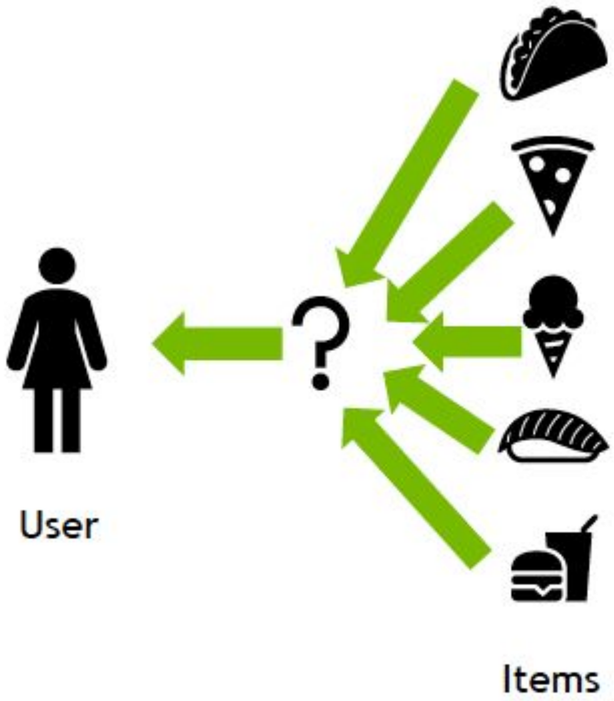
Topic Analysis in Recommendation Systems

By John Bogacz & Nicholas Pang

What are Recommendation Systems

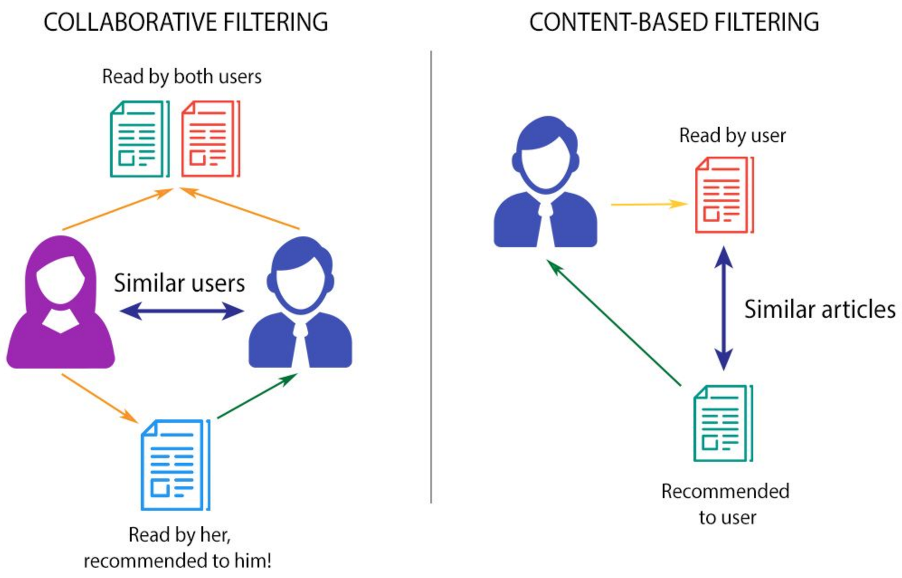


Motivation

Online services provide a plethora of options

- Filtering takes time, and becomes harder the larger the dataset is
- Humans may not recognize their own patterns

Background on Recommendation Systems



Existing solutions

- Collaborative Filtering
  - Users with similar activity are likely to have similar preferences
- Content-Based Filtering
  - Users are likely to like items similar to their previous activity
- Hybrid Filtering
  - A combination of collaborative and content-based

Challenges & problems

- Cold Start
- Latency
- Sparsity
- Diversity
- Scalability

[3],[11],[12]

Matrix Factorization

	Feature 1	Feature 2
User 1	?	?
User 2	?	?
User 3	?	?
User 4	?	?
User 5	?	?

x

	Item 1	Item 2	Item 3	Item 4	Item 5
Feature 1	?	?	?	?	?
Feature 2	?	?	?	?	?

