International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Oniomania Similarity Search with Recommended System

K.Karpaga Priyaa, C.Saranya, PA.Yogaalakshmi, M.Haripriya

Abstract— ***Users shopping online prefer to quickly access products that they are in need of among a list of similar products. The similarity metrics in use, only contemplates on the product attributes which doesn't always meet the user expectations. The current system relies on solving this by performing Θ-similarity and r-nearest neighbor, where product similarity is considered only if it satisfies a user-preference list, by hitting the reverse top-k queries results. However, the products expressed here are generally based on the user's preference, which can break the user independence of the product. To conquer these drawbacks, we use a more advanced system which represents feature based products such as Usercentric Model, where we have two phases. At first, we gather opinion based data about the feature, and subsequently these feature similarities are ranked. The ranking is predicted by collaborative filtering recommendation algorithm. Next, the rank generated from the above algorithm will be compared with existing data to get the top-k best product by performing No-Random Access algorithm (NRA) in second phase. Through this users are granted the top.***

Index Terms: Opinion mining, Collaborative filtering algorithm, NRA algorithm.

I. INTRODUCTION

Similarity computation among products is a very common task in data management. The estimation of similarity metrics is performed on several instances such as comments on youtube videos, number of likes-dislikes on a post, the reviews for a product etc, to understand and analyse the similarity in the preferences and tastes among social media users. In the existing system, two products being compared are considered to be similar if both match the preferences of the user, i.e it is a product in the results of the reverse top-k set. The user is provided with a fixed set-of queries about the characteristics of the products which needs to be answered.
These responses by the users' are expressed as a vector of weight values for the characteristics of the product. The reverse top-k query is run on these vectors to obtain the similarity metrics which represents a set of vendee who would like the product belonging to the top-k sets [1]. This existing system neglects the consideration of user opinion in the process for similarity computation. Hence, in out proposed system we consider this aspect by obtaining the user opinions in the form of reviews about the product only after the user having used that product. The opinions/reviews are the content obtained from the users about the product, its services and policies. Users always consider the positive and the negative opinions about the product in terms of the reviews and likes-dislikes on the product before making a purchase [2]. Sentiment analysis replacing the traditional web based surveys methods used by organizations to obtain information from the people about products and services.

A direct correlation of one product with another based on the opinion of the user is now trending. We qualitatively compare the two products with their existing features to rank them. The users prefer the reviews of the products to be transparent and explicitly provided, however, they are forced to scroll through a number of irrelevant reviews while looking for information on a specific feature of interest. Thus, we propose this system to minimize the work the user needs to perform by showing how we work on opinion mining to find the best product among the sets of products. In first phase, we follows by extracting the Public opinion about the features from the reviews and establishing feature similarity through K-means clustering algorithm and ranking through collaborative filtering algorithm. . In Second phases, we follows by performing the No-Random Access algorithm on the resulting sets to find best product

II. EXISTING SYSTEM

Different similarity metrics have been proposed in market analysis, to evaluate the similarity between the products. So far the similarity metric used is cosine similarity to estimate the similarity between the products. But the similarity among the products is estimated for user compatibility which has to take user preferences as a value. In this work they introduced a new user-centric approach for similarity computation, which capitalizes on rankings of products based on user preferences. The user preference is obtained by providing set of predefined question about the product to the user and users have to answer the sequestions.

By this way users' preferences are obtained and they are expressed as vectors. Two products tend to be similar if they meet the same user preferences, which in turn results in comparable inverse top-k sets. The reverse top-k queries are used for similarity judgements which results a set of customers who likes a given product that belongs to top-k sets.

Manuscript published on 30 September 2019
* Correspondence Author
K.Karpaga Priyaa*, Assistant professor, CSE department, Sri Sairam Engineering College, Chennai, Tamilnadu, India.
C.Saranya, Assistant professor, CSE department, Sri Sairam Engineering College, Chennai, Tamilnadu, India.
PA.Yogaalakshmi, Student, Engineering Degree in CSE, Sri Sairam Engineering College, Chennai, Tamilnadu, India.
M.Haripriya, Student, Engineering Degree in CSE, Sri Sairam Engineering College, Chennai, Tamilnadu, India.

