An Item-Oriented Algorithm on Cold-start Problem in Recommendation System

Mohammad Daoud
[1]Department of Computer Science and IT, Singhania University, Jhunjhunu, Rajasthan- 333515, India
[2]Centre for Education Technology, Indian Institute of Technology, New Delhi-110016, India

S.K Naqvi
FTK-Centre for Information Technology, Jamia Millia Islamia, New Delhi-110025, India

Tahir Siddqi
Systems and Paradigms Consulting, New Delhi-110025, India

ABSTRACT
Recommending new items is an important, yet challenging problem due to the lack of preference history for the new items. To handle this problem, the existing system uses the popular core techniques like collaborative filtering, content-based filtering and combinations of these. In this paper, we propose a market-based approach for seeding recommendations for new items in which new items will reach the audience quickest. To support this approach we purposed the algorithm that match the new item specification (features) with the existing item and identify whether these features are available in existing item sets or not. The proposed system identifies the user opinion on new item feature those are available in existing item set and generates the quality report of newly launched item (which is not purchased yet).

Keywords
Cold start problem; ecommerce, recommendation system, opinion

1. INTRODUCTION
Recommender frameworks are currently an essential piece of online site or E-business. They are exceptionally valuable in prescribing things or items to client or consumer as indicated by their preferences. The source of recommendation can be followed back to systems as cognitive science, rough guess hypothesis, data recovery and management science. The profits of having a recommender framework are cross-selling, personalization, keeping the consumers opinion on products and customer retention. Some of the websites that use recommenders are Amazon, eBay, CDNow, MovieLens, MovieFinder. Amazon (amazon.com) used the collaborative filtering approach for recommendation, in which the system recommends new items to the consumer by analyzing items bought by similar consumers. Pandora Radio (pandora.com) used the content based approach, which proposes items with comparative substance to the items favored by the target consumers or user. Netflix (netflix.com) used the hybrid approaches, in this system used both the content based and collaborative approaches are utilized to give recommendations. These methodologies give the consumers various recommendations or suggestions [19].

The Collaborative filtering (CF) has been exceptionally effective in both information filtering domain and E-business [20]. While the CF recommenders have been connected in ecommerce. Researchers take a substitute known as product based CF which delivers recommendations for consumer by discovering products which are like the products the consumer preferred earlier, and would have a tendency to maintain a strategic distance from the products that are near the products the consumer didn't like in the recent past [21,22]. But One troublesome, however regular, issue for a recommendation framework is the cold start problem, where suggestions are needed for products that no one (in information set) has yet evaluated or rated. Pure collaborative filtering cannot help in a cold start setting, since no consumer inclination data is accessible to structure any premise for recommendations, where presence of consumers and things without evaluations is likely or have quite few ratings available. Collaborative filtering can’t work appropriately at all in such circumstance. The cold-start problem is additionally called new user problem or new item problem or new system problem [23]. New item can’t be recommended until a few consumers have evaluated it, new consumers are improbable given great suggestions due to the absence of rating or buy history and new framework that don't have the consumer past shopping exchange information.

Various methods exist for addressing the cold start problem. Some of these methods are based on association rules, clustering, classification and so on. Many hybrid recommenders additionally exist for solving this issue. However, content information can help cross over any barrier from existing things to new things, by surmising likenesses among them. Thus researchers make recommendations for new items that appear similar to other recommended items. But all these methods are depend on product similarity and consumer previous transaction data. The consumer shopping taste and requirements change over the time and system did not know also about consumer shopping aim that she/he purchasing product for him or some other person. So we could not depend (believe) on consumer previous transaction data. The content information suggest only similar item not exact item. In this paper we purposed the system on new item that recommend or condemned the item to consumer, on the basis of existing consumer opinion record. There is no need of consumer transaction data. The system is also have the capability short out the new system problem or those startup the new online shopping business.

In this work we present FSM: **F**eature-based item-**S**imilarity **M**odels. FSM learns similarity between Existing item/device feature set and new item feature set. Existing item is related to the same type item or related item (i.e have some common feature) of target item. For example mobile is similar to mobile but its contain some common feature with other device such as tablet, laptop, camera so on. these device are called related device/item. Purposed system identifies the new item feature from existing feature set. the identified feature the system also find (get) on the exiting consumer opinion[my paper reference]. The consumer opinion is helpful to identify the product quality. The system is responsible to identify the product new feature those are not added system information set till date. These features are called newly invented feature, otherwise maximum features are available in previous item or related item. The purposed system is minimized the cold start problem without using the previous history of consumer. The key contributions of the paper are: we present a market-based approach for seeding recommendations for new items to reduce item latency, in which the item will reach the audience quicker.