=

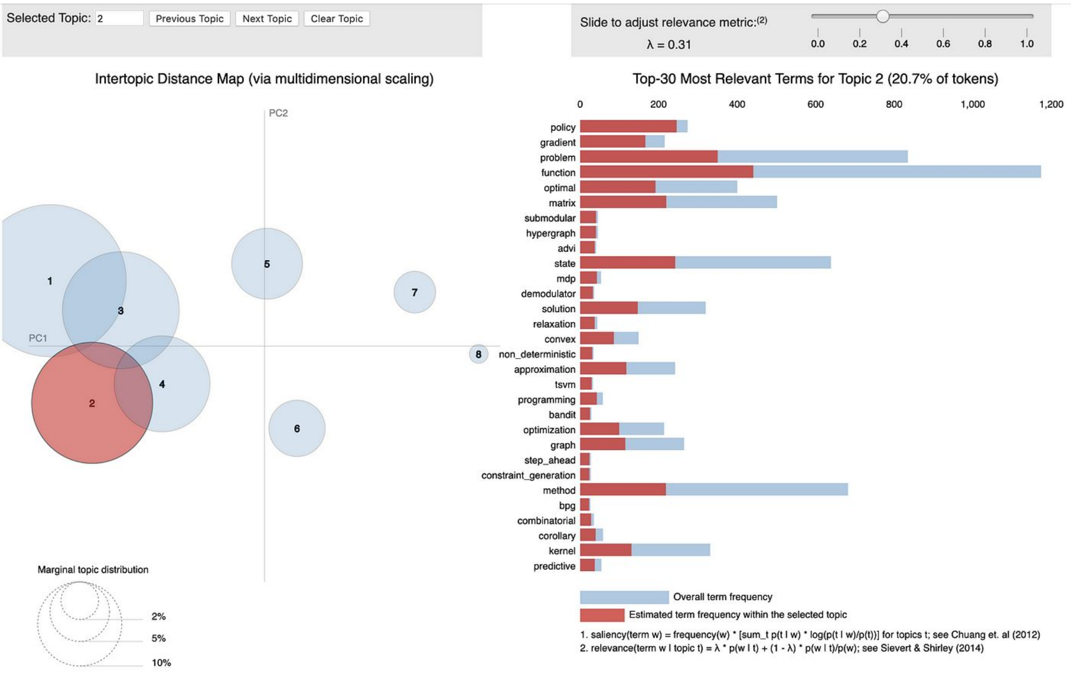
	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0?	3	0?	3	0?
User 2	4	0?	0?	2	0?
User 3	0?	0?	3	0?	0?
User 4	3	0?	4	0?	3
User 5	4	3	0?	4	0?

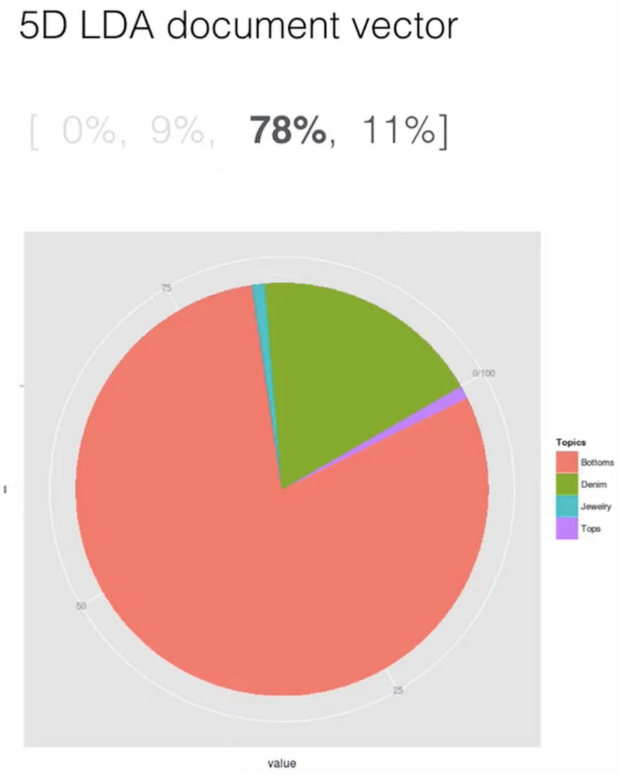
$$R(u,i) = p_u \cdot q_i^T + u_b + i_b + g_b$$

- p_u is the embedding of user u from P.
- q_i is the embedding of item i from Q.
- u_b and i_b are user and item biases.
- g_b is the global bias.

