Analysis (PCA) based feature extraction is carried out to remove redundant information. Four classical classifiers and representation entropy are used to evaluate the classification performance and information loss of the reduced set. The runtime results of different methods show that the proposed hybrid method is consistently much faster than the other three in almost all of the sets used. Meanwhile, the proposed method shows competitive classification performance, which has no significant difference basically compared with the other methods. The proposed method reduces the dimensionality of the raw data fast and it has excellent efficiency and competitive classification performance compared with the contrastive methods.

Feature selection is referred to the process of obtaining a subset from an original feature set according to certain feature selection criterion, which selects the relevant features of the dataset. It plays a role in compressing the data processing scale, where the redundant and irrelevant features are removed. Feature selection technique can pre-process learning algorithms, and good feature selection results can improve learning accuracy, reduce learning time, and simplify learning results .

However, some irrelevant features that are correlated to each other will also be integrated into a group, resulting in the existence of irrelevant features in the final subset. To solve this problem, a series of multi-stage or hybrid dimension reduction models were proposed to remove irrelevant, redundant, and noisy features .

However, feature selection or extraction operations in all these studies are carried out on the overall feature set or subset to filter out the irrelevant features or information. The mentioned clustering strategy is not combined FAST HYBRID DIMENSIONALITY REDUCTION METHOD 6 further. In fact, feature compression in every single cluster can better help to remove redundant information and cover the latent structure of the set. Therefore, we design to combine the above clustering strategy and hybrid operation and propose a novel fast hybrid dimension reduction method based on feature selection and grouped feature extraction. Firstly, two different feature selection methods are used to remove irrelevant features.

Then, the selected features are grouped into a certain amount of clusters based on the redundancy among them, in which the number of clusters is determined by intrinsic dimensionality estimation method. Finally, PCA based feature extraction is carried out in each cluster to remove redundant information. The proposed method incorporates the advantages of removing both irrelevant and redundant information of the data. Meanwhile, the latent structure of the set is also taken into account. The computational efficiency, classification performance and information loss in supervised learning problems are tested on 20 public datasets considering the diversity of data size and dimensionality. The results show that the proposed method has excellent efficiency and competitive classification performance compared with the contrastive methods.

**Problem Definition**

The embedded methods incorporate feature selection as a part of the training process and are usually specific to given learning algorithms, and therefore may be more efficient than the other three categories.

Traditional machine learning algorithms like decision trees or artificial neural networks are examples of embedded approaches. In Traditional machine learning algorithms generality of the selected features is limited and the computational complexity is large.

The wrapper methods use the predictive accuracy of a predetermined learning algorithm to determine the goodness of the selected subsets, the accuracy of the learning algorithms is usually high. However, the generality of the selected features is limited and the computational complexity is large.

The filter methods are independent of learning algorithms, with good generality. Their computational complexity is low, but the accuracy of the learning algorithms is not guaranteed.

Traditionally, feature subset selection research has focused on searching for relevant features. A well-known example is Relief which weighs each feature according to its ability to discriminate instances under different targets based on distance-based criteria function.

However, Relief is ineffective at removing redundant features as two predictive but highly correlated features are likely both to be highly weighted. Relief-F extends Relief, enabling this method to work with noisy and incomplete data sets and to deal with multiclass problems, but still cannot identify redundant feature.

So we require The hybrid methods as a combination of filter and wrapper methods ,in which by using a filter method we can reduce search space that will be considered by the subsequent wrapper. We will mainly focus on combining filter and wrapper methods to achieve the best possible performance with a particular learning algorithm with similar time complexity of the filter methods.

**On Clustering side below is the difference form exiting Hadoop:**

Apache Spark and Hadoop focused on processing data in parallel across a cluster, but the biggest difference is that it works in-memory.

Whereas Hadoop reads and writes files to HDFS, Spark processes data in RAM using a concept known as an RDD, Resilient Distributed Dataset. So Spark clustering is very fast as compared to Hadoop.

Spark can run either in stand-alone mode, with a Hadoop cluster serving as the data source, or in conjunction with Mesos. In the latter scenario, the Mesos master replaces the Spark master or YARN for scheduling purposes.