Table 1: Concepts and notation used

A ⟶	the set of all devices (same type devices and related devices for example: mobile, tablet, camera, laptop, iPod so on)
T_d ⟶	the device type, $T_d \in A$ or $T_d \subset A$ (mobile or tablet or camera or laptop so on)
d_i ⟶	device which $\in T_d$, i ⟶ integer value
V_{di} ⟶	set of existing device feature vector
V_d ⟶	new device feature vector
$R_{f,di}$ ⟶	Resultant bit matrix , initially $R_{f,di}$ ⟵ zero matrix
R_v ⟶	Vector

2. RELATED WORK

New item and new consumers represent a huge test to recommendation system. All in all these issues are alluded to as the cold start problem [1]. The literature is rich with diverse classes of routines for comprehending the cold start problem proposal Issue, for example, Measurable model-based approach [2], the comparing likelihood dispersion measurements are made as per the consumer, extend and introduce rates and high likelihood items are need to be recommended [3]. However there is still the issue of low exactness in proposals in these techniques.

The first cold start issues emerge in Collaborative filtering systems, where an item can't be prescribed unless some consumer has appraised it in the recent past. This issue applies to new items, as well as to obscure items, that is especially unfavorable to consumers with diverse tastes. All items considered the new-item issue is likewise regularly alluded to as the first-rater issue [4]. The other methodologies for distinguishing which of the new items may be applicable to a consumer is the consumer demonstrating methodology proposed by Billsus and Pazzani [5]. In this proposal, the group of items that a consumer preferred/loathed in the past was utilized as the preparation set to take in a model for that consumer so as to characterize new items. The items were represented by some feature (e.g., words on account of articles) and the learning algorithms utilized these features to make the consumers model. Billsus and Pazzani[5] tried different items with two different algorithms: naive Bayes and k-closest neighbor . Thought this approach was essentially planned to assign another item into the "significant" or "unimportant" class, it can be effectively summed up to register an importance score to every item, which can then be utilized to rank the new items keeping in mind the end goal to give back where its due most pertinent items for every purchaser. In consequent years, different analysts have researched the utilization of more exceptional consumer demonstrating strategies. The work done in [6] assembled customized consumer models in the connection of grouping news feeds. This work modeled short-term consumer requirements using the text-based features of the items recently viewed by the consumer, and modeled long-term user requirements using news topics/categories. The work done in [7] that make more precise content-based utilizer framework for classifying news report by exploiting topic taxonomies and topic synonyms. Fortuitously, content data can help to cross over any barrier in the middle of existing and new consumers, and in addition in the middle of existing and new items by inducing similitude's among them. There were numerous studies utilizing substance data to join with collective in different ways [11,12,13].

Latent factor models have developed as a popular technique for developing collaborative filtering based recommender frameworks. On account of the cold start problem, these strategies incorporated the item features in the factorization process. The most general of these systems was the regression-based latent-factor models (RLFM) [8] that can work in different circumstances including item cool start. RLFM generalized the typical latent factor representation of the preference matrix by adding another step in which the user/item features were transformed to the latent space using linear regression. An augmentation of the RLFM model was proposed in [9] where more adaptable relapse models were connected. On the other hand, the extra upgrades were limited. A methodology to learn attribute to feature mapping (AFM) was proposed in[10]. Item cold start was managed by first learning a factorization of the preference matrix into user and item latent factors. Park and Tuzhilin propose the clustering of long tail items (i.e. items with below a user defined number of ratings)[14]. Clustering is focused around a situated of inferred consumer and item qualities, for example, the normal rating, and popularity of items appraised and amiability of items. Their results recommend enhanced precision is achievable through clustering long tail items. In [15] author tries to address the cold start issue as an issue undertaking by proposing a pair wise preference regression model and in this manner minimizing the separation between the genuine rank of the items and the evaluated one for every consumer Work [16] takes after the thought of dynamically questioning consumer reactions through an introductory meeting procedure. [17] is focused around a meeting procedure, as well, by proposing a algorithm that figures out how to direct the meeting methodology guided by a choice tree with various inquiries at each one part. [18] plans to suggest applications to the Twitter consumers by applying latent Dirichlet allocation to create latent groups based on the users' supporters, so as to defeat the trouble of cold-start app recommendation. In [26] researcher purpose the model on cold start problem under this model they predict the maximum similarity (based on cosine similarity formula) between the existing products and new product or targeted product. In this research they selected limited product i.e those product have the maximum similarity with targeted or new product. They selected opinion only those features which belong to the new product feature only. Finally they find opinion (positive or negitive) of existing user.

