2021 International Wireless Communications and Mobile Computing (IWCMC) | 978-1-7281-8616-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IWCMC51323.2021.9498694

# A Parallel Coordinates Plot Method Based on Unsupervised Feature Selection for High-Dimensional Data Visualization

Jiaqi Lou*
School of Aerospace, Transport and Manufacturing
Cranfield University
Bedford, UK
e-mail: louis.lou@cranfield.ac.uk
*corresponding author

Ke Dong
School of Computer Science and Information
Hefei University of Technology
Hefei, China
e-mail: coliapaston@163.com

Maosen Wang
School of Aerospace, Transport and Manufacturing
Cranfield University
Bedford, UK
e-mail: maosen.wang@cranfield.ac.uk

***Abstract*—In the recent years, high-dimensional data visualization has become a challenging task in data science and machine learning. As one of the most effective methods for high-dimensional data visualization, Parallel Coordinates Plots (PCPs) demonstrate dimensional reduction by transforming features of multivariate data into 2D axes. Such approach, however, does not consider the irrelevant or redundant features such that each feature is projected into the axis in a fixed manner. This paper proposed a novel PCP introduced by an unsupervised feature selection called Laplacian Score, which can be used to improve the visualization performance of PCP by ranking the importance of attributes based on their locality preserving power. The experimental results demonstrated that the performance of PCP visualization can be improved by feature selection method. Furthermore, we proposed a flexible user interface based on PCP visualization and Laplacian Score.**

***Keywords*—High-Dimensional Data Visualization, PCP, Unsupervised Feature Selection, Laplacian Score**

## I. Introduction

Over the past decades, with the rapid development of public data, information security and big data technology, the high-dimensional data visualization has become a hot topic in many research fields such as business intelligence, public service, multimedia and spatial data processing etc. [1]. As an effective method in machine learning, dimensionality reduction generally can be used to transform the high-dimensional data into a lower-dimensional space with limited information loss. Principle Component Analysis (PCA), for instance, transforms the original correlated variables into linearly unrelated low-dimensional variables [2]. Moreover, Linear Discrimination Analysis (LDA) projects the labelled training datasets on a straight line in such a manner that projections of data with the same label cluster as close as possible, and the data with different labels are separated simultaneously [3]. In RadViz Visualizer, a series of data points are non-linearly projected from high-dimensional space into a 2D circular coordinate, in which the *n*-dimensions of datasets can be represented as *n*-radiuses in the system so that a multiple objects can be denoted by a single point [4].

As the most commonly used method for processing high-dimensional data, Parallel Coordinates Plots (PCPs) is proposed by Philbert Maurice d'Ocagne in 1885 [5]. Technologically, PCP can be visualized by transforming high-dimensional data into a finite two-dimensional coordinate system. Each dimension (or called feature) can be represented as a parallel axis in PCP, where each multi-dimensional data can be mapped into a polyline among the axes. The correlation between different attributes can thus be observed by analyzing the patterns of polylines and the points of intersections with the parallel axes.

A good scalability can be found in PCP, which can be explained by an observation that the increase in data dimensions is represented by the addition of parallel axes in the 2D coordinate. The pattern's recognizability can be also enhanced by the representation in the form of polylines [6]. However, the visualization performance of PCP is not satisfactory due to the overlapping of tremendous polylines in particular when the sample size is larger. Furthermore, the positive correlation of data can be more difficult to be discriminated than the negative correlation in PCP visualization [7].

Many works have been done to improve the performance of PCP in different high-dimensional data processing tasks. For example, Orientation-enhanced Parallel Coordinate Plots (OPCPs) was proposed to increase pattern and outliner discernibility to overcome the difficulty of distinguishing specific pattern [8]. Pargnostics (Parallel Coordinates Diagnostics) is developed based on screen-space metrics that quantify different visual structures to limit the overlapping phenomenon [9]. Similarly, the occlusion and overplotting in PCP can be improved by Smart Sketch-based & Data-based Brushing techniques by using interactive pattern search [10]. In order to enhance the capability of showing multivariate local positive and negative correlations of PCP, Indexed-Points Parallel Coordinates is developed by denoting generalized flat surface with the indexed points of multidimensions [11].

In this regard, the visualization performance of PCP can be improved by the feature selection. Feature selection is an effective technique which remains the correlated feature and remove the unrelated and redundant features simultaneously, so that the learning algorithm can be run in an effective manner. In comparison with the original PCP, this paper proposed a novel PCP with feature selection. To be specific, an unsupervised feature selection method called Laplacian

Score is utilized to rearrange the axes in PCP. That is, the visualization performance of PCP can be improved by the different combinations of features [12, 13].

