CSE4077- Recommender Systems

J Component – Review 1 Project Report

RECOMMENDATION BASED ON AMAZON FOOD REVIEW

By

19MIA1065	B. Phanindra Sai
19MIA1071	O. Naga Sai Kumar
19MIA1086	T. Siva Nikhil
19MIA1032	M. Jay Kumar Patel

M.Tech CSE with Specialization Business Analytics

Submitted to

Dr. A. Bhuvaneswari,
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and Engineering

VIT®

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and Engineering

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

FALL SEM 22-23

Worklet details

Programme	M.Tech with Specialization Business Analytics	
Course Name / Code	Recommender system/CSE4077	
Slot	E1+TE1	
Faculty Name	Dr. A. BHUVANESWARI	
Component	J – Component	
J Component Title	Recommendation based on Amazon Food review	
Team Members Name | Reg. No	19MIA1065	B. Phanindra Sai
	19MIA1071	O. Naga Sai Kumar
	19MIA1086	T. Siva Nikhil
	19MIA1032	M. Jay Kumar Patel

Team Members(s) Contributions – Tentatively planned for implementation:

Worklet Tasks	*Contributor's Names*
Data collection & literature survey	Sai Kumar, Jay Kumar Patel, Siva Nikhil, Phanindra Sai
Preprocessing	Jay Kumar Patel, Phanindra Sai
Model building	Phanindra Sai, jay Kumar Patel, Siva Nikhil, Sai Kumar
Visualization	Siva Nikhil, Sai Kumar
Technical Report writing	Sai Kumar, Siva Nikhil
Presentation preparation	Phanindra Sai, Jay Kumar Patel

ABSTRACT

Amazon sells lots of products worldwide and it plays a vital role in our life. now we are analyzing more on their food products. Considering that everyone has different purchase profile, a recommendation system is required to help and give a personalized suggestion products based on the user's preferences. In recent years, consumer interest in shopping online is increased globally with a focus on home delivery. We have data filled with reviews and the ingredients of food. We are trying out content based, popularity based, collaborative based filtering and SVM methods and we are finding out the best and their performances. We will be making the model using the reviews of the people who purchased past and their reviews.

1.Introduction

Almost all e-commerce websites allow users to rate the products or services which they received when shopping. These feedbacks serve as suggestions to other users and are influential to people's decisions on whether to buy the product. Therefore, exploring and understanding the ratings has become an important way to understand the need of users. Moreover, applying the data to build a better recommendation system is an integral part of the success of a company. Recommendation systems of Amazon brings more than 30% of revenues, and Netflix, where 75% of what people watch is from some sort of recommendation. Based on the Amazon Data, we built a recommendation system for Amazon users. We implemented Matrix Factorization, SVD, Deep Learning, content based, popularity based, collaborative based filtering. We compared different methods and made a combination of some methods to provide a better recommendation. The problem we are going to solve is how to help users select products which they may like and to make recommendation

to stimulate sales and increase profits. Firstly, we decided to choose the Amazon Fine Food Reviews dataset which consists of 568,454 food reviews Amazon users left up to October 2012 as our dataset. Secondly, our recommendation system is based on users rating prediction. We assume that users tend to like the products that have a score of greater than 4 and we will consider the highest 5 scores product as our recommendation candidates. Thirdly, we implemented several algorithms to predict the scores of each product for each user. 2.2   Distance Based Model Here we use the cosine-distance to give the similarity between vectors. Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. The similarity ranges from -1 to 1 where -1 means exactly opposite, 1 means exactly the same and in-between values indicating intermediate similarity or dissimilarity.

