From: **SHENG Rui** rshengac@connect.ust.hk
Subject: CSIC 5011: Project 1 Review [Rui SHENG-20787795]
Date: 19 April 2023 at 3:46 PM
To: datascience.hw@gmail.com

SR

(1)Qiqi Zhou
Summary:

This project is mainly to leverage machine learning combined with a dimension reduction technology (ie, PCA) to inpaint images. Additionally, the author tested the algorithm through the MNIST dataset. The robust result proves the efficiency of the proposed method.

Strengths:

The evaluation is very detailed and solid. First, the author discussed the performance of several concrete situations of the image inpainting. Then the author evaluated the robustness of the proposed method.

The hyperparameters are determined reasonably. Specifically, the author tested different hyperparameter setting and described the specific reason why the hyperparameter was set as a value.

Weaknesses:

The author just tested the efficiency of the proposed method in three kinds of damages: row, column, and cross damage. However, in the real world, more unregular damages should be considered.

The author did not display the performance number of the proposed method.

Scores:

Writing: 5.

Technology: 4. There is no performance evaluation of the proposed method. I think MSE or other performance metrics can be used.

Overall: 4.5.

Confidence: 3.

(2)Yingxue XU, Jiaxin ZHUANG, Fengtao ZHOU
Summary:

This project is mainly to leverage PCA to visualize the different classes of the US Crime Data. Specifically, the author has leveraged several methods (eg, SPCA, MDS) as the comparison methods.

Strengths:

The author used many methods (PCA, SPCA, MDS, ISOMAP, LLE) to compare the results, which provide rich context to help to understand the performance of the PCA.

The poster is very delicate, and the logic is smooth.

Weaknesses:

The author just relied on the figures to show the superiority of the PCA method compared with others. However, it will be more objective through objective statistical outcomes, such as purity and rand index.

The conclusion has a confusing point. The author proposed that the crime number was influenced by all the factors. However, there is no concrete experiment to test different factors. For example, I think the author should leveraged the controlled variable experiments to prove the necessity of all the factors.

Scores:

Writing: 4. The conclusion is a little confusing, which has been mentioned before.

Technology: 3.5. The comparison result is not solid. Besides, the author did not discuss the weaknesses and limitations of the proposed method.

Overall: 4.

Confidence: 3.

(3)Chris HC Nguyen and James M Shihua
Summary:

This project is mainly to use RPCA to extract the foreground and background from a video. Specifically, the videos from CCTV are used as the dataset, and the result has shown the efficiency of

the proposed method to detect and extract the moving objects from the background.

Strengths:

The method is very clear. The input and output of the algorithm is clearly presented, and pseudo-code was written. The whole process is very clear.

The topic is very interesting and challenging.

The experiment process, including iteration result and errors, is logically displayed.

The analysis of the results is very well organized and analyzes the possible causes of unexpected situations.

Weaknesses:

There is no discussion about the weaknesses and strengths of the proposed method.

Scores:

Writing: 5.

Technology: 4. There is no discussion about the weaknesses and strengths of the proposed method (the discussion part is not enough).

Overall: 4.5.

Confidence: 2.

(4)Ruochen MA; Jihong TANG; Yuyan RUAN; Zhi HUANG

Summary:

This project is mainly to test the performance of seven different dimensionality reduction methods through the SNPs dataset. Specifically, the author has proved the superiority of the random projection method compare with other six approaches.

Strengths:

The evaluation is solid. The author compared the results generated through PCA, MDS, t-SNE, ISOMAP, LLE, UMAP, robust PCA and random projection. Specifically, the author proved the superiority through quantitative method.

Weaknesses:

The author did not describe the reason why to validate the comparison results with a split of 7. Is the conclusion the same in other cases, such as 6, 8 or 9? Or is it most semantically effective to divide into 7 regions?

Scores:

Writing: 4. I am confused about the region number selection.

Technology: 4. The author did not discuss the weaknesses and limitations of the proposed method. For example, why to choose ARI. Is there any limitations to use this index?

Overall: 4.

Confidence: 3.

(5)CAI Bibi; QIU Zhenyu; WANG Zhiwei*

Summary:

This project is mainly to employs statistical topic modeling to analyze the NIPS words dataset and uncover the key themes discussed in the papers. Specifically, the authors apply the latent Dirichlet allocation (LDA) model to simulate the generation of each word in the document, and then infer the topics based on the weights of individual words. Additionally, the study utilizes various clustering and dimension reduction techniques, including K-means, MDS, and tSNE, to investigate the relationships between topics and papers, as well as to trace the trends and popularity of different topics over time.

Strengths:

The explanation of Latent Dirichlet Allocation and the process of fitting the LDA Model are elucidated with lucid equations and illustrations.

Weaknesses:

The method is not novel. The approach employed in this study adopts a bag-of-words model, which is known for its limitations in accurately modeling natural language. As a result, the findings regarding the trends of topics should be interpreted with caution.

Scores:

Writing: 5.

Technology: 4. The approach is not convincing. The state-of-the-art should be considered.

Overall: 4.5.

Confidence: 2.