RECOMMENDATION SYSTEMS IN E-COMMERCE WEBSITES

By Radhika Bailurkar (2019430003)

Faculty Incharge: Prof. Nataasha Raul

COMPARATIVE ANALYSIS

Sr No	Year	Method	Performance Metric	Features	Gaps Identified
1	2016	KMeans , Apriori, kNN	Estimation Values	Business Type Code,Business Name, Estimation Value	Sparsity will result if one customer gives ratings to business that are not frequented by other customers
2	2018	BAT Algorithm(BA, ABC Algos), five fold cross validation	MAE,Precision, Recall, F1 Score, RMSE, 6.9%BA>ABC	Swarm Size, Iteration and Run	In this work, the authors present a different scheme to conquer collaboration problems by giving weights to items whilst evaluating the similarity between users. The weights are assigned and iteratively improved using the algorithms mentioned.
3	2016	NewRec, TimeRec, HF,CF	MAE and RMSE	Data Sparsity, Influence of Time, Influence of no. of Users, Comparison among algorithms	Influences of context and user interaction behaviour
4	2013	Resnick model, Average model,SR,TR ,ST,STR Model	Precision,MAE,RMSE	Preference similarity, recommendation trust, and social relation	Experiment was performed on a small scale. Although the social networking service is increasingly promising, the offered features of social commerce are still limited to prestigious e-commerce sites.
5	2007	MinHash Clustering, PLSI, covisitation counts	Precision-Recall Curve	User click history, Clicked story on News Feed	Increased Cost, Sometimes while analyzing live traffic results may differ acc to Algorithms
6	2018	MLP,CNN	MAE,RMSE	Rating Matrix and Item Review	

7	2018	PCA, Karhunen-loeve, Discrete Fourier transform	Precision, Recall, Speedup/Time, F1 metrics, MAE,RMSE	Attributes of user, or of the items or both.	Relevant Variable Selection, Periodicity and Seasonality
8	2018	K-means Algorithm, Apriori Algorithm	Quality Metrics – confidence distance CD and recommendation conformity RC	Order ID, Product ID, Product Name	Results were different from expected because of heterogeneous dataset.
9	2004	DC Tree Algorithm, HITS	Recommendation Accuracy, Shortcut Gain	No of Clicks on the website	Incomplete or Limited Information Problem,Incorrect Information Problem, Persistence Problem
10	2007	CBA-CB	Associative Rules, Decision Tree Metrics	Requirement Phrases such as: Stylish, Colorful screen, light,etc	Data fragmentation. Doesnt work well with continuous data.

DATASET USED

1. 8.Korean Business Dataset
2. Jester Dataset comprises of 4.1 million ratings given by 73k users for 100 jokes
3. GroupLens research group in University of Minnesota. There are over 70,000 users and 6,600 rated movies in the database of MovieLens site.
4. Yahoo! Shopping, online shopping website in Taiwan. Data was collected in the form of Questionnaire.
5. a. Movie Lens Dataset consists of movie rating data collected using a web-based research recommender system.

   b. The second dataset consists of a subset of clicks received on the Google News website over a certain time period.

   c. News Big, contains more records: 500, 000 users, 190, 000 unique items and about 10, 000, 000 clicks.