The main contributions of this work can be summarized as follows:

1. A novel PCP visualization based on an unsupervised feature selection called Laplacian Score is developed.
2. The visualization performance of PCP is significantly improved by feature selection method.
3. We proposed a flexible visualization operation interface which is accomplished by the combination of PCP and LS.
4. By introducing a supervised feature selection called Fisher Score, the performance of LS can be verified.

The rest of this paper is organized as follows. Section 2 introduces the related works and visualized demonstration of PCP. Section 3 describes the method of feature selection and insightful principles of LS in detail. In Section 4, experiment is conducted to obtain the visualization results of PCP with rearranged axes modified by LS. Finally, Section 5 concludes the article and points certain future works.

## II. Parallel Coordinates Plots (PCP)

Many different techniques have been developed in order to enhance the visualization performance of multivariate or high-dimensional data. Parallel Coordinates Plots (PCP), for instance, is the most commonly and widely applied method, which was initially proposed by Philbert in 1885. In late 20$^{th}$ century, Inselberg developed the early work of PCP by defining the general mathematical framework of the geometry of parallel coordinates and high-dimensional data. [14, 15]. This mathematical model consists of geometric entities such as 2D plane, parallel axis, polylines etc. The high-dimensional data is represented in a planer coordinate, and the features are denoted by the parallel axes in the coordinates of PCP. An example of PCP visualization is shown in Fig. 1.

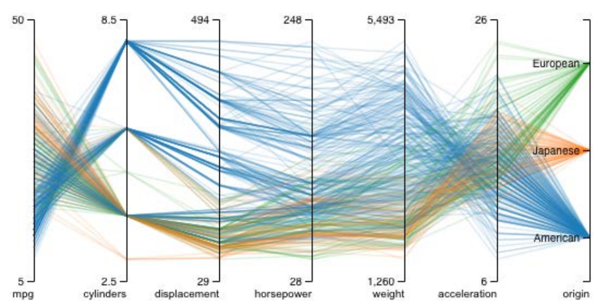



Fig. 1. An PCP visualization demonstration based on a car specification dataset.

The figure above shows the visualization of a multidimensional car dataset in PCP. The seven features of this dataset are denoted by seven parallel axes in the coordinates, and each polyline refers to an object in the dataset. By evaluating the coordinate of intersection of the polyline and the axis, the corresponding value based on a specific attribute of an object can be observed. The polylines are clustered in different colours to indicate the origin countries of the vehicles, which can be used to distinguish the data classification in PCP.

One of the main disadvantages of PCP is that the polyline-form representation covers tremendous number of pixels in visualization, especially when the scale of the given dataset is larger. Therefore, the phenomenon of overlapping of polylines can be observed among the parallel attribute axes [16]. In order to improve the visualization performance of PCP, many complementary methods are developed for clutter reduction. For instance, cluster-based ranking enhancement and proximity-based colouring method can be utilized to label and classify the polylines, different visualization effects can be generated in such manner [17]. In addition, PCP can be improved by combining with other visualization techniques such as Star Glyphs, RadViz Visualizer, and Scatter Plots etc [18].

However, such approaches only focus on clutter reduction by clustering or representation altering, the recognition of the nature of the original data is yet taken into consideration [19]. For a given sample, for example, the visualization performance in PCP might be strongly influenced by certain outliners and data noises. The visualization performance can thus be improved by filtering those relatively unimportant features [20, 21]. In this article, the visualization performance of PCP can be enhanced by an unsupervised feature selection.

## III. Feature Selection and Laplacian score

Notwithstanding a good capability can be found in PCP that denotes all the features by transforming high-dimensional attributes into planer coordinates, some of the features are irrelevant or redundant thus should not be studied. Moreover, the limitation of visualization space can lead to missing representation of parallel axes when a high-dimensional dataset is given. To tackle this issue, in this paper, an unsupervised feature selection technique called Laplacian Score is utilized to rank the importance of features according to the locality preserving power of each feature. The parallel axes, which denote the features in PCP, can be reorganized in such order so that those important features can remain in the visualization. Laplacian Score is primarily based on Laplacian Eigenmaps and Locality Preserving Projection [22-25]. Technically, Laplacian Score can be mainly expressed as the following steps:

### *A. Constructing the Adjacency Graph*

Given a data set $X=[x_1,x_2,...,x_m]\in$ with $m$ samples, and the corresponding feature set $F=[\mathrm{f}_1,\mathrm{f}_2,...,\mathrm{f}_D]^T\in$ . A graph with $m$ nodes is constructed. The node $i$ and $j$ is connected by an edge if and are close. This can be evaluated by the methods:

- -neighborhoods. [parameter $\in$ ]. If $\|x_i-x_j\|^2<$ .
- *K*-nearest neighbors. [parameter $k\in\mathrm{N}$ ]. If $i$ is among the *k*-nearest neighbors of $j$ or $j$ is among the *k*-nearest neighbor of $i$ .

Otherwise there is no connection in between the node $i$ and $j$ .

## B. Choosing the Weight

If two nodes are connected, a symmetric weight matrix is introduced, in which denotes the weight of the edge connecting the node $i$ and $j$. There are two variations defined below:

- Heat kernel. [parameter $t \in \mathrm{R}$ ].

$$W_{ij} = \mathrm{e}^{-\frac{\|x_i - x_j\|^2}{t}} \tag{1}$$

Where $t$ is a constant to control the spread of the neighbors.

- Simple-minded. [No parameter].

$$W_{ij} = 1$$

The weight is equal to 0 if the edge does not exist.

## C. Calculating Laplacian Score

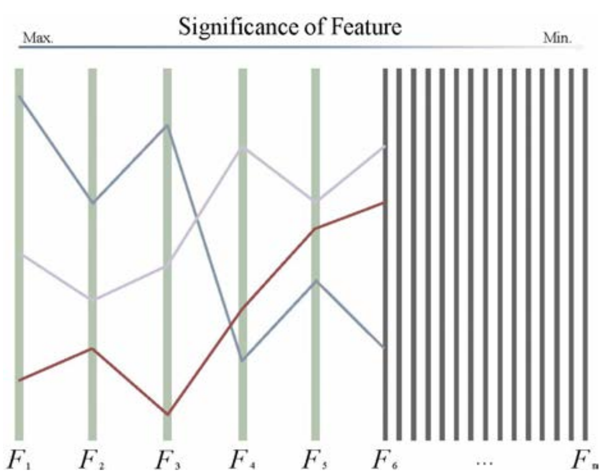
Thus, the Laplacian Score of the *d*-th feature is defined as:

$$L_r = \frac{\mathrm{f_r}^T L \mathrm{f_r}}{\mathrm{f_r}^T D \mathrm{f_r}} \tag{2}$$

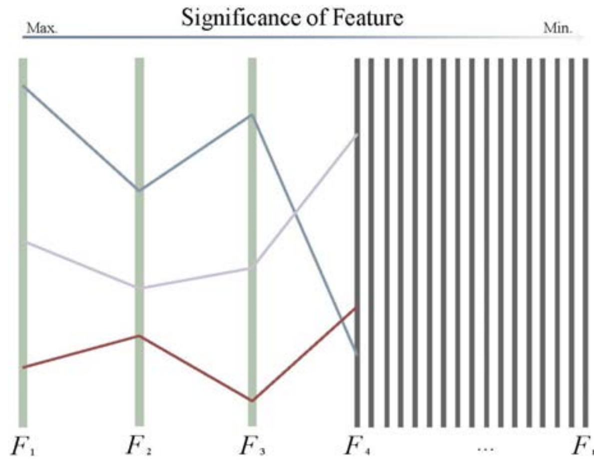
Where $D = diag(SI)$ , $L = D - S$ , $\mathrm{I} = [1,1,...,1]^T$ , $\mathrm{f_r} = \mathrm{f_r} - \frac{\mathrm{f_r}^T D\mathrm{I}}{\mathrm{I}^T D\mathrm{I}}\mathrm{I}$

# IV. EXPERIEMNT

Ranked by the Laplacian Score, the axes of PCP are rearranged in the order of the feature importance. The visualization operation interface is in a manipulated manner so that the number of the most important features that are visualized can be changed based on users' demand.




(a)




(b)

Fig. 2. An PCP visualization demonstration ranked by Laplacian Score: (Fig. 2a) with top 5 ( $F_6 \ldots F_n$ ) features shwon; (Fig. 2b) top 3 ( $F_4 \ldots F_n$ ) features shown.