**Research Gap**

There are three main models to feature selection: filter, wrapper, and embedded models. The filter models can achieve quick sorting of features to remove a large number of irrelevant or noise features. They usually have good generalization performance and high computational efficiency as they are independent of the classification algorithm. Therefore, these feature select base methods can effectively remove the irrelevant features, but the potential structure of the original sets will be destroyed most likely. Feature selection based on Fisher score (Malina, 1981)[13], information gain (Quinlan, 1986)[14], mutual information (Peng, Long, & Ding, 2005)[15], and Gini index (W. Shang, et al., 2007)[16] are few classical methods.

Mr. M. Senthil Kumar, Ms. V. Latha Jothi M.E in” A Fast Clustering Based Feature Subset Selection Using Affinity Propagation Algorithm”- Traditional approaches for clustering data are based on metric similarities, i.e., nonnegative, symmetric, and satisfying the triangle inequality measures using graphbased algorithm to replace this process a more recent approaches, like Affinity Propagation (AP) algorithm can be selected and also take input as general non metric similarities [1].

Priyanka M G in “Feature Subset Selection Algorithm over Multiple Dataset”- here a fast clustering based feature subset selection algorithm is used. The algorithm involves (i) removing irrelevant features, (ii) constructing clusters from the relevant features, and (iii) removing redundant features and selecting representative features. It is an effective way for reducing dimensionality. This FAST algorithm has advantages like efficiency and effectiveness. Efficiency concerns the time required to find a subset of features and effectiveness is related to the quality of the subset of features. It can be extended to use with multiple datasets [2].

**Probable Chapter of Final Thesis**

**Chapter 1: Introduction**

**Chapter2: Literature Review**

**Chapter 3: Problem Definition and Research Gap**

**Chapter 4: Aim and Objective**

**Chapter 5: Software and Hardware Specification**

**Chapter 6: Methodology**

**Chapter 7: Implementation Results**

**Chapter 8: Conclusion**

**Chapter9: References**

**Literature Survey**

We have entered an era of big data with the typical characteristics of large data set size and high dimensionality (Wang, Wang, & Chang, 2016)[5]. It brings us huge challenges to extract useful information from massive data. Compared with the difficulty of data reduction, the curse of dimensionality (Golay & Kanevski, 2017; Maeda, 2014)[6] may be more difficult to solve. Generally, there are a large number of irrelevant and redundant features in high-dimensional data set and it increases the difficulty of data processing, knowledge mining, and pattern classification. As the key method to solve this problem, dimensionality reduction can filter out some noise and redundant information by reducing the original high-dimensional space to the low-dimensional intrinsic space (Golay & Kanevski, 2017; Maeda, 2014)[6]. On the premise of effectively reducing the dimensionality, it is an effective and reasonable way of dimension reduction to retain the implicit rules or topological structure in the original data space. It helps to extract meaningful insights from the original data set, reduce the complexity of data processing,

release the computational burden of the computer, and also helps to improve the stability and interpretability of the learning model (Wang, et al., 2016)[5].

In addition, dimension reduction also provides useful bases for effective and clear data visualization. In general, dimensionality reduction methods can be divided into two types: feature extraction (Choi, Shin, Lee, Sheridan, & Lu, 2017; GenaroDaza-Santacoloma, et al., 2009; Subasi & Gursoy, 2010; Sun, Gang, Bo, Zhang, & Zhang, 2017) [7]based type and feature selection (Das, Sengupta, & Bhattacharyya, 2018; Dessì & Pes, 2015; Ferreira & Figueiredo, 2012; Kolhe & Deshkar, 2017)[8] based type.

Feature extraction methods are usually based on feature transformation, essentially projecting high-dimensional data into low-dimensional subspace.

This kind of dimension reduction methods generally preserve the original relative distance between features and help to cover the potential structure of the original data, so they will not cause a large loss of information. However, when encountering such data sets containing a large number of irrelevant features, the effect of dimensionality reduction is usually poor because almost all features are inevitably taken into account in projection. Principal Component Analysis (PCA) (Hotelling, 1933)[9], Multi-Dimensional Scaling (MDS) (Kruskal & Wish, 1978)[10], Is ometric Mapping (ISOMAP) (Tenenbaum, Silva, & Langford, 2000)[11], and Locally Linear Embedding (LLE) (Roweis & Saul, 2000) [12] are typical feature extraction based dimensionality reduction methods. Feature selection methods sort the original features according to specific criteria and select the top-ranked features to form a subset.

