Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

USING A NOVEL TWITTER SIMULATOR FOR FINDING THE EFFICIENT ALGORITHM FOR A FOLLOWEE-FOLLOWER RECOMMENDER SYSTEM

Sepideh Banihashemi

Department of Computer Science
Ryerson University
245 Church Street, ENG-234
Toronto, ON M5B 2K3, CANADA

ABSTRACT

A new simulator of a Twitter recommender system for suggestion of a followee (i.e., a user to follow) is proposed in this paper. This simulator simulates a distributed processing environment such as cloud computing for parallel running of a recommendation algorithm that can be used to test both scalability and accuracy of the employed algorithm. By simulation and comparison of three algorithms (similarity, clustering, and neural network), the similarity and clustering algorithms have higher accuracy compared to the machine learning-based algorithm. To test each algorithm, Twitter user's tweets were changed to TF·IDF and the bag-of-words NLP approaches. In parallel run of employed algorithms by adding additional processing nodes, the scalability of recommender system is measured. This paper shows that the simulator of recommender system developed in this work, is an effective tool for testing scalability and accuracy of the algorithms employed in Twitter recommendation systems when deployed in cloud computing.

1 INTRODUCTION

In online social network services such as Twitter, the users are posting information in short messages which are called tweets. In such a network, a user that follows another user is considered as a follower while the other user would be perceived as a followee (Armentano et al. 2012). Recommender systems are applied to look for the users' preferences and create a list of recommendations for a specific user (Jannach et al. 2010). In terms of Twitter, a followee recommender system can generate a list of users that another user might be interested in following. An initial version of the simulator for the Twitter recommender system is used in this work, with the results of generating the artificial tweets reported in (Li and Abhari 2017).

2 SIMULATION AND RESULTS

The simulations were run on a computer with 4.0 GHz and 16GB RAM. The machine's Operating System was Windows 10. Java was installed on the machine to be able to run the simulator. The performance measurement was obtained from simulations of distributed environment (with JADE) on this machine.

The following data set with its features summarized in Table 1, is used in simulation of three algorithms and the results are shown in Table 2. It is observed that similarity algorithm has the highest accuracy compared to the others in single and multi-node environments. Although the running time of similarity algorithm was much faster than the others, scalability test for all algorithms was performed by increasing the number of nodes while measuring the time and the accuracy maintained the same value. The results in Weka software is for the purpose of the validation of the algorithms employed in recommender system which are inline with the simulator results showing the recommender system (R.S.) accuracy in the distributed processing environment.

Table 1: Dataset used for the simulations.

Dataset	Number of Tweets	Number of Followers	Number of Followees	Number of Unique Words
LDS_Set	94,547	36	4	86,658

Table 2: Accuracy results for K-Means clustering, ANN – multilayer perceptron and similarity method.

Number of Nodes	**Simulator R.S. based K-Means**	**Weka K-Meams**	**Simulator R.S. based MLP**	**Weka MLP**	**Simulator R.S. based Similarity**
1	32.5%	35%	27.5%	25%	32.5%
2	30%	33.29%	24.5%	24.4%	40%
4	25%	49.30%	28.84%	23.90%	40%
8	32.5%	50.96%	NA	NA	35%

NA: The number of folds is greater than the number of instances in each node.

3 CONCLUSIONS

In this paper, for testing accuracy and scalability of different algorithms used in social networks recommender system, a simulator capable of simulating a cloud-based recommender system was developed. The Twitter data simulator is built on top of a distributed processing system for which the benefits were also explained. Finally, for the simulated recommender system, a similarity-based algorithm was found to be the most effective one compared to other two algorithms. Clustering algorithm has the second-best accuracy which is a very common method in collaborative recommender systems. However, clustering is subject to scalability problem, so it is used the same as other algorithms in a distributed environment to simulate Twitter followee recommender system deployed in the cloud environment. The distributed data and multi-node were applied to show the performance of running each algorithm in parallel which is a common method for providing scalability in cloud computing. The results show how much adding the nodes will change the accuracy of running each algorithm in parallel and distributed manner. The simulated multi-node computing and underlying backbone can easily be used for performance measurement of the other machine-learning based applications which are developed to be used by cloud computing, multi-node processors or multi-server systems (such as Spark) in real systems.

REFERENCES

Armentano, M. G., D. Godoy, and A. Amandi. 2012. "Topology-Based Recommendation of Users in Micro-Blogging Communities". *Journal of Computer Science and Technology* 27(3):624-634.

Jannach, D., M. Zanker, A. Felfernig, G. Friedrich. 2010. *Recommender Systems: An Introduction*. 1st ed. New York: Cambridge University Press.

Li J., A. Abhari. 2017. "Generating Stochastic Data to Simulate a Twitter User". In *Proceedings of the 20th Communications Networking Symposium. Society for Computer Simulation International*, April 23rd – 26th, Virginia Beach, Virginia, 102-112.