In the first visualization of Fig. 2a, top five features are selected to be shown in PCP visualization interface based on the Laplacian Score each feature obtained. The axes which denote *6*th to the *n*th feature in the ranking are folded without detailed information of data visualized. The operation interface reshaped into the second visualization when only the top three features are chosen to be visualized (see Fig. 2b). It can be observed that the fourth and fifth features in the ranking are folded in the same manner as the rest of the low-ranked features, and three axes denotes the top three features are visualized with detailed polylines in this scenario.

In this study, Iris dataset are employed to verify the improved performance of PCP based on the Laplacian Score ranking. The Iris dataset consists of 150 samples, and each sample can be represented by four attributes i.e. sepal length, sepal width, petal length and petal width. The dataset was classified into three categories: setosa, versicolor and virginica. Based on the original PCP, four attributes are visualized as parallel axis in the x-direction in the default manner. The y-axis denotes the corresponding value of each feature. Polylines among parallel axes are illustrated with three different colors, which represent the labels it processed and the clustering visualization in the dataset.

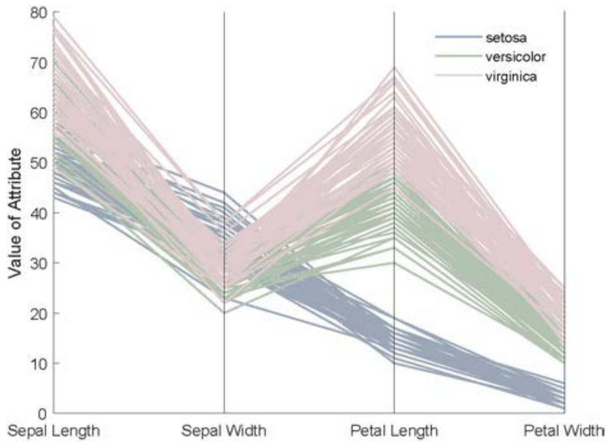
The Laplacian Score calculated by each feature is reported in Tab. 1:

TABLE I. LAPLACIAN SCORE OF EACH FEATURE IN IRIS DATASET

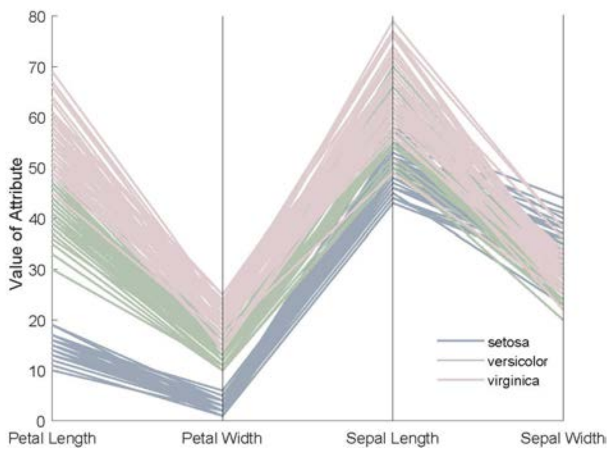
| No. | Feature | Laplacian Score | Ranking |
|---|---|---|---|
| 1 | Sepal Length | 0.9045 | 3 |
| 2 | Sepal Width | 0.5253 | 4 |
| 3 | Petal Length | **0.9744** | 1 |
| 4 | Petal Width | <u>0.9327</u> | 2 |

Bold indicates the best performance and underline indicates the second-best performance.

It can be found in Tab. 1 that the "petal length" excels the other features in Laplacian Score, following by "petal width". The feature "sepal width" receives the lowest score of 0.52 only, while the scores of the other features all exceed 0.9.




(a)




(b)

Fig. 3. The iris dataset visualizaed by (Fig. 3a) original PCP, and (Fig. 3b) the updated PCP with rearranged axes based on Laplacian Score.

The default generated PCP visualization of Iris dataset is shown in Fig. 3a. By a sharp contrast, the reranked PCP based on Laplacian Score can be found in Fig. 3b. The "petal length", which achieved the highest Laplacian Score (0.9744), displayed a clear pattern of intersections with extremely few overlapping phenomenon of the polylines compared with the other axes. The data demonstrate a good ability in clustering on this feature according to its label, and the clusters are separated well so that the reliability can be ensured when classifying a new-introduced sample. Therefore, "petal length" can be viewed as the most valuable attribute to be studied in this experiment. The original axes of PCP are reordered such that the axis denoted by "petal length" is organized as the first parallel axis in PCP visualization. As the axis sequencing to the right-hand-side, overlapping and clutter is gradually shown among the polylines on the axes which denotes the features with lower scores.

