

Predicting online Grocery Ordering Intention

Rohit Rathish
Department of Computer
Engineering,
V.E.S. Institute of
Technology,
University of Mumbai,
Mumbai-400074, India.

Yash Jahagirdar
Department of Computer
Engineering,
V.E.S. Institute of
Technology,
University of Mumbai,
Mumbai-400074, India.

Yogesh Gupta
Department of Computer
Engineering,
V.E.S. Institute of
Technology,
University of Mumbai,
Mumbai-400074, India.

Vaibhav Jain
Department of Computer
Engineering,
V.E.S. Institute of
Technology,
University of Mumbai,
Mumbai-400074, India.

Abstract— E-commerce has seen a tremendous growth in the past decade. Online grocery shopping is one such application of E-commerce. An important feature of an online grocery system is to come up with suitable recommendations, which can help the user make quick decisions, so that they don't have to spend additional time, browsing the website. Generating visual aids and reports acts as another plus point, for these websites. In order to implement this, almost all major online shopping sites use Recommendation systems. The aim of the paper is to give an overview of a smart cart application which is a predictive model application and which aims to provide germane item recommendations based on purchase history and user interests based on a dataset. The model which uses a recurrent neural network model for predicting future orders of customers in grocery shopping has been described in detail. The paper also describes the application that analyses the customer's grocery purchasing habits and predicts which previously purchased products will be in a customer's next order. The application will also have a recommendation system that predicts the items hence making the user's buying experience more robust.

Keywords—E-Commerce, Grocery, Recommendation System, Neural Networks

I. INTRODUCTION

There has been an increasing demand for e-commerce sites, in the past decade. Online grocers, especially have increased in popularity. Each of these sites are using Recommendation systems and algorithms. The aim of this paper is to create an online grocery recommendation system that can come up quickly with relevant recommendations for the user and display interesting patterns for companies which can be used for market research.

II. LITERATURE SURVEY

Various research papers, articles and patents have been studied before embarking on methodology. In one technical paper the slope one and min hash algorithms have been proposed for item-item collaborative filtering in the grocery recommendation system [1]. In another paper, the prediction of a recommendation system using association rule mining and analyzing past orders of customer and mapping the selected item with similar item is achieved [2]. Likewise in random walk algorithm [3] the rank of a page or an item is decided by the probability that the item will be decided. Also, in the patents [7] and [8] the recommendation systems are produced by taking into account user's interests and past history of item viewing while recommending items. In this systems, the item to item mapping occurs and by using set of tables the item to

item mapping is achieved. Papers [4] and [5] give an overview of the f optimization algorithm and recurrent neural networks (RNN). The [6] paper gives information regarding different algorithms used for pattern mining and compares them across a set of parameters. Thus, in the literature survey various algorithms and methods related to recommendation systems were studied and based on that methodology is prepared.

III. METHODOLOGY

The dataset given consists of aisles, departments, orders and products dataset. It also consists of 3 million orders purchased by myriad of anonymized users.

Before exploration of data, the data is cleaned to make exploration of data as well as training the data for neural networks possible. Data cleaning is the process of detecting and correcting inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty data. This is done using dplyr and tidyr packages of R language.

The cleaned data thus obtained is explored using exploratory data analysis. In exploratory data analysis, the correlations between different factors are studied and the various x-y plots such as scatter plots, histograms are drawn and necessary inferences are identified. This exploratory data analysis is used to visualize data and find different data insights which helps in the next steps.

This data exploration will be displayed in the form of graphs, reports.

After data exploration, RNN is to be used to form a recommendation system application which predicts the grocery order of a customer based on his/her history and interests identified using dataset given.

The RNN and its layers are trained using a training dataset consisting of orders ordered by customers.

Using f1 measure and optimizing it, the RNN can be trained accordingly.

For implementing RNN, Python language with NumPy package is used. Also, TensorFlow, a package is used to train and test the neural network.

For testing the neural network, a test, dummy dataset is created and is used to test the accuracy of the neural network, thus providing the accuracy of the recommendation system.

If the accuracy of the system is not up to 80% then the RNN is to be retrained and retested.