The query processing is carried out in 2novel approach Θ-similarity and m-nearest neighbor algorithm to find the upper and lower bounds which in turn utilized for user centric similarity calculations. In this work, they use Jaccard coefficient to perform similarity estimations linking the final sets of the reverse top-k queries.

By dividing them in upper and lower this system provides significant result and provides efficient similarity amid the products. But they expressed in most general way of user preferences. They haven't considered the user reviews where the people reveal about the product qualities and services

III. PROPOSED SYSTEM

Estimating user preferences about the stock is significant in e- commerce application but it's important for a user to have the self-report about the product. Thus the proposed system integrates user opinion by claiming k-means clustering algorithm. This system works by 2 phases, (refer fig1) at first feature similarity is established by implementing the opinion mining and ranking is done by k-means clustering algorithm later rank evolved from the above is collated with existing work to acquire the top k best product by operating them with NRA threshold algorithm in second phase.

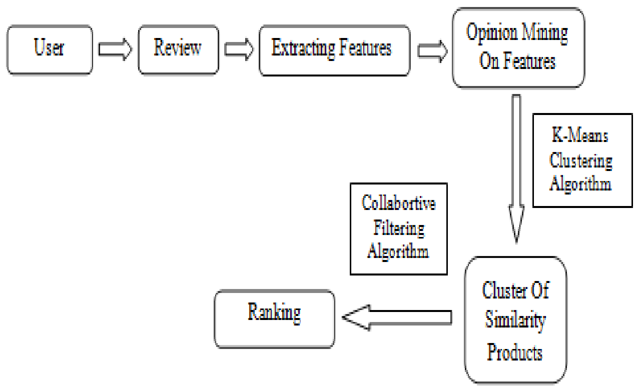


Fig 1: Proposed system

Extracting the features from the reviews: (Phase 1.1) Level 1:

The syntactical form of reviews are analyzed in the Pre-processing phase [2]. The pre- processing involves,

1. pos (post of speech) tagging,
2. Chunking,
3. Stop-word removal,
4. Stemming.

The explicit features are identified by parsing the system recursively which identifies the parts. The necessary features can be identified using Sentiment analysis .The proper feature selection is then prevailed by using the NLP (Natural Language Processing) which usually express product features as product P and feature F in phrases like or 'P has F' or 'F of P'. In Fig 2, While extracting we may extract unnecessary features which must be [2] cleansed by pruning the whole set and combining the similar features.

Public opinion about the features:(Phase 1.2) Level 2:

The polarity between each opinion and its feature is found by dicriminating the opinions and non-opinions in a single review. The word semantic orientation is performed

Task---Given a set of positive, negative, neutral semantic orientation (SO) labels, a set of reviews and tuples, each tuple with opinion term w connected with feature f for a given statement s, each tuple (w, f, s) is assigned an SO label.

Solution --- *We use 3-step approach for the solution*

1. In the collection of reviews provided, we find the SO label in the evaluation for each term w.

2. In the collection of reviews and SO labels, we find the SO label for each word-feature pair.

3. In the collection of SO labels, we find the SO label for each input tuple for the review.

The above approach for the solution is secured by using the unsupervised collective classification mechanism type of machine learning algorithm, later polarity is customized by comparing against two baseline methods **PMI++** and **Hu++**.