A new framework is introduced that decouples relevance analysis and redundancy analysis. We develop a correlation-based method for relevance and redundancy analysis, and conduct an empirical study of its efficiency and effectiveness comparing with representative methods [3].

Yanxia Zhang, Ali Luo, and Yongheng Zhao in” An automated classification algorithm for multi-wavelength data” we applied a kind of filter approach named Relief to select features from the multi-wavelength data. Then we put forward the naive Bayes classifier to classify the objects with the feature subsets and compare the results with and without feature selection, and those with and without adding weights to features. The result shows that the naive Bayes classifier based on Relief algorithms is robust and efficient to preselect AGN candidates [4].

**Aim & Objective**

Our Aim is to design a feature based clustering module which will be give competitive classification performance, it should reduces the dimensionality of the raw data fast and should have excellent efficiency and competitive classification performance compared with the contrastive methods.

Feature selection is an important approach for reducing the dimension of high-dimensional data. In recent years, many feature selection algorithms have been proposed. However, most of them only exploit information from the data space. They often neglect useful information contained in the feature space, and typically do not exploit information about the underlying geometry of the data.

To overcome these problems, our aim is to introduce novel fast hybrid dimension feature selection methods based on the feature selection framework of joint embedding learning, sparse regression, and subspace learning, and extend the framework by introducing the feature graph.

One objective for both feature subset selection and feature extraction methods is to avoid overfitting the data in order to make further analysis possible. The simplest is feature selection, in which the number of gene probes in an experiment is reduced by selecting only the most significant according to some criterion such as high levels of activity

Second Objective is improving the classification accuracy by removing both irrelevant and redundant information of the data.

Third objective is to demonstrate that standard FS methods can be designed in these Big Data platforms and still can prove to be useful when dealing with big datasets, boosting both performance and accuracy

**Methodology**

The spark.ml package aims to provide a uniform set of high-level APIs built on top of DataFrames that help users create and tune practical machine learning pipelines

FS methods can be broadly categorized as :

1. Wrapper methods, which use an evaluation function dependent on a learning algorithm . They are aimed at optimizing a predictor as part of the learning process.

2. Filtering methods, which use other selection techniques as separability measures or statistical dependences. They only consider the general characteristics of the dataset, being independent of any predictor .

3. Embedded methods, which use a search procedure which is implicit in the classifier/regressor .

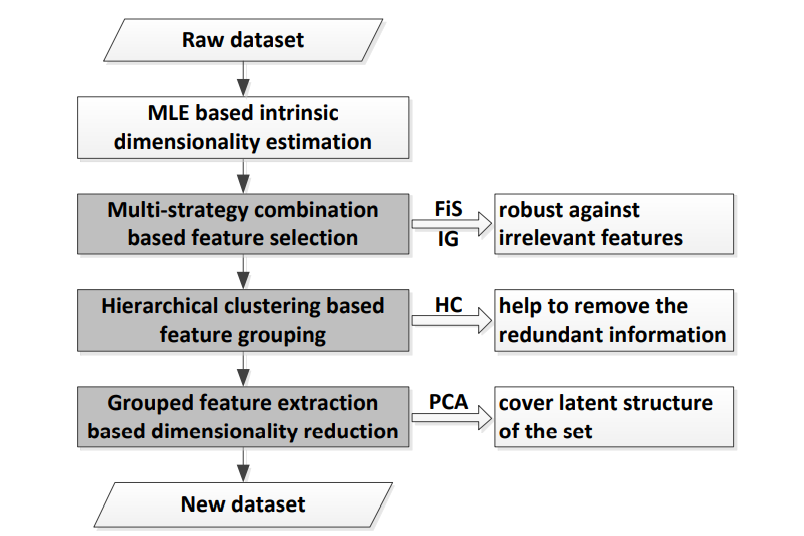
Filter methods usually result in a better generalization due to its learning independence. Nevertheless, they usually select larger feature subsets, requiring sometimes a threshold to control them. Regarding complexity, filters are normally less expensive than wrappers. In those cases in which the number of features is large (especially for Big Data), it is indispensable to employ filtering methods as they are much faster than the other approaches.

