

05. SHEN, Xinwei and YANG, Yunfei. Representation learning on gene expression data.

Summary of the report.

In this report, they explore three levels of representation learning methods: (i) linear dimension reduction methods: PCA and robust PCA, (ii) non-linear dimensionality techniques: manifold learning, and (iii) deep representation methods: VAE, on a single-cell gene expression dataset. They implement various methods and stress on a common problem in single-cell expression analysis, cell clustering, to demonstrate their representation power. In conclusion, manifold learning and VAE outperform linear dimensionality reduction approaches on this dataset.

Describe the strengths of the report.

Various methods are used and compared.

Describe the weaknesses of the report.

N.A.

Evaluation on quality of writing (1-5):

4

Evaluation on presentation (1-5):

4

Evaluation on creativity (1-5):

4

Confidence on your assessment (1-3):

2

01. SHEN_YANG_report

1. Summary

The report used different methods to realize the representation learning on gene expression data. three levels of representation learning methods: (i) linear dimension eduction methods: PCA and robust PCA, (ii) non-linear dimensionality techniques: manifold learning, and (iii) deep representation methods: VAE, on a single-cell gene expression dataset. The found that manifold learning and VAE outperforms linear methods.

2. Describe the strengths of the report.

A bunch of algorithms were studied and clear presentation of the results were illustrated.

3. Describe the weaknesses of the report.

No obvious weakness was found

4. Evaluation on quality of writing (1-5): 5

The report is clearly written and well organized.

5. Evaluation on presentation (1-5): 4

6. Evaluation on creativity (1-5):4

7. Confidence on your assessment: 2

Group 5

Summary of the report

Compare 3 representation learning methods: PCA and robust PCA, manifold learning and deep representation methods.

Strength

Had three levels of learning methods tested and compared, and combine representation methods with clustering methods.

Weakness

The understanding of the dataset itself is limited, but it's understandable their focus is on the technical part.

Evaluation on quality of writing (1-5): 4

Maybe a table of final score could be better present the performance than seperately writing in paragraph.

Evaluation on quality of presentation (1-5): 4

Maybe a screen recording application will help with the video quality;

Evaluation on quality of creativity (1-5): 5

It's impressive to see a broader choice of representation learning methods from 3 levels.

Confidence on your assessment (1-3): 3

Comment on paper 5

In paper 5, Representation learning on gene expression data, the author explored three levels of representation learning methods: (i) linear dimension reduction methods: PCA and robust PCA, (ii) non-linear dimensionality techniques: manifold learning, and (iii) deep representation methods: VAE, on a single-cell gene expression dataset. They implement various methods and stress on a common problem in single-cell expression analysis, cell clustering, to demonstrate their representation power and give detailed analysis and comparison of different approaches as well as different settings of hyperparameters. In conclusion, manifold learning and VAE outperform linear dimensionality reduction approaches on this dataset.

The author describes the methods and the reasons for considering them. They demonstrate the utility of various methods on real data and analyzes the results comprehensively in the end.

Strength: The strength of this paper is the author compared three different types of clustering and each of them contains some methods. The result is clear and convinced.

Weakness: The weakness of this paper is the pictures and charts are not so rich.

Evaluation on quality of writing: 4. The writing is clear. Pictures and charts are used in this paper.

Evaluation on presentation: 4. The paper is well organized and clear.

Evaluation on creativity: 3.

Confidence on your assessment: 3.

Representation learning on gene expression data

- **Summary of this report:** Three levels of representation learning methods are explored. The detailed analysis and comparison of different approaches as well as different settings of hyperparameters are also given in the report.
- **Describe the strengths of the report:** This report clearly shows that the comparisons of different approaches. The analysis are well thought out and convincing.
- **Describe the weaknesses of the report:** Some descriptions of the figures in this report are not clear.
- **Evaluation on presentation: 5**
- **Evaluation on Clarity and quality of writing (1-5): 4**

Details of typos:

(1) The most commonly used dimensionality reduction technique **include(s)** the principal component analysis (PCA),

- **Evaluation on creativity (1-5): 5**

The PCA is correctly used. Moreover, the report is well reported.

- **Overall ratings: 4.5**
- **Confidence on your assessment: 2**

- Representation learning on gene expression data

Summary:

Compare different representation learning methods, the linear, nonlinear and neural network-based algorithms.

Strength of the project:

Very comprehensive comparison covering 3 major types of dimensionality reduction method. Very clear explanation

Weakness of the project:

Regarding Figure 4, the VAE's ARI should also be added to the graph as a fair comparison resembling Figure 3.

Evaluation on Clarity and quality of writing (1-5):	5
Evaluation on Technical Quality (1-5):	4.5
Overall rating:	5
Confidence on your assessment:	2

05. SHEN, Xinwei and YANG, Yunfei. Representation learning on gene expression data

- **Summary:** This report explores three levels of representation learning methods: (1) linear dimension reduction methods: PCA and robust PCA, (2) non-linear manifold learning and (3) deep representation methods: VAE on single-cell gene expression dataset. They conclude that VAE and manifold learning outperform linear models.
- **Strength:** This report performs intensive experiments and provides insightful analysis trying to explain the result.
- **Weakness:** It would be better if the authors can provide more details about the evaluation metric.
- **Evaluation:**

	Writing	Presentation	Creativity	Confidence
Score	5	5	5	2

4 Representation learning on gene expression data

4.1 Summary

Shen Xinwei et al. construct a three-level study on gene expression data included PCA and r PCA, manifold learning and VAE.

4.2 Strength and Weakness

The strength is the comparison between different level of study, and raise a concept of showing their representation power.

4.3 Score

4.3.1 Clarity and Quality of Writing

I appreciate their effort on explaining all the important concept in their report, even unprofessional people can understand their report. The structure of the report is clear apparently. I will give them 5/5 on this aspect.

4.3.2 Presentation

Clear presentation. 5/5

4.3.3 Creativity

They compare the representation power of three levels of algorithm. 5/5

Overall

5/5

05.SHEN_YANG_report

Summary:

In this report, three levels of representation learning methods: (i) linear dimension reduction methods: PCA and robust PCA, (ii) non-linear dimensionality techniques: manifold learning, and (iii) deep representation methods: VAE, were performed on a single-cell gene expression dataset. It shows their representation power in the cell clustering.

Strengths:

Different clustering methods are applied to get a comprehensive analysis. And the clustering results are vivid.

Evaluation on Clarity and quality of writing (1-5):	5
Evaluation on Technical Quality (1-5):	4
Overall rating:	5
Confidence on your assessment:	2

05. SHEN, Xinwei and YANG, Yunfei. Representation learning on gene expression data.

In this work, author explored the property of linear and non-linear dimension reduction methods and one deep learning method on data clustering and visualization. Results visualized the clustering results in 2D and demonstrated that the neural network method VAE outperformed the dimension deduction methods.

Strengths: This work is well organized, and author conducted enough evaluation on method comparison and visualization.

Weakness: The experiments are designed with weak association, the linear methods PCA and RPCA are only discussed in eigenvalue cumulation, while not included in clustering comparison.

Evaluation on quality of writing (5): This report is well written with adequate references, and the figures demonstrated the results clearly.

Evaluation on presentation (4): Good presentation conveys authors' idea clearly.

Evaluation on creativity (3): As discussed in Weakness part, although authors conducted enough experiments, the association between experiments are not strong, and the idea of experiment design is lack of creativity.

Confidence on your assessment(2)