2. Literature Survey

Sl no	Title	Author / Journal name / Year	Technique	Result
1	Diet-Right: A Smart Food Recommendation System	Faisal Rehman Journal of researchgate 2017	ACO, Cloud Computing	The result shows that the highest accuracy is achieved with 110 ants. It is quite evident that when we increase the number of ants, the accuracy is also increased. Moreover, it is observed that the accuracy remains constant between 80 to 100 ants.
2	Food recommender systems for diabetic patients: a narrative review	Somaye Norouzi, Mohsen Nematy. Journal of researchgate 2017	CFRS, KBRS and CARS	Rule- based reasoning and semantic web such as food ontology and the combination of both were the most popular techniques applied to develop food recommender systems
3	A Personalized Food Recommender System for Zomato	Mansi Goel, Ayush Agarwal. Journal of arvix 2019		Best performance (0.90 F-score) is obtained on manually-annotated ground-truth dataset.
4	Recommendation System for Grocery Store Considering Data Sparsity	NatsukiSanoa, NatsumiMachino Journal of sciencedirect. 2015	SVD-type recommendation based on real POS data	The F-value of the best recommendation method for product category recommendation

				is increased 5.24 times compared to the product item method.
5	Online Grocery Recommendation System	Suja Panicker Journal of researchgate 2016	slope-one and min hash algorithms	Total number of common elements=7 Total number of elements=12 So, 7/12=0.58 That means, similarity between User1 and User2 is 58%.
6	Amazon.com Recommendations Item-to-Item Collaborative Filtering	Greg Linden, Brent Smith, and Jeremy York Journal of UMD 2003	Item-to-Item Collaborative Filtering, search based model	a good recommendation algorithm is scalable over very large customer bases and product catalogs, requires only subsecond processing time to generate online recommendations, is able to react immediately to changes in a user's data, and makes compelling recommendations for all users regardless of the number of purchases and ratings. Unlike other algorithms, item-to-item collaborative filtering is able to meet this challenge.

| 7 | Amazon Food Review Classification Using Deep Learning and Recommender System | Z Zhou, L Xu Journal of stanford Systems, 2009 | Feed-forward Neural Network, LSTM. | Model RMSE Popular(baseline) 1.7372 Collaborative Filtering 1.4538 Matrix Factorization 1.1198 |

3. Dataset and Tool to be used (Details)

https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews

This dataset consists of reviews of fine foods from amazon. The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories.

4. **The Jupyter Notebook** is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

5. **Python** is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems.

6. **Github Repository Link (where your j comp project work can be seen for assessment)**

REFERENCES

[1] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and trends in information
retrieval 2.1-2 (2008): 1-135.

[2] Lam, Savio LY, and Dik Lun Lee. "Feature reduction for neural network based text categorization."
Database Systems for Advanced Applications, 1999. Proceedings., 6th International Conference on. IEEE,
1999.

[3] Sundermeyer, Martin, Ralf Schlter, and Hermann Ney. "LSTM Neural Networks for Language Modeling."
INTERSPEECH. 2012.

[4]Schafer, J. Ben, et al. "Collaborative filtering recommender systems." The adaptive web. Springer Berlin
Heidelberg, 2007. 291-324.

[5]Lops, Pasquale, Marco De Gemmis, and Giovanni Semeraro. "Content-based recommender systems: State
of the art and trends." Recommender systems handbook. Springer US, 2011. 73-105.

[6]Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 8 (2009): 30-37.

[7]Van den Oord, Aaron, Sander Dieleman, and Benjamin Schrauwen. "Deep content-based music recommendation." Advances in Neural Information Processing Systems.
2013.

[8] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global Vectors for Word Representation. In EMNLP (Vol. 14, pp. 1532-1543).

[9] Christopher Olah, Understanding LSTM Networks, retrieved from
http://colah.github.io/posts/2015-08-
Understanding-LSTMs/

[10]Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2015). Session-based Recommendations with

Recurrent Neural Networks. arXiv preprint arXiv:1511.06939.

[11]Chellapilla, K., Puri, S., & Simard, P. (2006, October). High performance convolutional neural networks

for document processing. In Tenth International Workshop on Frontiers in Handwriting Recognition. Suvisoft.

[12]Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing

internal covariate shift. arXiv preprint arXiv:1502.03167. Chicago

[13]Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. The Journal of

Machine Learning Research, 13(1), 281-305. Chicago

[14]Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple

way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.

6

[15]Derks, E. P. P. A., Pastor, M. S., & Buydens, L. M. C. (1995). Robustness analysis of radial base function

and multi-layered feed-forward neural network models. Chemometrics and Intelligent Laboratory Systems,

28(1), 49-60.