Assignment 3 – Paper Critique

Name: Pavan Balaji Kumar

Student Number: 300169572

The paper : https://link.springer.com/article/10.1007/s10994-018-5765-6

1. What is the paper about? (Research problem)

Feature selection is one of the main methods used in the pre-processing step to reduce a dataset's dimension. It selects the critical features that are required to predict the output of a model. Feature selection reduces the noise associated with the data and eliminates the redundant features present in the dataset. Supervised, semi-supervised and unsupervised are the significant types of feature selection. The paper "**Unsupervised feature selection based on kernel fisher discriminant analysis and regression learning"** describes a new unsupervised feature selection technique.

2. Why is it important? (Application and relevance)

As the amount of data produced increases every day, the information (features) available to solve the problem has also increased. This gives rise to a problem known as "**The curse of Dimensionality**." When a particular dataset has many features, some features may be useless in predicting a machine learning model's output and redundant. This negatively impacts the model's performance and the time complexity of the optimization function. Features selection helps in getting rid of the unimportant features and reduces the dimensions of the dataset. Reducing the number of features to consider reduces the training time significantly. Having a small number of features may also increase the explainability of the model.

3. How does the paper extend the state-of-the-art? (Novelty)

Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Joint Embedding Sparse Regression Analysis (JELSER) are some of the unsupervised feature selection and dimensionality reduction methods that are widely used and efficient. These techniques often use either **Manifold learning**, which preserves the non-linear property of the data or **Discriminative learning**, only focusing on the linear projection of data. As both these techniques have their shortcomings, the paper proposes a new method**, Kernel fisher discriminant analysis and regression learning (KFDRL).** This method uses both the manifold and discriminant information by joining global discriminant analysis with spectral clustering and regression. This enables the algorithm to better separate data than the other techniques already present, resulting in better performance than the others.

4. How does the proposed approach address the research problem? (Methodology)

The feature selection process starts by combining Linear Discriminant analysis and spectral clustering. The method uses a scaled cluster assignment matrix **H,** obtained from the spectral embedding matrix, to group the dataset into the desired number of clusters. This step ensures the preservation of the manifold information of the dataset. A between-cluster scatter matrix and within-cluster scatter matrix is defined to ensure that different clusters are as far apart as possible, and the data points in the same cluster are tighly packed. The LDA uses a Gaussian kernel **K** to ensure the technique can work on both linear and non-linear data. The dataset is then centred using **Cn.** This combined model is defined by minH $tr($ HTGH) where $\boldsymbol{G} = \boldsymbol{C}n - \boldsymbol{C}Tn(\boldsymbol{C}n + \mu\boldsymbol{K}-1)-1\boldsymbol{C}n$.

Secondly, there is a regression term that is used to convert the datapoint to their embeddings. This transformation is done using **W,** a transformation matrix in which each row represents each feature in the dataset. These row vectors help us in selecting the essential features. When $L_{2,1}$-*norm* regularisation is applied to help the algorithm fit and generalize better also makes the transformation matrix **W** sparse aiding the selection of more discriminant features. H and **W** are calculated in an unsupervised manner using spectral clustering and regression. The features are selected by scoring the row vectors of **W** and ranking them in descending order. The importance of the feature is determined by how high it ranks. The required number of features can be selected from the highest-ranking features.

5. What are the main conclusions? (Results.)

In the paper, the proposed algorithm's performance was compared against well-known algorithms like JELSR, LapScore, SPEC, MRSF, MCFS, DFSC and LSPE against ten datasets given in Table 2 in the paper. The convergence rate of the algorithm was tested on four datasets from UCI. From Fig 1. It was seen that the value of the cost function decreased rapidly in the first three iterations across all four datasets, indicating the increased efficiency KFDLR algorithm to find an optimal solution for the objection function. When KFDRL was used to select essential features from the AT&T dataset for the face recognition task, it was observed that the images constructed using 4096(40% of total) features retains enough information for face recognition tasks. This indicated that the algorithm was capable of selecting the most important features while ignoring the redundant ones.

K-means clustering and Nearest Neighbourhood classification was performed with features selected using KFDRL and other algorithms mentioned above. Metrics like accuracy and NMI were used to compare the algorithms' performance. The KFDRL algorithm had higher accuracy in the clustering task than other algorithms for all datasets except the Sonar dataset and better NMI except for the Isolet and Umist dataset. According to the paper, the decrease in performance on specific datasets may be due to the dataset's underlying structure being discriminant. Since KFDRL uses both manifold and discriminant learning, the performance is affected when the dataset is strictly manifold or discriminant in structure. The classification accuracy obtained using the features selected by KFDRL was better than the other algorithms, even with a small number of features in most cases, as seen in table 6.

6. What are the limitations of the work? (Reflection on limitations and soundness of conclusions.)

Though KFDRL performs better than other state-of-the-art algorithms, it also has its shortcomings. Time complexity is one of the primary limitations. As seen in table 5, the KFDRL algorithm takes a lot more time to run than the other algorithm discussed in this paper. Tuning the hyperparameters for the objective function can be a very time-consuming task due to the long runtime. The algorithm's sensitivity towards large datasets and the optimization function getting stuck at local optima are also some of the limitations addressed in this paper.

7. How can this work be extended? (Your own suggestions of potential future work.)

Reducing the time complexity of the algorithm would be the best step to take moving forward. Having reduced runtime will make the use of this algorithm justifiable. Since the algorithm uses both manifold and discriminant information algorithm's performance, it took a hit when the data was of a particular structure. Extending the algorithm in a way that it can adapt to these structures and maintain the same performance will be a great advantage.

References:

[1] Shang, Ronghua, Yang Meng, Chiyang Liu, Licheng Jiao, Amir M. Ghalamzan Esfahani, and Rustam Stolkin. "Unsupervised Feature Selection Based on Kernel Fisher Discriminant Analysis and Regression Learning." *Machine Learning* 108.4 (2018): 659-86. Print.