Reviewer: LAI Yanming

12. DUNDA, Gerry Windiarto Mohamad.  
Statistical Analysis on Authors and Word Trend of NIPS Papers from 1987 to 2017.

Summary:

The research analyzed NIPS papers from 1987 to 2017 and identified trends in the

usage of keywords, research themes, and publication patterns. The research found

that the word usage of machine learning has undergone significant changes in focus

and emphasis over the past three decades, with some words becoming more popular

than others. Additionally, the study found a weak positive correlation between

community size and publication count.

Strengths:

1. The authors analyze NIPS papers from 1987 to 2017 using statistical techniques to identify word usage trends and community detection within the research community. By analyzing the frequency of keywords and topic modeling techniques, we uncover the prominent research themes over the years and track how they have evolved.
2. The investigation indicate the importance of considering community size when studying publication patterns in scientific communities and highlight the benefits of using advanced community detection methods like Leiden for more precise and efficient analysis of complex networks.
3. The authors provide clear visulizations and conduct a detailed analysis.

Weaknesses:

1. The authors could add short descriptions to MDS and mention its advantages and drawbacks for the sake of thoroughness.
2. Maybe the authors can also apply some other data-reduction methods and make a comparison with MDS.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality: 4

Overall rating: 4

Confidence: 3

13. LI Haobo, CHEN Zixin, TENG fei, SHENG Rui.  
NIPS Conference Papers 1987-2015 Data Set.

Summary:

The impact of academic writing style on paper acceptance is a

major concern for researchers. There has been a noticeable shift in

language usage over the past few decades, making it essential for

researchers to stay up-to-date with these changes and adapt their

writing style accordingly. However, the sheer volume of research papers across

various areas makes it impractical to analyze the progression of

word usage. To address this issue, the authors propose a visual analytics

approach aimed at helping researchers explore the evolution of

word usage across time in accepted papers.

Strengths:

1. The study is well visualized, making the results very intuitive and easy to understand.
2. The authors give a detail analysis to the experiment results.

Weaknesses:

1. It would be better to add a section for conclusion at the end of the poster.
2. The authors could add short descriptions to the methods they apply and mention their advantages and drawbacks for the sake of thoroughness.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality: 4

Overall rating: 4

Confidence: 3

1. CAI Bibi; QIU Zhenyu; WANG Zhiwei.  
   Topic Modeling For NIPS Words.

Summary:

The authors perform statistical topic modeling to analyze the NIPS words dataset. Specifically, they use the latent Dirichlet allocation (LDA) to introduce the latent variable topic to model the generative process of each word in the document and then figure out the topics based on the weights of each word inner each topic. They further use clustering and dimension reduction methods, such as K-means, MDS, and tSNE, to analyze this latent representation to gain more insights into this dataset.

Strengths:

1. The authors use LDA to obtain the word distribution as the representation of the latent topics, and the topic proportion as the representation of each paper. With the topic representation and topic proportion, they figure out the meaning of each topic and analyze the papers’ topics. They further use clustering and dimension reduction methods, such as K-means, MDS, and tSNE, to analyze the output of the LDA to gain more insights into this dataset.
2. The authors present sufficient visulizations and give a comprehensive analysis to the experiment results.

Weaknesses:

1. The authors could add short descriptions to the clustering and dimension reduction methods they apply and mention their advantages and drawbacks for the sake of thoroughness.
2. Maybe it would be better to add a brief section for conclusion at the end of the report.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality: 5

Overall rating: 5

Confidence: 3

1. LI Qichao and HUANG Haohan.  
   Exploration of PCA family for handwritten digit classification.

Summary:

The principal component analysis (PCA) is usually utilized as a method to undergo dimension reduction, projecting high-dimensional dataset into a low-dimensional affine space. In this mini-project, the authors try to compare the performance of different kinds of PCA, including original PCA, sparse PCA and kernel PCA with different kinds of kernel functions, by utilizing the data reduced through these methods to carry out the classification tasks. Here the authors select the random forest classifier considering its simplicity. Their experiments show that PCA is an effective tool for dimension reduction. They perform RF classifier on the data reduced by different PCA and conclude that KPCA with linear kernel function shows the best performance in handwritten digits dataset.

Strengths:

1. Many pictures and tables are used to visulize the experiment result, making the report easy to understand.
2. The authors compare the performance of different kinds of PCA, including original PCA, sparse PCA and kernel PCA with different kinds of kernel functions, by utilizing the data reduced through these methods to carry out the classification tasks.
3. The authors gives a detailed introduction to PCA and its variant methods.

Weaknesses:

1.The experiment of reconstructing image by using PCA components is lacked. Hence the ability of PCA that reducing the dimensionality of the dataset while preserving the essential information in the data is not sufficiently explored.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality: 4

Overall rating: 4

Confidence: 3

1. JIA Guangnan.  
   EXPLORING THE EFFECTIVENESS OF PCA ON HANDWRITTEN DIGIT DATASET.

Summary:

The authors explore the effectiveness of PCA on the handwritten digit dataset. This project investigate how well PCA can reduce the dimensionality of the dataset while preserving the essential information in the data and the application of PCA, K-means clustering, and logistic regression to handwritten digit recognition. They compute the reconstruction error of the PCA model on the test data using the mean squared error metric. The reconstructed digit images are visually similar to the original images, even when using a relatively low number of principal components, indicating that PCA can capture the essential information in the data and use it to reconstruct the original images.

Strengths:

1.The authors systematically explored the impact of different numbers of PCA components on the quality of reconstructed images, the reconstruction error, and the performance of K-means clustering and logistic regression. They find that selecting an appropriate number of PCA components is crucial to balancing the trade-off

between preserving information and reducing dimensionality while minimizing the reconstruction error.

2.They compare the performance of K-means clustering and logistic regression on the reduced dataset with different numbers of PCA components and show that logistic regression consistently outperforms K-means clustering in terms of classification accuracy

Weaknesses:

1. The authors could add short descriptions to the K-means clustering and logistic regression and mention their advantages and drawbacks for the sake of thoroughness.
2. It would be better for the authors to give references to the different methods they have used in the project.
3. Maybe the authors can give a more detailed description to the image reconstruction part.

Evaluation on Clarity and quality of writing: 4

Evaluation on Technical Quality: 5

Overall rating: 4

Confidence: 3