6. a. MovieLens-1M

b. Amazon Android Apps

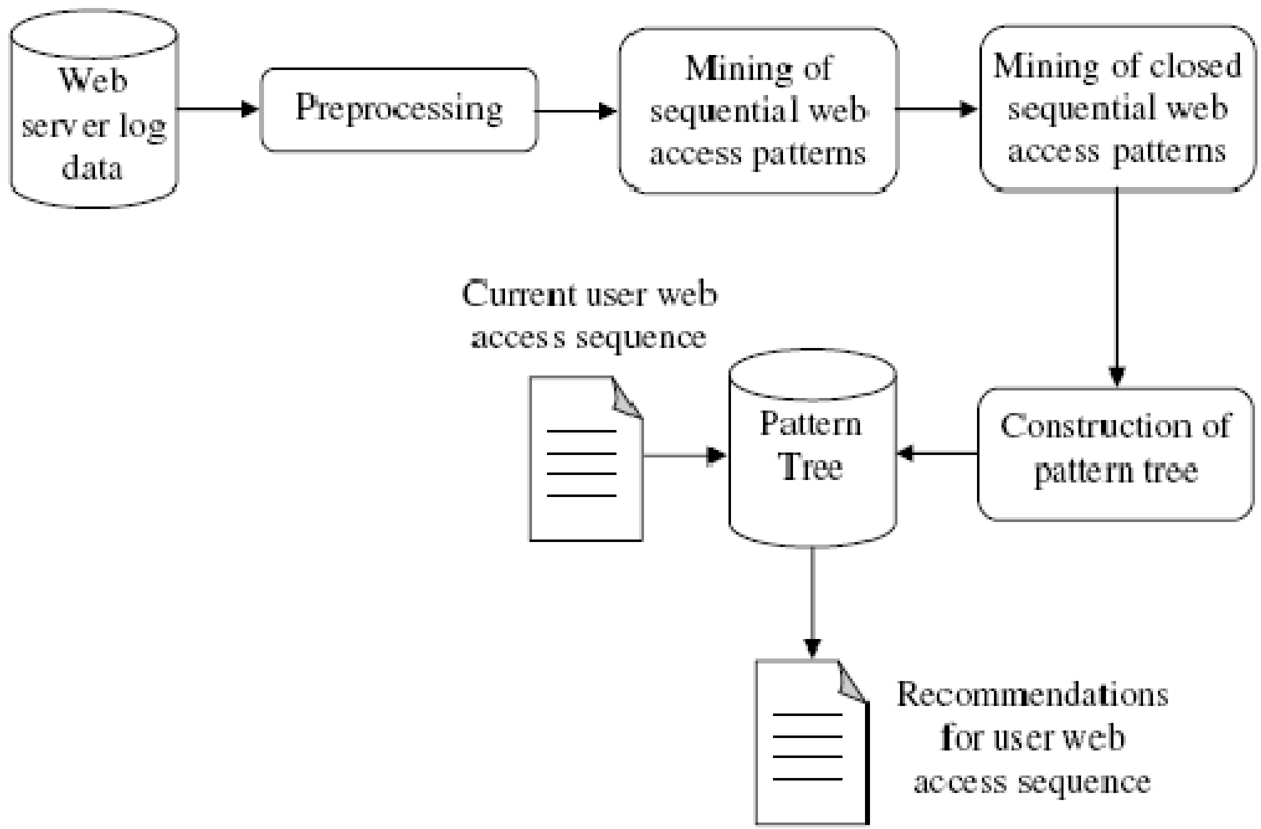
c. Amazon Instant Video

d. Amazon Digital Music

7. MovieLens Dataset
8. Kaggle Competition
9. Results from web site of the Computing Science Department of the University of Alberta, Canada. Data was collected for 8 months.
10. Web Database

WORKFLOW MODEL

(Proposed)



IMPLEMENTATION

We are given a data set of items, with certain features, and values for these features The task is to categorize those items into groups. To achieve this, we will use the kmeans algorithm; an unsupervised learning algorithm.K-Means clustering intends to partition *n* objects into *k* clusters in which each object belongs to the cluster with the nearest mean. This method produces exactly *k* different clusters of greatest possible distinction. The best number of clusters *k* leading to the greatest separation (distance) is not known as a priori and must be computed from the data. The objective of K-Means clustering is to minimize total intra-cluster variance, or, the squared error function.

1. Specify number of clusters K.
2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

- Compute the sum of the squared distance between data points and all centroids.
- Assign each data point to the closest cluster (centroid).
- Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

Kmeans Squared Error Function

number of clusters number of cases

centroid for cluster j

case i

objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \underbrace{\left\| x_i^{(j)} - c_j \right\|^2}_{\text{Distance function}}$$

REFERENCES

1. JinHyun Joo, SangWon Bang, GeunDuk Park, "Implementation of a Recommendation System using Association Rules and Collaborative Filtering", *Information Technology and Quantitative Management (ITQM 2016)*, 2016

2. Sambhav Yadav, Sushama Nagpal, “An Improved Collaborative Filtering Based Recommender System using Bat Algorithm”, *International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*,2018
3. Zhijun Zhang, Gongwen Xu, and Pengfei Zhang, ”Research on E-Commerce Platform-Based Personalized Recommendation Algorithm”, *Applied Computational Intelligence and So Computing*,2016
4. Yung-Ming Li, Chun-Te Wu, Cheng-Yang Lai, ”A social recommender mechanism for e-commerce: Combining similarity, trust and relationship”, *Decision Support Systems*,2013
5. Abhinandan Das, Mayur Datar, Ashutosh Garg, ”Google News Personalization: Scalable Online Collaborative Filtering”, *Track: Industrial Practice and Experience*,2007
6. Meshal Alfarhood, Jianlin Cheng, ”DeepHCF: A Deep Learning Based Hybrid Collaborative Filtering Approach for Recommendation Systems”, *17th IEEE International Conference on Machine Learning and Applications*, 2018
7. Janani Balasubramanian, Soumya Koppaka, Chinmay Rane, and Nataasha Raul, ”Parameter Based Survey of Recommendation Systems”, 2018
8. Kutuzova Tatiana, Melnik Mikhail, “Market basket analysis of heterogeneous data sources for recommendation system improvement”, *7th International Young Scientist Conference on Computational Sciences*, 2018
9. Jia Li and Osmar R Zaiane, ”Combining Usage, Content, and Structure Data to Improve Web Site Recommendation”, 2004
10. Xi-Zheng Zhang, "Building Personalized Recommendation System in E-Commerce using Association Rule-Based Mining and Classification," *2007 International Conference on Machine Learning and Cybernetics, Hong Kong*, 2007, pp. 4113-4118.
11. Gediminas Adomavicius, and Alexander Tuzhili, ”Towards the Next Generation of Recommender Systems:A Survey of the State-of-the-Art and Possible Extensions”
12. Neha Verma, Devanand, Bhavna Arora, ”Experimental Analysis of Recommendation System in e-Commerce”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE),ISSN: 2278-3075, Volume-8 Issue-8S3*, June 2019
13. J. Ben Schafer, Joseph A. Konstan, John Riedl,”E-Commerce Recommendation Applications”,*E-Commerce Recommender Applications*
14. Francisco, G. S., Rafael,”An integrated approach for developing e-commerce applications”, *Expert Systems with Applications*, 2005
15. Xiaojun, Liu, “An improved clustering based collaborative filtering recommendation algorithm”,*Cluster Computing*,2017