3. PROPOSED WORK

In this model initially we prepared the data for purposed system, first we make the feature set of the existing item/device (V_{di}), the feature order will be remain same for all items. Those features exist in feature set the value of these features is 1 otherwise 0.If new feature has occurred of any particular new item/device, then the new feature adjusted at the end position of V_{di} and the value become 0 of the existing items related these features.

We assume that a new item has launched in the market, in that case the new item has some new feature those don't exist in our data set or these features are available in different device (same type device or related device) here only two possibilities available for new items, these areas follows.

$$\left\{\begin{array}{l} \textit{if } len(V_d) == len(V_{di}) \\ \textit{then ,} \\ V_{di} \textit{ remain same.} \\ \textit{else } len(V_{di}) < len(V_d) \\ \textit{then , } V_{di} = V_d \end{array}\right. \qquad (1)$$

The equation (1) shows the two conditions in first condition both vector are equal length. That means new device feature vector (V_d) are available in existing feature vector set (V_{di}). V_d contains all those features that are available in V_{di} . V_d is made to parallel of V_{di}. So V_d feature order remains same to the V_{di} and those features are available in V_d in comparison of V_{di} the value is 1 otherwise 0 . The second condition the length of V_d is greater than V_{di} that means V_d contain some new features. If some new feature has occurred, then these added the end position of V_d. so len (V_{di}) < len(V_d) and updated existing feature set ($V_{di} = V_d$). See the whole representation in figure [1].

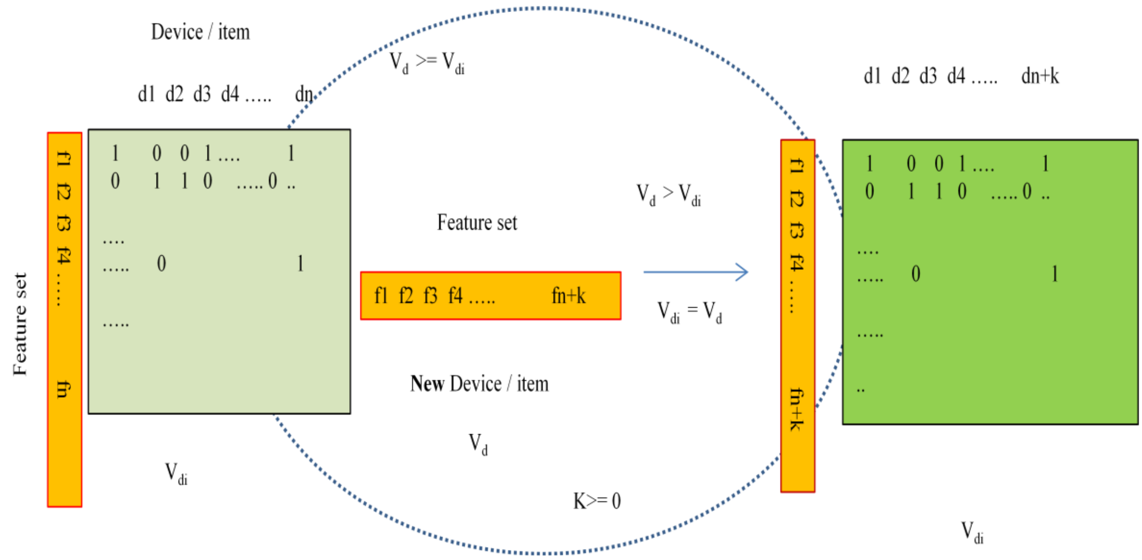

Fig. 1

We have already discussed above the types of devices one is similar type and another is related type device see in table 1. We purposed the Algorithm on new cold start problem that shows in algorithm 3.1. First, we find the V_d in

similar device type (T_d). For example, suppose V_d , is mobile, then V_d first find the its specification in other mobile (V_{di}). We performed the bitwise AND operation between V_{di} and V_d and bitwise OR operation with $R_{f,di}$ and store these values in $R_{f,di}$ .we used the OR operation between each device d_i (d_1 OR d_2 OR d_3 OR d_4 so on) of $R_{f,di.}$ if we obtained the unit vector (R_v), then we break our process because we find the all feature of V_d in V_{di} .

