International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-8, June 2020

Product Quantization using Regression

Vasanthi Vadlamudi, Divya Sree Ravi, Rishita Dhulipalla, Rohith Vamsi Danduboyina, K.S. Vijaya Lakshmi

Abstract: ***Approximate Nearest Neighbor (ANN) has developed an immense demand for many tasks. This ANN methodology was being used for product quantization. These product quantization methods were being used for e-commerce sites. However, this quantization maybe sometimes misleading due to a lack of accuracy in technique. So, we managed to increase the accuracy of quantization by adding Logistic Regression in the process. This helps to increase the accuracy of the method by having a probability value. This helps to make correlated items much more accurate when compared to pure quantization. This method is helpful for e-commerce sites for efficiency in the prediction of purchase by the customer.***

Keywords: ***Approximate nearest neighbor, product quantization, quantization, regression.***

I. INTRODUCTION

Nowadays E-Commerce sites have achieved great success in winning customers attention.E-commerce made customers life more convenient because they can just click and browse the products they need. It could help customers by saving time. Consumers save their time and energy through online purchases it has gained an immense power in daily life. As the customers for online purchasing were increased the website owners are planning to make more profits out of it by recommending the products related to customer interests. They are collecting the customers shopping data and knowing the correlated items of their purchase. This online product quantization has developed an immense craze in E-Commerce sites for predicting the customer purchase. These predictions will help the developers to recommend only the products related to customer's likes and avoid the products that are not related to customers. Details like customer age, purchase history, and price range are considered to generate these items that can be recommended. These recommendations help to increase the profits to the site since most of the customers purchase the items that are recommended to them by the website. The past methods included online sketching and hashing but this quantization made a huge difference in the approach.

II. RELATED WORK

In Online Hashing [1], it was proposed that online hashing updates the model frequently. The hash function is expressed as a structured prediction and then propose a prediction loss function. There is an updated hash model in every step impelled by optimal hash codes, which helps to lead to zero prediction loss.

[2]Structure Sensitive Hashing apprehends the two types of structures in data. SSH uses discriminative functions which quantize data into the prototypes of clusters combined with unique binary codes.

In PQ for ANN[3], there is an approximation of the Euclidean distance. To avoid exhaustive search an inverted file system is combined which resulted in high efficiency. It improved memory usage and search efficiency.

In online product quantization by D. Xu team [4], they have presented online PQ method for real-time data. Besides, they added partial codebook update so that there is an update in time cost. Also, to loss bound there is online PQ over sliding window approach for the real-time data.

For nearest neighbor search in a dynamic database, hashing methods [8] were being used. They have techniques for the new data being updated, without training stored data points again. Online Hashing [1], Adapt Hash [7] and Online Supervised Hashing [9], require label information, which might not be possible in most of the real-world applications. Stream Spectral Binary Coding (SSBC) [5] and Online Sketching Hashing (OSH) [6] do not require labels, which are the current online unsupervised hashing methods.

III. METHODOLOGY

In this methodology first, the Architecture of our proposed system is explained and then about the Data Collection, pre-processing, segmentation, and the procedure was explained briefly. All the steps are represented step by step from starting to ending of the process. In module A we can see the fig.1 where the method starts with giving data as input followed by data preprocessing later by attribute selection and then by ANN approach along with logistic regression which results in the prediction of purchase. The probability values helps in having the accuracy in the prediction. In fig.4 we can have a clear difference of prediction by using regression. Below there are architecture diagram, dataset details, procedure and significance of method.

Revised Manuscript Received on June 30, 2020.
***** Correspondence Author
Vasanthi Vadlamudi*, Department of Computer Science and Engineering, VR Siddhartha Engineering College, Vijayawada, Kanuru. Email: vasanthivadlamudi@gmail.com
Divya Sree Ravi, Department of Computer Science and Engineering, VR Siddhartha Engineering College, Vijayawada, Kanuru. Email: divyasreeravi1@gmail.com
Rishita Dhulipalla, Department of Computer Science and Engineering, **VR** Siddhartha Engineering College, Vijayawada, Kanuru. Email: rishitharishi41099@gmail.com
Rohith Vamsi Danduboyina, Department of Computer Science and Engineering, VR Siddhartha Engineering College, Vijayawada, Kanuru. Email: rohithvamsi0101@gmail.com
K.S.Vijaya Lakshmi, Assistant professor at the Department of Computer Science and Engineering, VR Siddhartha Engineering College, Vijayawada, Kanuru. Email: vijaya@vrsiddhartha.ac.in