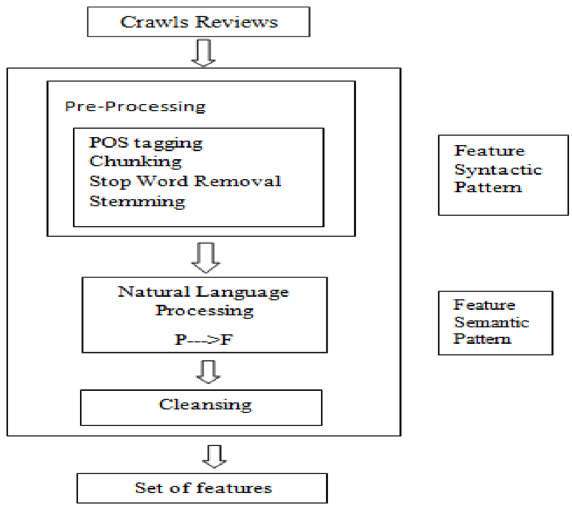


Fig 2: Workflow of level 1 (Extracting the features from the reviews)

Establishing feature similarity through k-means clustering algorithm: (Phase 1.3) Level 3:

The similarity among the features is found in the content based filtering algorithm. Each product is compared with other products. The product p1 with its extracted feature based opinion is compared with other products to exhibit similarity (Refer fig 3). The opinion mined features of product p1 is compared with all others products if the same opinion for the features are found in the other then they are grouped together. ([3],[4]) Then the same feature products are clustered into a group by using semi supervised k-means clustering algorithm.

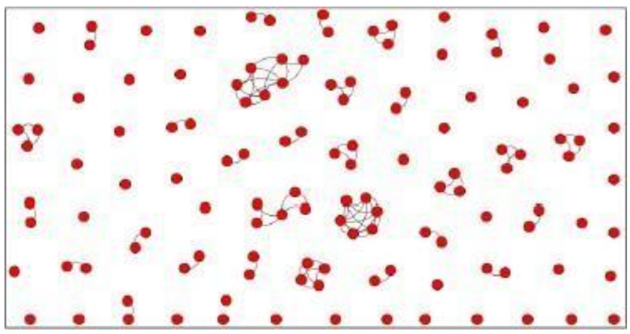


Fig 3: Working of K-means Clustering Algorithm

OPEN ACCESS
International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Ranking them through collaborative filtering algorithm: (Phase 1.4) Level 4:

The clustered similar products are then recommended to the user through collaborative filtering algorithm. The clusters are passed as the input and rank prediction is done. The clusters are mapped based on the user opinion by memory based CF. The similar users are found by utilizing user based collaborative filtering algorithm. [7] The items are mapped based on the user taste and the rank is predicted based on the user's interest about the product. Then we recommend these items to similar set of users (refer Fig4).

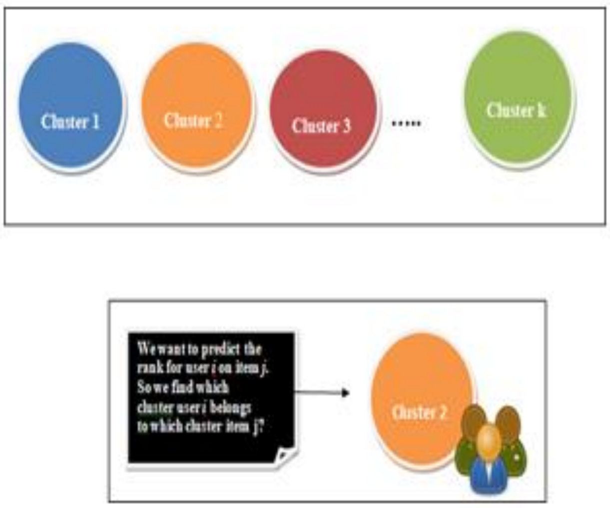


Fig 4: Ranking through CF

Performing the No-Random Access algorithm on the results sets to find best product: Phase 2 (Level 5):

The top-k best product is obtained by performing NRA algorithm. The NRA aggregates multiple values into single values (refer fig 5). The ranks produced from the recommended system and user centric search are aggregate through NRA where the ranks of a product in both systems are added and mean value is found. After getting a globalized rank we perform NRA where a threshold value is given and the ranks above the threshold are the top-k best products.