PCA is one of the most popular dimensionality reduction method proposed by Hotelling (Hotelling, 1933). With the variance of data as the standard, PCA measures the amount of information contained in the data. Higher variance corresponds to a larger amount of information. The computation of PCA includes singular value decomposition and projection transformation. The original high-dimensional data is mapped to a linear subspace formed by a few number principal components with relatively larger eigenvalues. Thus the correlation among the original dimensions is eliminating and the dimension of the data is reduced. In this paper, PCA is used to reduce the redundancy in every cluster obtained through hierarchical clustering, in which the features share much redundant information and are highly compressible.

A pictorial summary of the proposed dimensionality reduction method processing is presented in Figure 2. The processing starts with the raw datasets.

Next step is intrinsic dimensionality estimation FAST based on MLE (maximum likelihood estimation )to determine the ultimate dimension. After the above estimation, multi-strategy combination based feature selection is used to filter irrelevant and noise features. We use two kinds of feature ranking methods FiS and IG to assign a score to each feature and a series of low scored features will be filtered out. Next, the remaining features will be clustered into different groups according to the redundant information among them by Spark Mlib.

The number of groups is determined by the results of the above intrinsic dimensionality estimation. Hierarchical clustering based on the maximal information compression index is used to perform this task. Finally, the feature extraction method PCA is applied to further refine the selected feature space. The first principal component of each feature group is used as the corresponding representative feature vector, which not only removes redundant information in the group but also covers the latent structure of the set. To evaluate the proposed method comprehensively, runtime, classification accuracy, and information loss are recorded to be compared with the other conditions or methods

****

**Fig2: Proposed Methodology**

**Project Completion Schedule**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Number** | **Milestone Name** | **Milestone Description** | **Timeline** Week no. from the start of the project | **Remarks** |
| 1 | Requirements Specification | Complete specification of the system (with appropriate assumptions). A document detailing the same should be written and a presentation on that be made. | 2-3 | Attempt should be made to add some more relevant functionality other than those that are listed in this document. |
| 2 | Technology familiarization | Understanding of the technology needed to implement the project. | 4-5 | The presentation should be from the point of view of being able to apply it to the project, rather than from a theoretical perspective. |
| 3 | Database creation | A database having a table to add the fields details should be created. | 5-7 | It is important to finalize on the database at this stage itself so that development and testing can proceed with the actual database itself. |
| 4 | High-level and Detailed Design | Listing down all possible scenarios cancellation, and then coming up with flow-charts or pseudo code to handle the scenario. | 7-9 | The scenarios should map to the requirement specification (ie, for each requirement that is specified, a corresponding scenario should be there). |
| 5 | Implementation of the front-end of the system | Implementation of the various screens to be added. | 10-12 | During this milestone period, it would be a good idea for the team (or one person from the team) to start working on a test-plan for the entire system. This test-plan can be updated as and when new scenarios come to mind. |
| 6 | Integrating the front-end with the database | The front-end developed in the earlier milestone will now be able to update the database. In short, the system should be ready for integration testing. | 12-13 |  |
| 7 | Integration Testing | The system should be thoroughly tested by running all the testcases written for the system (from milestone 5). | 14-15 | Another 2 weeks should be there to handle any issues found during testing of the system. After that, the final demo can be arranged. |
| 8 | Final Review | Issues found during the previous milestone are fixed and the system is ready for the final review. | 16-18 | During the final review of the project, it should be checked that all the requirements specified during milestone number 1 are fulfilled (or appropriate reasons given for not fulfilling the same) |

**Reference**

[1] Mr. M. Senthil Kumar, Ms. V. Latha Jothi M.E,”A Fast Clustering Based Feature Subset Selection Using Affinity Propagation Algorithm”, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014 Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT’14).

[2] Priyanka M G “Feature Subset Selection Algorithm over Multiple Dataset” Proceedings of IRF International Conference, Goa, 16th March-2014, ISBN: 978-93-82702-65-8.