IV. SYSTEM OVERVIEW

ER diagram of the system is as follows:

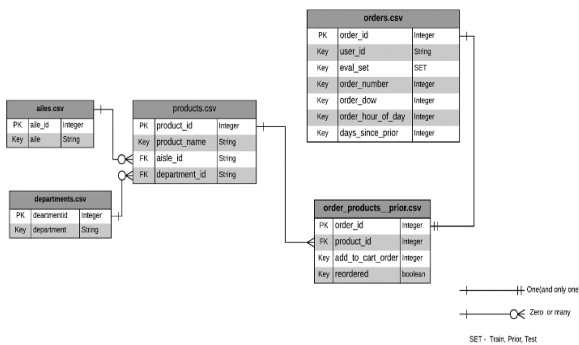


Fig1. Er Diagram

The model aims to predict the future orders of the customers, so there is an abundant need of data for the developers/admins in order to develop a proper working system. Here, the admin will get the much needed data from the various data stores maintained at the admin side. This data comes in the form of various customer details, order details and department details which are updated as soon as customer places his current order. This data is fed to be cleaned and is segregated into training data and test data. The training data is used to train the algorithm (RNN) and the test data is used to test and check whether the algorithm is working efficiently or not. The processing of the data using RNN and then optimising of the data using F1 forms the basis for the development of the recommendations. Once the system is working correctly, recommendations are then deduced and sent to the customer for his/her ease.

Architectural Diagram of the system is as follows:

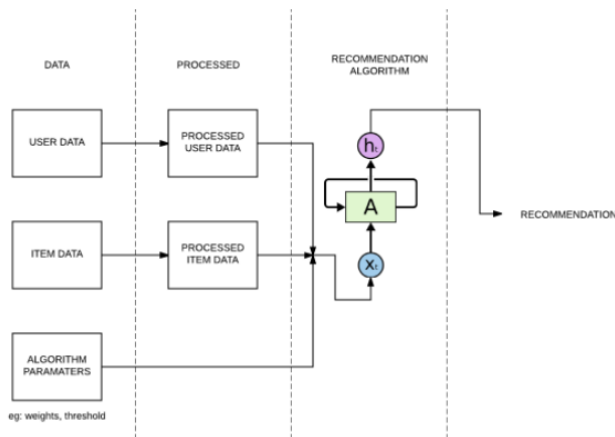


Fig2. System Architecture

As recurrent neural network has been chosen to perform the analysis and prediction, it does expects **data**. This data has been provided by the customers ordering the items which are getting stored in the databases as well as the data already available which is the items data. As this data is increased, it is processed further to segregate the data into training data which is used to train the algorithm and the test data which tests the data after the algorithm has been devised. As we are using neural networks, so algorithmic parameters like weights, thresholds are also considered. Training user data, training item data, algorithmic parameters are fed to the training algorithm as input which then devices the input and puts forth an output. The test data is then used to device the accuracy of the algorithm. Algorithm is revised if it is unable to clear the test.

V. CONCLUSION

The recommendation system which has been proposed, is based upon neural networks and is able to provide suitable predictions about the future orders of the customer.

VI. ACKNOWLEDGMENT

The authors wish to acknowledge Prof. Mrs. Abha Tewari for her guidance and encouragement.

VII. REFERENCES

- [1] Prof. Suja Panicker, Shubham Fulzele, Prayag Mane, Akshay Rathi and Chetan Pakhale. "Online Grocery Recommendation System." International Journal for Scientific Research and Development 4.4 (2016): 318-321.
- [2] Rudin, C., Letham, B., Salleb-Aouissi, A., Kogan, E., & Madigan, D. (2011, December). Sequential event prediction with association rules. In *Proceedings of the 24th annual conference on learning theory* (pp. 615-634).
- [3] Li, M., Dias, B. M., Jarman, I., El-Deredy, W., & Lisboa, P. J. (2009, June). Grocery shopping recommendations based on basket-sensitive random walk. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1215-1224). ACM.
- [4] Nan, Y., Chai, K. M., Lee, W. S., & Chieu, H. L. (2012). Optimizing F-measure: A tale of two approaches. *arXiv preprint arXiv:1206.4625M*. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
- [5] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modelling.
- [6] Shamila Nasreen, Muhammad Awais Azam, Khurram Shehzad, Usman Naeem, Mustansar Ali Ghazanfar(2014).Frequent Pattern Mining Algorithms for Finding Associated Frequent Patterns for Data Streams: A Survey. *doi:10.1016/j.procs.2014.08*
- [7] Jacobi, J. A., Benson, E. A., & Linden, G. D. (2011). *U.S. Patent No. 7,908,183*. Washington, DC: U.S. Patent and Trademark Office.
- [8] Linden, G. D., Smith, B. R., & Zada, N. K. (2005). *U.S. Patent No. 6,912,505*. Washington, DC: U.S. Patent and Trademark Office.