## V. Conclusion and Future Work

In this article, we presented a novel approach for PCP visualization based on Laplacian Score of feature selection and ranking. The original principle of PCP is studied, and the disadvantages and related proposed methods are discussed. The mathematical approach of Laplacian Score and the process of reordering the axes of attributes in PCP is introduced in detail. Finally, an experiment is conducted based on typical Iris dataset to verify the enhancement in the performance of visualization compared with the original PCP.

In the future, an interactive user interface can be developed based on the current work. The interface can automatically calculate the Laplacian Score for each attribute of the imported dataset and visualize the rearranged parallel plots according to their LS. A flexible interface is also required to visualize any number of feature axes with the highest ranking, while the rest attributes can be folded in the visualization based on the users' command. For reliability, our experiment is conducted based on this unsupervised feature selection method of Laplacian Score. However, many supervised and semi-supervised algorithms can be introduced in future experiments. For example, Fisher's Score, *k*-NN classification and class correlations etc. can be utilized to introduce more indicators in the process of feature selection based on multiple approaches such as weighting and clustering etc. [26-28]. The development of wireless communication has promoted the real-time nature of data transmission [29-36].By this means, the performance of Laplacian score can thus be compared and combined with other algorithm to further enrich the methodology of feature selection for the ranking purpose in PCP.

## References

[1] H. Nguyen and P. Rosen, "DSPCP: A Data Scalable Approach for Identifying Relationships in Parallel Coordinates," *IEEE Transactions on* Visualization *and Computer Graphics,* vol. 24, no. 3, pp. 1301-1315, 1 March 2018

[2] J. Johansson and C. Forsell, "Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 579-588, 31 Jan. 2016

[3] H. Guo, H. Xiao and X. Yuan, "Scalable Multivariate Volume Visualization and Analysis Based on Dimension Projection and Parallel Coordinates," *IEEE* Transactions *on Visualization and Computer Graphics,* vol. 18, no. 9, pp. 1397-1410, Sept. 2012

[4] G. Ellis and A. Dix, "Enabling Automatic Clutter Reduction in Parallel Coordinate Plots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 717-724, Sept.-Oct. 2006

[5] J. H. T. Claessen and J. J. van Wijk, "Flexible Linked Axes for Multivariate Data Visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2310-2316, Dec. 2011

[6] J. Blaas, C. Botha and F. Post, "Extensions of Parallel Coordinates for Interactive Exploration of Large Multi-Timepoint Data Sets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1436-1451, Nov.-Dec. 2008

[7] T. N. Dang, L. Wilkinson and A. Anand, "Stacking Graphic Elements to Avoid Over-Plotting," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1044-1052, Nov.-Dec. 2010

[8] R. G. Raidou, M. Eisemann, M. Breeuwer, E. Eisemann and A. Vilanova, "Orientation-Enhanced Parallel Coordinate Plots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 589-598, 31 Jan. 2016

[9] A. Dasgupta and R. Kosara, "Pargnostics: Screen-Space Metrics for Parallel Coordinates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1017-1026, Nov.-Dec. 2010

[10] R. C. Roberts, R. S. Laramee, G. A. Smith, P. Brookes and T. D'Cruze, "Smart Brushing for Parallel Coordinates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 3, pp. 1575-1590, 1 March 2019

[11] L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 6, pp. 1997-2010, 1 June 2018

[12] S. Niijima and Y. Okuno, "Laplacian Linear Discriminant Analysis Approach to Unsupervised Feature Selection," *IEEE/ACM*

*Transactions on Computational Biology and Bioinformatics*, vol. 6, no. 4, pp. 605-614, Oct.-Dec. 2009

[13] Q. Huang, D. Tao, X. Li, L. Jin and G. Wei, "Exploiting Local Coherent Patterns for Unsupervised Feature Ranking," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B (Cybernetics), vol. 41, no. 6, pp. 1471-1482, Dec. 2011

[14] R. Huang, W. Jiang, G. Sun, "Manifold-based constraint laplacian score for multi-label feature selection Pattern Recognit," *Lett.*, pp. 346-352, 2018

[15] A. Alalga, K. Benabdeslem, N. Taleb. "Soft-constrained Laplacian Score for Semi-Supervised Multi-Label Feature Selection," *Springer-Verlag*, pp. 75-98, 2016

[16] J. Lee, D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information Pattern Recognit," *Lett.*, pp. 349-357,2013

[17] P. B. Brazdil, C. Soares, "A comparison of ranking methods for classification algorithm selection." *11th European Conference on Machine Learning*, pp. 63-75, 2000

[18] L. Feng, J. Zhou, S.-L. Liu, N. Cai, and J. Yang, "Analysis of journal evaluation indicators: an experimental study based on unsupervised Laplacian score," *Scientometrics*, vol. 124, no. 4, pp. 233–254, 2020.