[3]Lei Yu, Huan Liu,” Efficient Feature Selection via Analysis of Relevance and Redundancy” Journal of Machine Learning Research 5 (2004) 1205–1224 Submitted 1/04; Revised 5/04; Published 10/04.

[4]Yanxia Zhang\*a, Ali Luo\*a, and Yongheng Zhao\*a,” An automated classification algorithm for multi-wavelength data” National Astronomical Observatories, Chinese Academy of Sciences, China.

[5] Wang, L., Wang, Y., & Chang, Q. (2016). Feature selection methods for big data bioinformatics: A survey from the search perspective. Methods, 111, 21-31.

[6] Golay, J., & Kanevski, M. (2017). Unsupervised feature selection based on the Morisita estimator of intrinsic dimension. Knowledge-Based Systems, 135, 125-134.

[7]Choi, S., Shin, J. H., Lee, J., Sheridan, P., & Lu, W. D. (2017). Experimental demonstration of feature extraction and dimensionality reduction using memristor networks. Nano Letters, 17, 3113-3118.

[8] Das, A. K., Sengupta, S., & Bhattacharyya, S. (2018). A Heterogeneous Feature Models using rough set theory based genetic algorithm. Applied Soft Computing, 65, 400- 411.

[9] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24, 417-441.

[10]Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. Beverly Hills and London: Sage Publications.

[11]Tenenbaum, J. B., Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290, 2319-2323.

[12]Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290, 2323-2326.

[13]Malina, W., . (1981). On an extended fisher criterion for feature selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 3, 611-614.

[14]Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1, 81-106.

[15] Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of maxdependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 1226-1238.

[16] Shang, W., Huang, H., Zhu, H., Lin, Y., Qu, Y., & Wang, Z. (2007). A novel feature selection algorithm for text categorization. Expert Systems with Applications, 33, 1-5.

[17] A. Khotanzad, Y.H. Hong, Rotation invariant image recognition using features selected via a systematic method, Pattern Recognition 23 (1990) 1089-1101.

[18] N. Vasconcelos, Feature selection by maximum marginal diversity: optimality and implications for visual recognition, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, pp. 762-769.

[19] N. Vasconcelos, M. Vasconcelos, Scalable discriminant feature selection for image retrieval and recognition, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2004.

[20] J.Y. Choi, Y.M. Ro, K.N. Plataniotis, Boosting color feature selection for color face recognition, IEEE transactions on image processing 20 (2011) 1425- 1434. [14] A. Goltsev, V. Gritsenko, Investigation of efficient features for image recognition by neural networks, Neural Networks 28 (2012) 15-23.

[21] D.L. Swets, J.J. Weng, Efficient content-based image retrieval using automatic feature selection, in: Proceedings of International Symposium on Computer Vision, 1995.

[22] D.L. Swets, J.J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Transactions on pattern analysis and machine intelligence 18 (1996) 831-836.

[23] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, A simultaneous feature adaptation and feature selection method for content-based image retrieval systems, Knowledge-Based Systems 39 (2013) 85-94.

[24] D.D. Lewis, Y. Yang, T.G. Rose, F. Li, Rcv1: A new benchmark collection for text categorization research, Journal of machine learning research 5 (2004) 361-397.

[25] L.P. Jing, H.K. Huang, H.B. Shi, Improved feature selection approach TFIDF in text mining, in: Proceedings of International Conference on Machine Learning and Cybernetics, 2002, pp. 944-946.

[26] S. Van Landeghem, T. Abeel, Y. Saeys, Y. Van de Peer, Discriminative and informative features for biomolecular text mining with ensemble feature selection, Bioinformatics 26 (2010) 554-560.

[27] G. Stein, B. Chen, A.S. Wu, K.A. Hua, Decision tree classifier for network intrusion detection with GA-based feature selection, in: Proceedings of the 43rd ACM Southeast conference, 2005, pp. 136-141.

[28] F. Amiri, M.R. Yousefi, C. Lucas, A. Shakery, N. Yazdani, Mutual information-based feature selection for intrusion detection systems, Journal of Network and Computer Applications 34 (2011) 1184-1199.