[1],[2],[5]

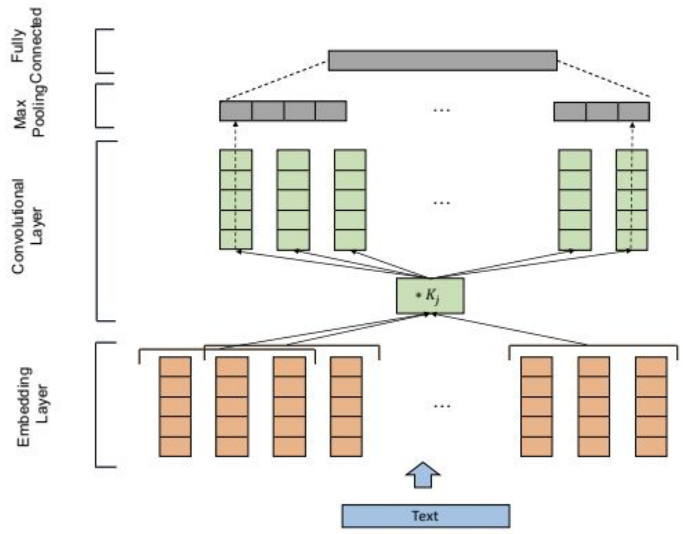
Latent Dirichlet Allocation (LDA)





[10]

Neural Attentional Rating Regression with Review-level Explanations (NARRE)



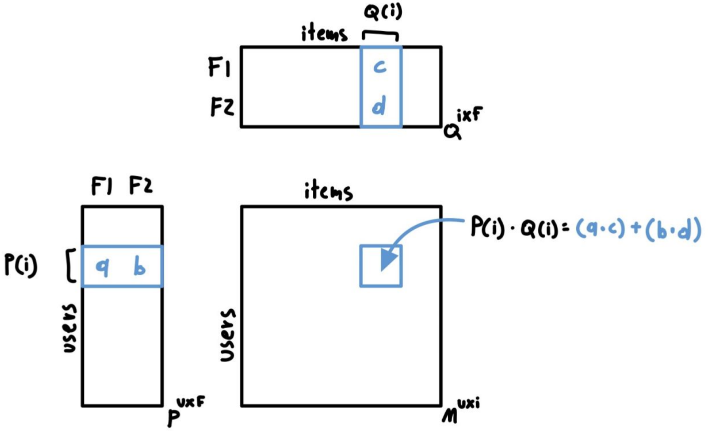
Reviews play an active role in predicting ratings

- Uses attention mechanisms & text processing models to provide rankings for reviews
- Those rankings help determine the weight of the review in rating calculation

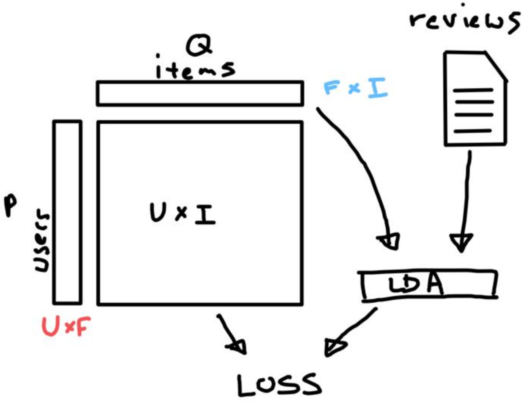
[9]

New Design & Components

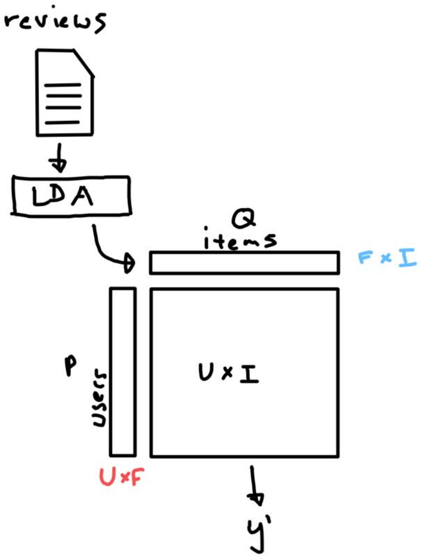
Matrix Factorization (original)