[19] N.-Q. Doan, H Azzag and M Lebbah, "Hierarchical Laplacian Score for unsupervised feature selection", *International Joint Conference on Neural Networks (IJCNN)*, pp. 1-7, 2018

[20] J. Liang and D. Huang, "Laplacian-Weighted Random Forest for High-Dimensional Data Classification," *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, Xiamen, China, pp. 748-753, 2019

[21] J.-Q. Wang, Y.-H. Li and J.-F. Chen, "Label Reconstruction based Laplacian Score for semi-supervised feature selection," *International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, Shengyang, pp. 1112-1115, 2013

[22] L.-P Yang, W.-G Gong, X.-H. Gu, W.-H. Li, X. Du, "Complete Discriminant Locality Preserving Projections for Face Recognition," *Journal of Software*, vol. 21, pp. 1277, 2010.

[23] M. Belkin, P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," *Advances in Neural Information Processing Systems*, MA, Cambridge: MIT Press, vol. 14, 2002.

[24] K. R. Kavitha, K. Neeradha, Athira, K. Vyshna and S. Sajith, "Laplacian Score and Top Scoring Pair Feature Selection Algorithms," *2020 Fourth International* Conference *on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 214-219, 2020

[25] X. He, D. Cai, and P. Niyogi. "Laplacian score for feature selection." *Advances in* Neural *Information Processing Systems 18,* 2005

[26] P. Padungweang, C. Lursinsap and K. Sunat, "Univariate Filter Technique for Unsupervised Feature Selection Using a New Laplacian Score Based Local Nearest Neighbors," *2009 Asia-Pacific Conference on Information Processing*, Shenzhen, pp. 196-200, 2009

[27] K. Benabdeslem and M. Hindawi, "Efficient Semi-Supervised Feature Selection: Constraint, Relevance, and Redundancy," *IEEE Transactions on Knowledge and* Data *Engineering*, vol. 26, no. 5, pp. 1131-1143, May 2014

[28] J. Xu, B. Tang, H. He and H. Man, "Semisupervised Feature Selection Based on Relevance and Redundancy Criteria," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 1974-1984, Sept. 2017

[29] B. Rong, Y. Qian, K. Lu, H. Chen and M. Guizani, "Call Admission Control Optimization in WiMAX Networks," IEEE Transactions on Vehicular Technology, vol. 57, no. 4, pp. 2509-2522, July 2008.

[30] N. Chen, B. Rong, X. Zhang and M. Kadoch, "Scalable and Flexible Massive MIMO Precoding for 5G H-CRAN," in IEEE Wireless Communications, vol. 24, no. 1, pp. 46-52, February 2017.

[31] B. Rong, Y. Qian and K. Lu, "Integrated Downlink Resource Management for Multiservice WiMAX Networks," in IEEE Transactions on Mobile Computing, vol. 6, no. 6, pp. 621-632, June 2007.

[32] S. Sun, M. Kadoch, L. Gong and B. Rong, "Integrating network function virtualization with SDR and SDN for 4G/5G networks," IEEE Network, vol. 29, no. 3, pp. 54-59, May-June 2015.

[33] Y. Wu, B. Rong, K. Salehian and G. Gagnon, "Cloud Transmission: A New Spectrum-Reuse Friendly Digital Terrestrial Broadcasting Transmission System," IEEE Transactions on Broadcasting, vol. 58, no. 3, pp. 329-337, Sept. 2012.

[34] Z. Zhang et al., "6G wireless networks: Vision, requirements, architecture, and key technologies," IEEE Veh. Technol. Mag., vol. 14, no. 3, pp. 28-41, Sep. 2019.

[35] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," IEEE Netw., vol. 34, no. 3, pp. 134-142, May/June 2020.

[36] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," IEEE Commun. Mag., vol. 58, no. 3, pp. 55-61, Mar. 2020.