[29] A. Alazab, M. Hobbs, J. Abawajy, M. Alazab, Using feature selection for intrusion detection system, in: Proceedings of International Symposium on Communications and Information Technologies (ISCIT), 2012, pp. 296-301.

[30] H. Liu, J. Li, L. Wong, A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns, Genome informatics 13 (2002) 51-60.

[31] H. Liu, H. Han, J. Li, L. Wong, Using amino acid patterns to accurately predict translation initiation sites, In silico biology 4 (2004) 255-269. [26] Q. Song, J. Ni, G. Wang, A fast clustering-based feature subset selection algorithm for high-dimensional data, IEEE Transactions on knowledge and data engineering 25 (2013) 1-14.

[32] G. Li, X. Hu, X. Shen, X. Chen, Z. Li, A novel unsupervised feature selection method for bioinformatics data sets through feature clustering, in: Proceedings of IEEE International Conference on Granular Computing, 2008, pp. 41-47.

[33] Y.F. Gao, B.Q. Li, Y.D. Cai, K.Y. Feng, Z.D. Li, Y. Jiang, Prediction of active sites of enzymes by maximum relevance minimum redundancy (mRMR) feature selection, Molecular BioSystems 9 (2013) 61-69.

[34] D.S. Huang, C.H. Zheng, Independent component analysis-based penalized discriminant method for tumor classification using gene expression data, Bioinformatics 22 (2006) 1855-1862.

[35] C.H. Zheng, D.S. Huang, L. Zhang, X.Z. Kong, Tumor clustering using nonnegative matrix factorization with gene selection, IEEE Transactions on Information Technology in Biomedicine 13 (2009) 599-607.

[36] H.J. Yu, D.S. Huang, Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 10 (2013) 457-467.

[37] L. Wang, J. Yu, Fault feature selection based on modified binary PSO with mutation and its application in chemical process fault diagnosis, Advances in Natural Computation (2005) 439-439.

[38] T.W. Rauber, F. de Assis Boldt, F.M. Varejão, Heterogeneous feature models and feature selection applied to bearing fault diagnosis, IEEE Transactions on Industrial Electronics 62 (2015) 637-646.

[39] K. Zhang, Y. Li, P. Scarf, A. Ball, Feature selection for high-dimensional machinery fault diagnosis data using multiple models and Radial Basis Function networks, Neurocomputing 74 (2011) 2941-2952.

[40] M. Vasconcelos, N. Vasconcelos, Natural image statistics and lowcomplexity feature selection, IEEE Transactions on pattern analysis and machine intelligence 31 (2009) 228-244.

[41] T. Khoshgoftaar, D. Dittman, R. Wald, A. Fazelpour, First order statistics based feature selection: A diverse and powerful family of feature seleciton techniques, in: Proceedings of 11th International Conference on Machine Learning and Applications (ICMLA), 2012, pp. 151-157.

[42] J. Gibert, E. Valveny, H. Bunke, Feature selection on node statistics based embedding of graphs, Pattern Recognition Letters 33 (2012) 1980-1990.

[43] M.C. Lane, B. Xue, I. Liu, M. Zhang, Gaussian based particle swarm optimisation and statistical clustering for feature selection, in: Proceedings of European conference on evolutionary computation in combinatorial optimization, 2014, pp. 133-144.

[44] J. Tang, S. Alelyani, H. Liu, Feature selection for classification: A review, Data Classification: Algorithms and Applications, (2014) .

[45] S. Alelyani, J. Tang, H. Liu, Feature Selection for Clustering: A Review, Data Clustering: Algorithms and Applications 29 (2013) 110-121.

[46] J.R. Vergara, P.A. Estévez, A review of feature selection methods based on mutual information, Neural computing and applications 24 (2014) 175-186.

[47] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, Recent advances and emerging challenges of feature selection in the context of big data, Knowledge-Based Systems 86 (2015) 33-45.

[48] J.C. Ang, A. Mirzal, H. Haron, H.N.A. Hamed, Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection, IEEE/ACM transactions on computational biology and bioinformatics 13 (2016) 971-989.

[49] R. Sheikhpour, M.A. Sarram, S. Gharaghani, M.A.Z. Chahooki, A Survey on semi-supervised feature selection methods, Pattern Recognition 64 (2017) 141-158.