Regularization



Weight Initialization



[2]

Important Formulas

$$L(u,i) = (R(u,i) - y)^2$$

MSE

- Loss function used for regular MF and MF with LDA weight initialization

$$L(u,i) = (R(u,i) - y)^2 - \alpha \ell(T|\theta, \phi)$$

MSE w/ LDA

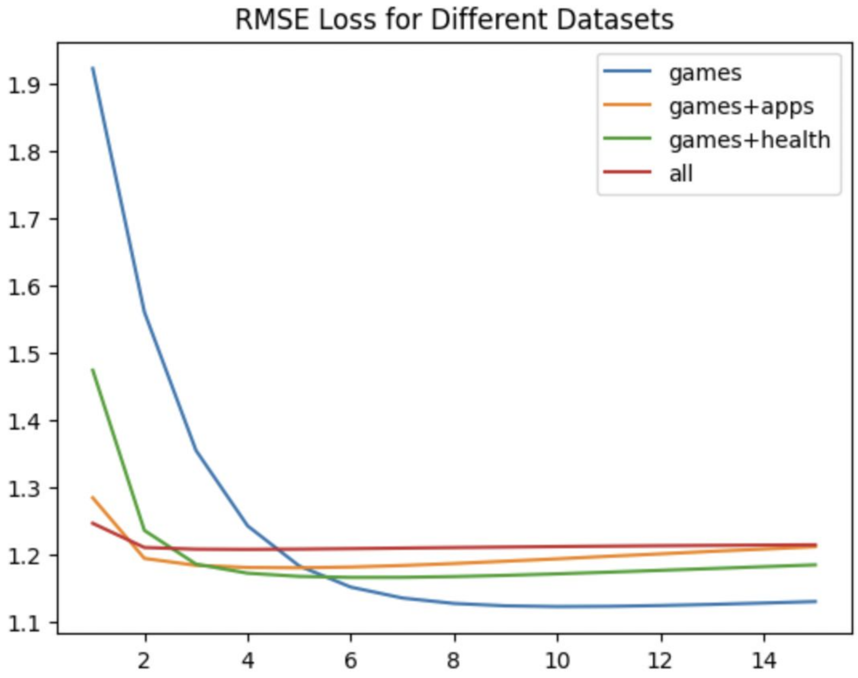
- ℓ is the likelihood of corpus T given the topic distribution per document θ and word distribution per topic ϕ
- α is a regularization parameter for the likelihood
- Only used in gradient descent

Dataset

TABLE I
SIZES OF PROPOSED DATASETS

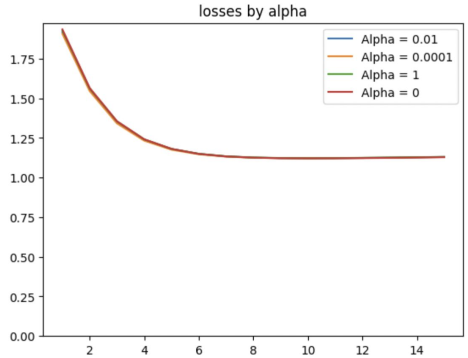
Category	#Users	#Items	#Reviews
Video Games	24303	10672	231780
Android Applications	87271	13209	752937
Health and Personal Care	38609	18534	346355

Results - LDA on Different Datasets



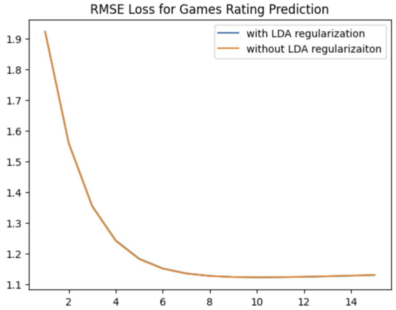
- More diverse datasets led to worse losses overall
- Combining datasets tended to lead to faster convergence
- Combining games with health led to better results than games with apps