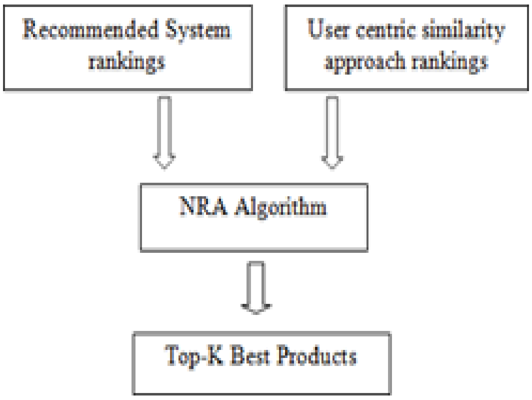


Fig 5: Top-k Best product through NRA algorithm

IV. EXPERIMENTAL SETUP & RESULTS

DATASET:

The dataset consists of reviews of various products from amazon, flipkart, snapdeal. Reviews include information of product and user ratings, a plain text feedback. There are various information about types of laptop and mobile phones.

- **Item**:Wether it's a product, film, webpage, or piece of d ata, something that is recommended.
- **User**: An individual rating of products and in turn, receiving suggestions for new items.
- **Rating**: An expression of a user's choice for products

V. ALGORITHM:

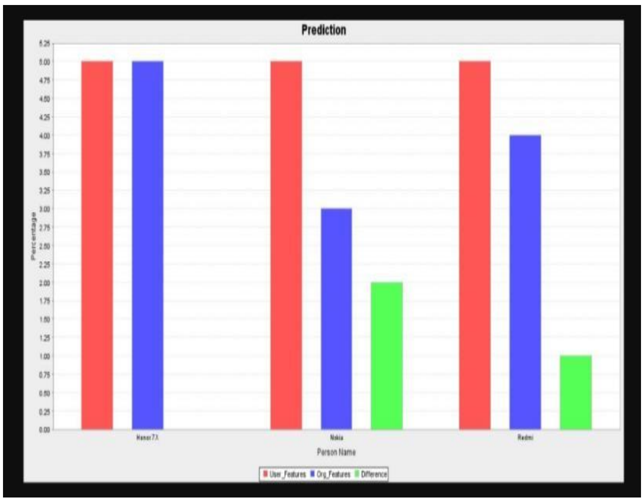
Input: An item set I, a user set U, and a rating matrix R ∈ RM×N. A set of rated items

Iu⊆ I for each user u ∈ U. The maximal number of iterations max Iteration and error threshold ϵ.

Output: A ranking ˆτu of items for each user u ∈ U.

for u ∈ Udo
for v ∈ U and u = v do
Pu, Pv ← TopKProDist(Iu, Iv , R) /* Eq.6*/
sim(u, v) ← Similarity(Pu, Pv) /* Eq.7*/
end
Nu ← SelectNeighbors({sim(u, v)}v∈U/u)
end
for u ∈ U do
t = 1
repeat
ε = 0
Initialize (φ0 u)
for g ∈ GTu k do
φ u,g ← Update(Nu, sim, R) /* Eq.13 */
ε+ = (φt u,g − φt−1 u,g)2
end
t ← t + 1
until t > max Iteration or ε<ϵ;
for t ∈ Tu do
P(t) ← Aggregation({φu,g}g∈G Tu
K
)
end
τˆu ←
end

EXPECTED OUTCOME RESULT:



Retrieval Number C6851098319/2019©BEIESP
DOI: 10.35940/ijrte.C6851.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VI. RELATED WORK

[5] Shuaiqiang Wang [et.al At 2017] Collaborative filtering has often been used to provide custo mers with individualized services. The interpretation of this strategy is to use the user item matrix to discover homogene ous users or products so that the system can exhort users.Mo st of the similarities used in this strategy are still based on si milarity algorithms such as cosine, mean squared difference and coefficient of pearson correlation. In the event of cold u ser circumstances, these techniques are difficult. This paper conferred a new user similarity model to improve recommendation performance when there are insufficient ratings available to individual users to calculate the similarities. The paradigm takes into account not only the local user ratings context data, but also the global user tone preference. The similarity correlation results shows the dominance of the ne similarity model with good performance.

