CSCM35: Big Data and Data Mining Coursework 1

Andy Gray
445348

06/05/20

1 Summary of Concept (or Problem Statement)

Your two page summary must contain a conceptual summary of the assigned paper. Which category of vector field visualization does this paper fall into? What are the authors trying to achieve at a conceptual level? What is the goal of the work?

2 Contributions

What is (are) the contribution(s) of the paper? Are they given explicitly or do they have to be extracted by the reader? In other words, what is the novelty? What do the authors present that is new (never been presented prior to the publication of this paper)?

3 Related Work Summary

What is the most important piece of related work relevant to the given paper? In other words, what previous research does this build on? The current work may build on more than one previous research paper.

4 Summary of Implementation

Your summary contains a brief description of theimplementation. You are not expected to under-stand all of the implementation detail of any paper but the basic strategy is conveyed.

5 Summary of Results

The summary contains a brief description of the results. Are the performance times given? Are data set sizes provided? What are the characteristics of the data set tested?

6 Analysis and Discussion

What are the disadvantages of the proposed method? Does it have any flaws or shortcomings? Does it have any limitations? This information is more difficult to extract from the paper as the authors may being trying to hide it.

7 References

Your two page summary contains references and uses citations. It cites the paper you have been given. You are encouraged to use images and/or figures in your summary.

8 Conclusion

References

[1] Agrawal, R., Imieliński, T., and Swami, A. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (1993), pp. 207–216.

[2] Agrawal, R., Srikant, R., et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (1994), vol. 1215, pp. 487–499.

[3] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., and Varoquaux, G. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning* (2013), pp. 108–122.

[4] Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.

[5] Han, J., Pei, J., and Kamber, M. *Data mining: concepts and techniques*. Elsevier, 2011.

[6] Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in science & engineering 9*, 3 (2007), 90–95.

[7] jingjingslides. jingjingslides slides. In *Slide titles* (2020).

[8] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[9] Python Core Team. *Python: A dynamic, open source programming language*. Python Software Foundation, Vienna, Austria, 2020.

[10] Raschka, S. Mlxtend: Providing machine learning and data science utilities and extensions to python's scientific computing stack. *The Journal of Open Source Software 3*, 24 (Apr. 2018).

[11] Tan, P.-N., Steinbach, M., and Kumar, V. *Introduction to data mining*. Pearson Education India, 2016.