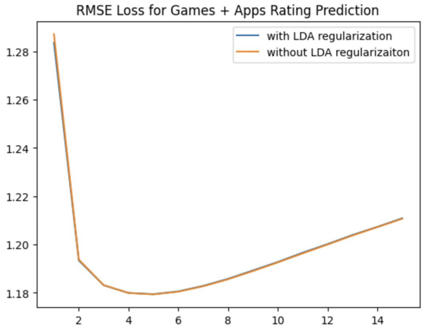
Results - LDA as Regularization

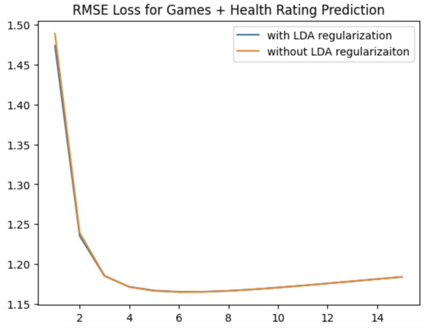


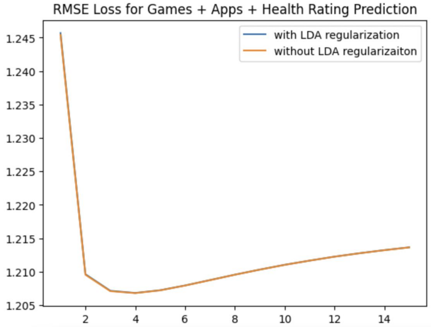
LDA as a regularizer does not appear to impact overall loss