3.1 The following is the Architecture we used for our methodology

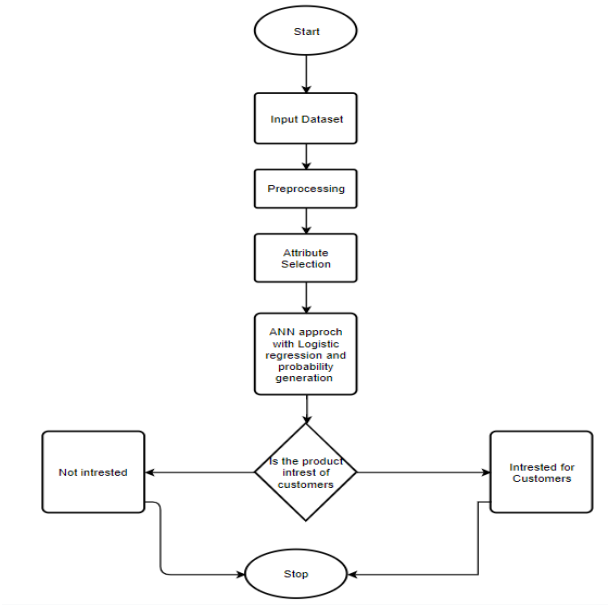

Fig. 1. Architecture for our proposed system.

A. Dataset Collection

This dataset fig .2 has been collected from Kaggle having customer shopping details which is the subset that contains 10 attributes with 152 tuples. It has the product id, customer reviews etc.

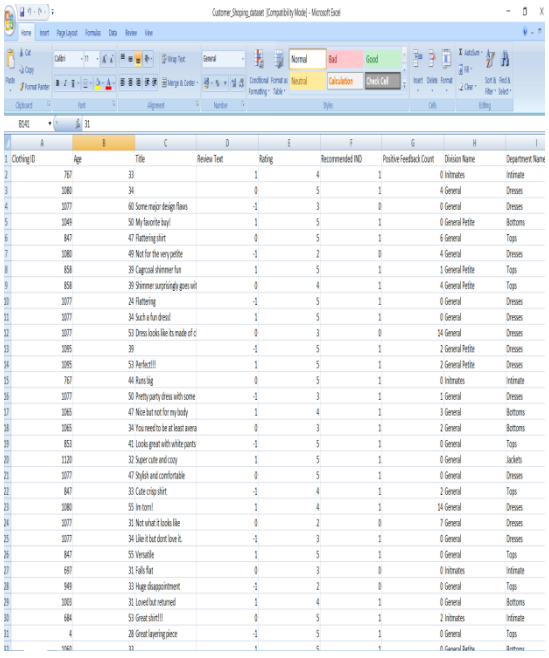

Fig. 2. Dataset consisting of different attributes

B. Procedure for Product Quantization

1. The collected dataset is preprocessed and unwanted data is removed completely by manual data preprocessing (since we have taken static dataset) for better analysis results. Real-time data preprocessing technique works best for dynamic data.
2. The attributes that are required for the process are chosen such as clothing id, rating, review text, etc.
3. The complete data is reviewed and the ANN approach is used to generate the correlated items.
4. This is the general approach for the quantization of products but since we are trying to increase the accuracy of the system we include the regression concepts.
5. Calculating the approximate probability of occurrence of items will help to make the prediction much more efficient since the introduction of logistic regression on entire data

Clothing_ID	Score	Class Name	Score	
949.0	Sweaters	2.0	10	1
89.0	Sleep	4.0	1	0
1003.0	Skirts	4.0	2	0
1049.0	Pants	5.0	7	1
1120.0	Outerwear	5.0	2	0
767.0	null	1.0	2	0
697.0	Lounge	3.0	4	0
858.0	Knits	5.0	40	1
966.0	Jackets	4.0	3	0
767.0	Intimates	5.0	3	0
910.0	Fine gauge	5.0	2	0
1077.0	Dresses	3.0	32	1
847.0	Blouses	5.0	34	1

Fig. 3 Results of correlated items without regression

Clothing_ID	Score	Class Name	Score	
949.0	Sweaters	2.0	10	1
89.0	Sleep	4.0	1	0
1003.0	Skirts	4.0	2	0
1049.0	Pants	5.0	7	1
1120.0	Outerwear	5.0	2	0
767.0	null	1.0	2	0
697.0	Lounge	3.0	4	1
858.0	Knits	5.0	40	1
966.0	Jackets	4.0	3	1
767.0	Intimates	5.0	3	1
910.0	Fine gauge	5.0	2	0
1077.0	Dresses	3.0	32	1
847.0	Blouses	5.0	34	1

. Fig. 4 Results of correlated items after applying regression

C. Significance of Method

This model includes different techniques like customer segmentation, purchase prediction. Logistic Regression is used to predict that given an appropriate customer cluster will be interested to purchase the product or not. A binary predictor is the best choice for identifying the final product behavior of the data because we can have only categorical values.

