IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

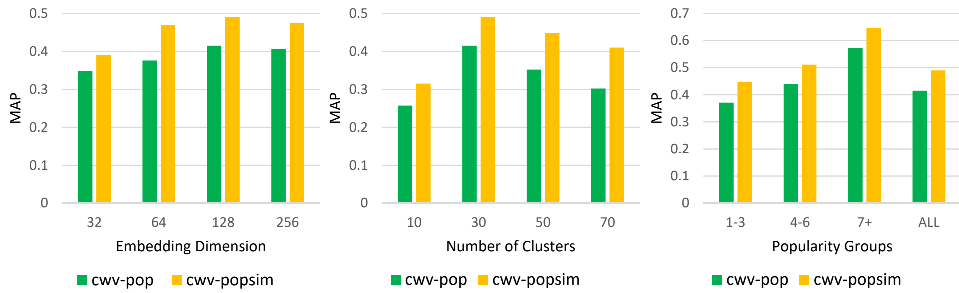

Fig. 4. Experimental results on hyperparameter analysis (the left is the embedding dimension, and the middle is the number of clusters), and the effect of actor popularity groups (right) between cwv-pop and cwv-popsim methods.

The results of changing cluster numbers (10, 30, 50, 70) are exhibited in Figure 4 (middle). We obtain higher MAP scores when the number of clusters is 30. Although the empirical results show that the performance is a bit sensitive to the embedding dimension and the number of clusters, we think the choice of such hyperparameters should rely on the text corpus and how different types of movies and roles are considered. They should be adjusted and tailored to the collected datasets.

We also demonstrate how the popularity of actor/actress influences the recommendation performance. We group the input role descriptions based on the popularity of their corresponding ground-truth actors/actresses, in which the popularity is defined as the number of played movies. Four popularity groups, including 1-3, 4-6, 7+, and ALL, are considered. The results in MAP scores are presented in Figure 4 (right). We can clearly find that role description groups with popular actors/actresses lead to better performance, especially on the "7+" group. We think such results bring an essential implication. Popular actors/actresses are discussed more frequently in social media (PTT), have more news articles, and are described using more sentences. They have much richer training texts. Therefore, we can generate more effective word embeddings to depict these actors/actresses. In contrast, the unpopular actors have fewer text descriptions, and thus their recommendation leads to worse performance.

In short, we arrange insights found in the experimental studies. (1) To have a more accurate recommendation of movie actor/actress, we need to collect more text descriptions for learning better embeddings. (2) The text descriptions should be collected from diverse sources (e.g., news, social media, and Wikipedia). (3) An effective mechanism to select representative query terms from the input role description is also crucial. (4) Both actor popularity and embedding similarity are equally important in recommending actors. (5) It is also essential to filter out irrelevant actors, and our embedding clustering and voting is a potential strategy.

V. Conclusions

This paper presents a novel actor recommendation problem based on unstructured text data for the movie industry. We collect a multi-source dataset, and propose a word embedding-based approach to deal with the task. Promising experimental results encourage future effort on improving actor recommendation. The empirical study also provides a list of insights on description-based actor recommendation. Ongoing work is to exploit the semantic match technique [10] to directly learn the matching function between descriptions of actors and roles.

References

[1] L. Chen, W. Wu, and L. He, *Personality and Recommendation Diversity*, 2016, pp. 201–225.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 4171–4186.

[3] Y. Hu, Z. Wang, W. Wu, J. Guo, and M. Zhang, "Recommendation for movies and stars using yago and imdb," in *2010 12th International Asia-Pacific Web Conference*, 2010, pp. 123–129.

[4] N. Jakob, S. H. Weber, M. C. Müller, and I. Gurevych, "Beyond the stars: Exploiting free-text user reviews to improve the accuracy of movie recommendations," in *Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*, ser. TSA '09, 2009, pp. 57–64.

[5] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002.

[6] P.-H. Li, T.-J. Fu, and W.-Y. Ma, "Remedying bilstm-cnn deficiency in modeling cross-context for ner," in *Proceedings of AAAI International Conference on Artificial Intelligence (AAAI)*, 2020.

[7] A. Liu, Y. Liu, and T. Mazumdar, "Star power in the eye of the beholder: A study of the influence of stars in the movie industry," *Marketing Letters*, vol. 25, no. 4, pp. 385–396, 2014.

[8] R. Meng, S. Zhao, S. Han, D. He, P. Brusilovsky, and Y. Chi, "Deep keyphrase generation," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, pp. 582–592.

[9] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.

[10] J. Rao, L. Liu, Y. Tay, W. Yang, P. Shi, and J. Lin, "Bridging the gap between relevance matching and semantic matching for short text similarity modeling," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 5369–5380.

[11] M. Solaimani, S. Salam, L. Khan, P. T. Brandt, and V. D'Orazio, "Apart: Automatic political actor recommendation in real-time," in *Social, Cultural, and Behavioral Modeling*, 2017, pp. 342–348.

[12] W. T. Wallace, A. Seigerman, and M. B. Holbrook, "The role of actors and actresses in the success of films: How much is a movie star worth?" *Journal of Cultural Economics*, vol. 17, no. 1, pp. 1–27, 1993.

[13] C. Weng, W. Chu, and J. Wu, "Rolenet: Movie analysis from the perspective of social networks," *IEEE Transactions on Multimedia*, vol. 11, no. 2, pp. 256–271, 2009.

[14] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, 2019.