- Changing alpha value does not lead to faster or better convergence

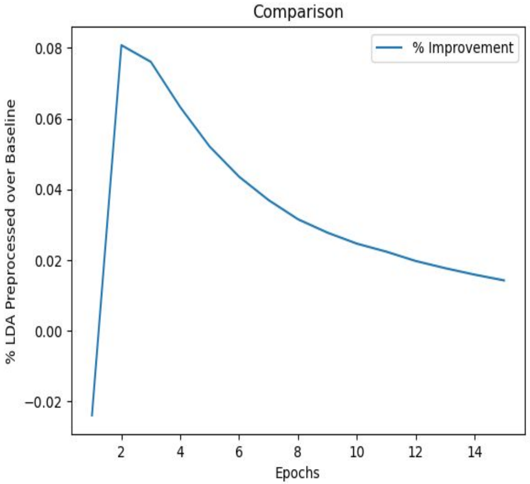


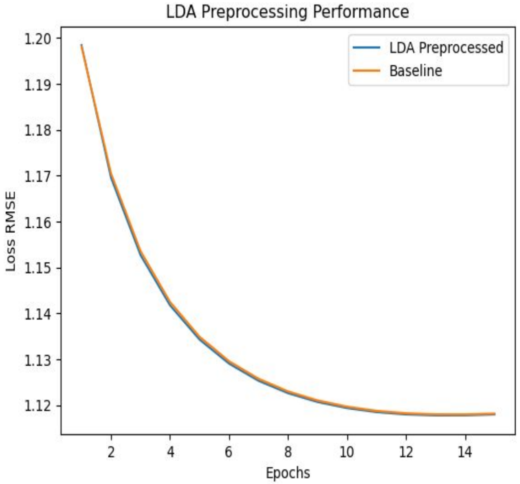


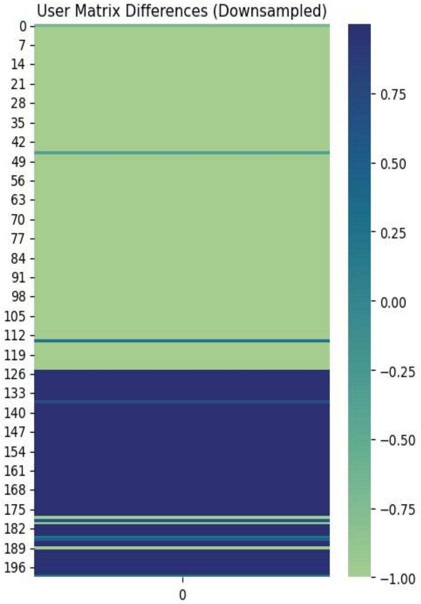


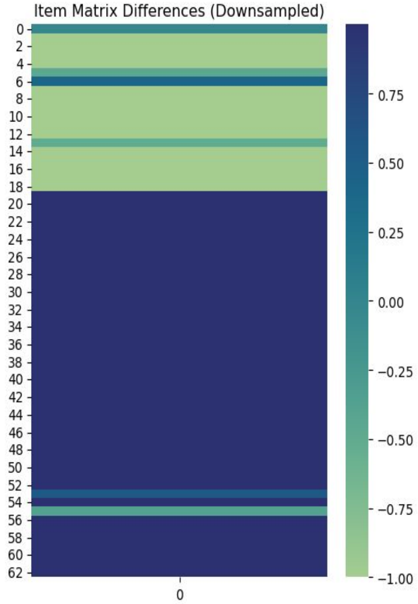


Results - LDA as Weight Initialization









- Faster convergence
- No significance improvements on overall accuracy in the long-term

Discussion

Pros

- LDA can help reach faster convergence
- Weight initialization is faster than regularization

Cons

- Time complexity increase with LDA is significant
- No loss improvements

Potential Future Directions

- More features
    - Since time is a limiting factor, running more diverse datasets with more latent factors was difficult. Combining the model with multithreading and GPUs for more optimal parameters may reveal different results
- Different LDA's for users and items
    - Users and items may have different subsets of relevant terms. Having a different LDA model for each may yield better corpus likelihood and potentially more accurate results.
- Content-based analysis rather than review-based
    - Creating topic distributions from item descriptions may be more accurate than creating them from reviews.

Work Cited

[1] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural Attentional Rating Regression with Review Level Explanations," Proceedings of the 2018 World Wide Web Conference on World Wide Web WWW '18. ACM Press, 2018.

[2] J. McAuley and J. Leskovec, "Hidden factors and hidden topics," Proceedings of the 7th ACM conference on Recommender systems. ACM, Oct. 12, 2013.

[3] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," Journal of Big Data, vol. 9, no. 1. Springer Science and Business Media LLC, May 03, 2022.

[4] X. Wang, Y. Chen, J. Yang, L. Wu, Z. Wu, and X. Xie, "A Reinforcement Learning Framework for Explainable Recommendation," 2018 IEEE International Conference on Data Mining (ICDM). IEEE, Nov. 2018.

[5] D. Wu, X. Luo, M. Shang, Y. He, G. Wang, and M. Zhou, "A Deep Latent Factor Model for High Dimensional and Sparse Matrices in Recommender Systems," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 7. Institute of Electrical and Electronics Engineers (IEEE), pp. 4285–4296, Jul. 2021.

[6] L. Zheng, V. Noroozi, and P. S. Yu, "Joint Deep Modeling of Users and Items Using Reviews for Recommendation." arXiv, 2017.

[7] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," Discourse Processes, vol. 25, no. 2–3. Informa UK Limited, pp. 259–284, Jan. 1998.

[8] H. Christian, M. P. Agus, and D. Suhartono, "Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF)," ComTech: Computer, Mathematics and Engineering Applications, vol. 7, no. 4. Universitas Bina Nusantara, p. 285, Dec.31, 2016.

[9] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural Attentional Rating Regression with Review-level Explanations," Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18. ACM Press, 2018.

[10] R. Parizotto, B. L. Coelho, D. C. Nunes, I. Haque, and A. Schaeffer Filho, "Offloading Machine Learning to Programmable Data Planes: ASystematic Survey," ACM Computing Surveys, vol. 56, no. 1. Association for Computing Machinery (ACM), pp. 1–34, Aug. 26, 2023.

[11] Tyagi, Neelam. "6 Dynamic Challenges in Formulating the Recommendation System." Analytics Steps, www.analyticssteps.com/blogs/6-dynamic-challenges-formulating-imperative-recommendation-system. Accessed 17 Oct. 2023.

[12] Roy, Deepjyoti, and Mala Dutta. "A Systematic Review and Research Perspective on Recommender Systems - Journal of Big Data." SpringerOpen, Springer International Publishing, 3 May 2022, journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00592-5.