[6] Kunpeng Zhang Ramanathan Narayanan [et .al at 2010] online shopping are preferred by enormous number of customer because of its cost and reliability. As the amount of products purchased online rises, buying ch oices based on images and brief item descriptions are gradua lly becoming enormous for customers. On the other hand, purchasing decisions can be made using the features of the customer reviews and the information derived from comparing the various products. To enrich the decision making of purchasing the products, online retailers such as Amazon.com1 add reviews of purcha sed products to thecustomers. To help other customers, these reviews have become a contrasting and reliable source. Traditionally, the products are ranked based on their quality. However, each of the products has multiple features and hence, the user preferences vary as each customer may be interested in different product features. Thus, the products are ranked using a feature-based ranking technique. In a product category, they first recognize product characteristics and analyze their comparative use and frequencies. We then, recognize the subjective and comparative phrases for each feature in the review. These features are authorized for sentimental orientations. They model the relationships among products by using the information gathered from customer reviews. Relative quality of products was determined by mining these graphs

VII. CONCLUSION

A recommender system e-commerce personalized web sites. In our work, we have explored a feature similarity method k-means clustering algorithm, between the features extracted from user opinions. And we used atypical No-Random Access algorithm to find the top-k best product among the several products. The outcome of the proposed system has desirable enhancement for the users to interact with the website. The result of this system provides top-k best product which helps the new users. The results showed our approach is more definite than traditional system.

REFERENCE

1. Akrivi Vlachou , Christos Doulkeridis , Yannis Kotidis, Kjetil Nørvåg, “Reverse Top-k Queries,” Department of Computer Science, NTNU, Trondheim, Norway.
2. Muhammad Zubair Asghar, Aurangzeb Khan, Shakeel Ahmad, Fazal Masud Kundi, “A Review of Feature Extraction in Sentiment Analysis,” Journal of Basic and Applied Scientific Research in 2014.
3. Zhongwu Zhai, Bing Liu, Hua Xu and PeifaJia, “Grouping Product Features Using Semi-Supervised Learning with Soft- Constraints,” Dept. of Comp. Sci. University of Illinois at Chicago.
4. Zhongwu Zhai, Bing Liu, Hua Xu, PeifaJia, “Clustering Product Features for Opinion Mining,” Dept. of Comp. Sci., University of Illinois at Chicago.
5. Shuaiqiang Wang, Shanshan Huang, Tie-Yan Liu, Jun Ma And Zhumin Chen, Jari Veijalainen, “Ranking- Oriented Collaborative Filtering: A Listwise Approach,” University of Jyvaskyla, Shandong University, Microsoft Research Asia.
6. Kunpeng Zhang, Ramanathan Narayanan and Alok Choudhary, “Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking,” Electrical Engineering and Computer Science Department Northwestern University.
7. Shuaiqiang Wang, University of Jyvaskyla Shanshan Huang, Shandong University Tie-Ya Liu, Microsoft Research Asia Jun Ma and Zhumin Chen, Shandong University.” Ranking-oriented CF algorithms demonstrated significant performance gains in terms of ranking accuracy”

AUTHORS PROFILE

Karpaga Priyaa K has received Master of Engineering degree in CSE from Anna University. At present she is Assistant professor in CSE department in Sri Sairam Engineering College. Her area of interest is under Machine Learning and Data Mining.

Saranya C has received Master of Engineering degree in Software Engineering from Anna University. At present she is Assistant professor in CSE department in Sri Sairam Engineering College. Her area of interest is under Artificial Intelligence and Data Mining.

PA.Yogaalakshmi has received Bachelor of Engineering Degree in CSE from Sri Sairam Engineering College, Anna University.. Her area of interest is under Machine Learning and Data Mining

M.Haripriya has received Bachelor of Engineering Degree in CSE from Sri Sairam Engineering College, Anna University.. Her area of interest is under Machine Learning and Data Mining