The earlier methods of quantization will help to know the recommendations that can be given to the customers while this technique will help to make the sentiment analysis better the positive reviews which are considered neutral are managed and the prediction of purchase will be efficient. Let us take a look at the results, in fig.3 the clothing id 966.0 is not recommended but in fig.4 966.0 jackets is recommended because the text sentiment score, probability made it recommendable when regression is applied.

IV. RESULTS AND DISCUSSIONS

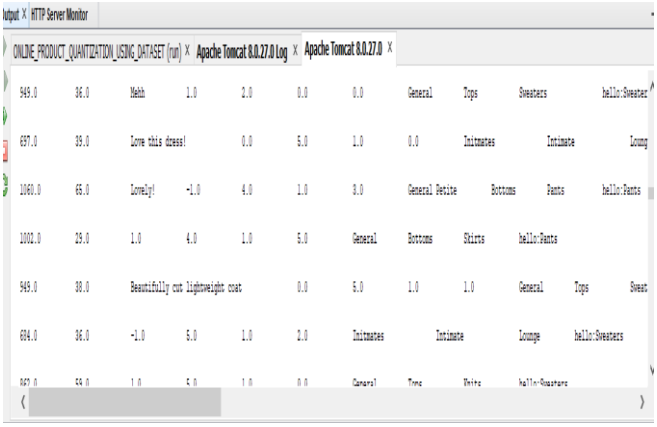

Fig. 5 The data is analyzed in the Net Beans

Fig. 6 GUI for input and viewing results

The above fig.5 and fig.6 shows the results and the GUI of the page layout.

V. CONCLUSION

The data which is collected for over a while is considered to analyze the customer interests and predict his next purchases through the process of product quantization which is already an efficient way of prediction has been improved slightly by adding logistic regression in the methodology which is a binary predictor which gives a straight away categorical values. This will help the owner to suggest only the products which are of customers interests so there will be less wastage of advertisement costs and even the customer will not be irritated with unnecessary stuff going on. This approach is an addition to the existing quantization. We can add features like generating reports based on customer ratings and reviews regarding a product and verify whether the product is worth placing on the website or not. So, that the trust of customers can be gained easily regarding the product quality.

REFERENCES

1. L. Huang, Q. Yang and W. Zheng, "Online Hashing," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 6, pp. 2309-2322, June 2018.
2. X. Liu, B. Du, C. Deng, M. Liu and B. Lang, "Structure Sensitive Hashing With Adaptive Product Quantization," in IEEE Transactions on Cybernetics, vol. 46, no. 10, pp. 2252-2264, Oct. 2016.
3. H. Jégou, M. Douze and C. Schmid, "Product Quantization for Nearest Neighbor Search," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 1, pp. 117-128, Jan. 2011.
4. D. Xu, I. W. Tsang and Y. Zhang, "Online Product Quantization," in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 11, pp. 2185-2198, 1 Nov. 2018.
5. M. Ghashami, A. Abdullah, "Binary coding in-stream", *CoRR*, vol. abs/1503.06271, 2015.
6. C. Leng, J. Wu, J. Cheng, X. Bai, H. Lu, "Online sketching hashing", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2503-2511, 2015.
7. F. Cakir, S. Sclaroff, "Adaptive hashing for fast similarity search", *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1044-1052, 2015.
8. Q. Yang, L. Huang, W. Zheng, Y. Ling, "Smart hashing update for fast response", *Proc. Int. Joint Conf. Artif. Intell.*, pp. 1855-1861, 2013.
9. F. Cakir, S. A. Bargal, S. Sclaroff, "Online supervised hashing", *Comput. Vis. Image Understanding*, vol. 156, pp. 162-173, 2017.

AUTHORS PROFILE

Vasanthi Vadlamudi pursuing 4/4 B.Tech in department of CSE at Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, AP. Her area of interest include data science, Internet of things and computer science.

Divya Sree Ravi pursuing 4/4 B.Tech Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, AP. Her area of interest include data science, Internet of things and computer science.

Rishita Dhulipalla pursuing 4/4 B.Tech Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, AP. Her area of interest include data science, web technologies and computer science.

Rohith Vamsi Danduboyina pursuing 4/4 B.Tech Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, AP. Area of interest include data science, Internet of things Animation Design and Visual Effects.

K. Sree Vijaya Lakshmi, M.Tech,(Ph.D), LMISTE, working as an Assistant Professor in the Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada since March 2010 and worked as a Programmer for 15 years in same Department and College from 1995 to 2010. Pursuing Ph.D. from JNTUK, Kakinada. Awarded a Certificate of Merit from DOEACC Society (Department of Electronics Government of India and Computer Society of India) on 31st July 2002. 7 papers were published in International Journals, 4 papers were presented in International Conferences and one paper was presented in the National Conference. All these presented papers were published as book chapters in proceedings of those Conferences.