Suppose that if we did not get the unit vector then, we find V_d in related device type.

In a second step initially we find the similarity between V_d and all related device. Those devices have the maximum similarity with V_d we give the first priority. For example supposes V_d = mobile, then we find similarity score with the related device (tablet, iPod, laptop, and camera so on). Again, we follow the process from the step-1 till we identify the unity vector. In the worst case our system is unable to find the unit vector at grand (final) level mining. However, our system reduced the cold start problem at the maximum level of the new item launch.

3.1 Algorithms

for each related device type T_d in A
    for each device d_i in device type T_d
      $R_{f,di}$ V_d bitwise AND operation V_{di} bitwise OR operation $R_{f,di}$
    end for
          $R_v \leftarrow (d_1$ OR d_2 OR d_3 OR d_4 OR ….. $d_n)$ where $d_i \in R_{f,d}$
            if R_v is unit vector
              break
            end if
  end for

3.2 Proposed System Architecture

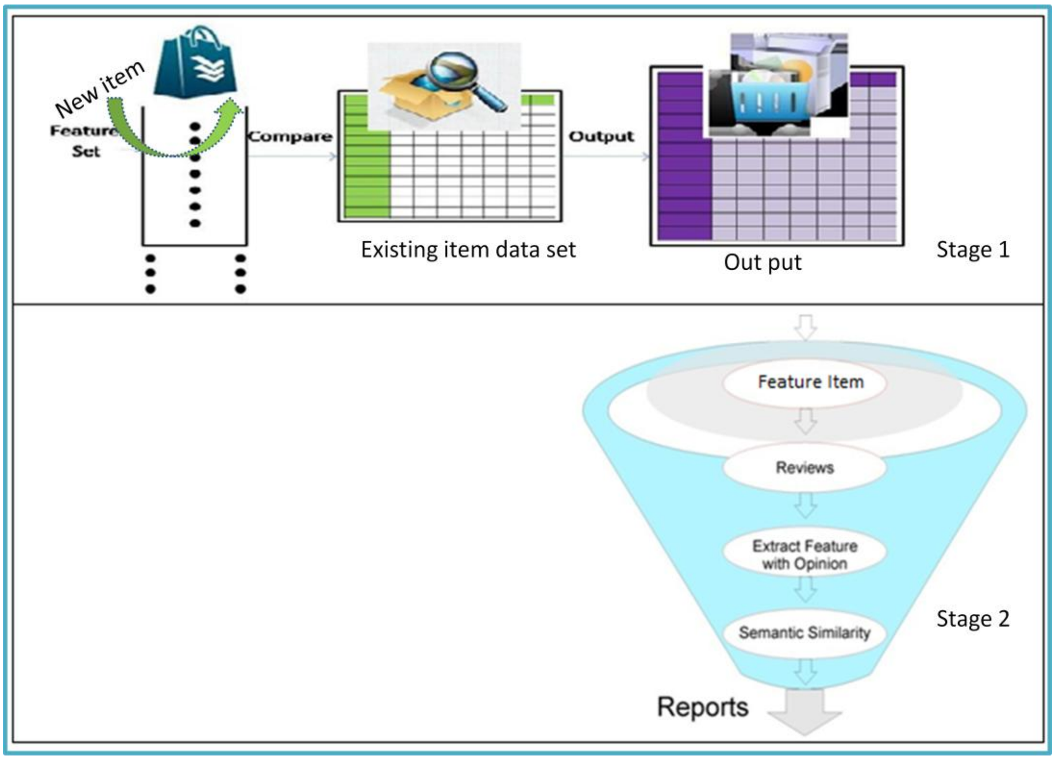

Fig. 2

The proposed system identifies the user opinion on new item feature those are available in a data set and also identify those features which are invented newly (not present in item features set). The purposed system architecture show in Fig.2. that is divided into two stages. In the first stage we find the new item feature in available data set, to perform the task we developed the algorithm. In the second stage we find the existing user opinion on these features. This task we are already done [24,25]. See the result in Fig [3] of existing user opinion on a product feature[24].

The proposed system identifies existing user opinion based on features of the new product. The system generates a report that's based on the user opinion and show all positive and negative opinion results see in figure [3]. Based on this report we can predict whether the newly launched product is recommendable to other users or not. The purpose of recommendation system is to reduce item latency, in which the item will reach the other users quicker.

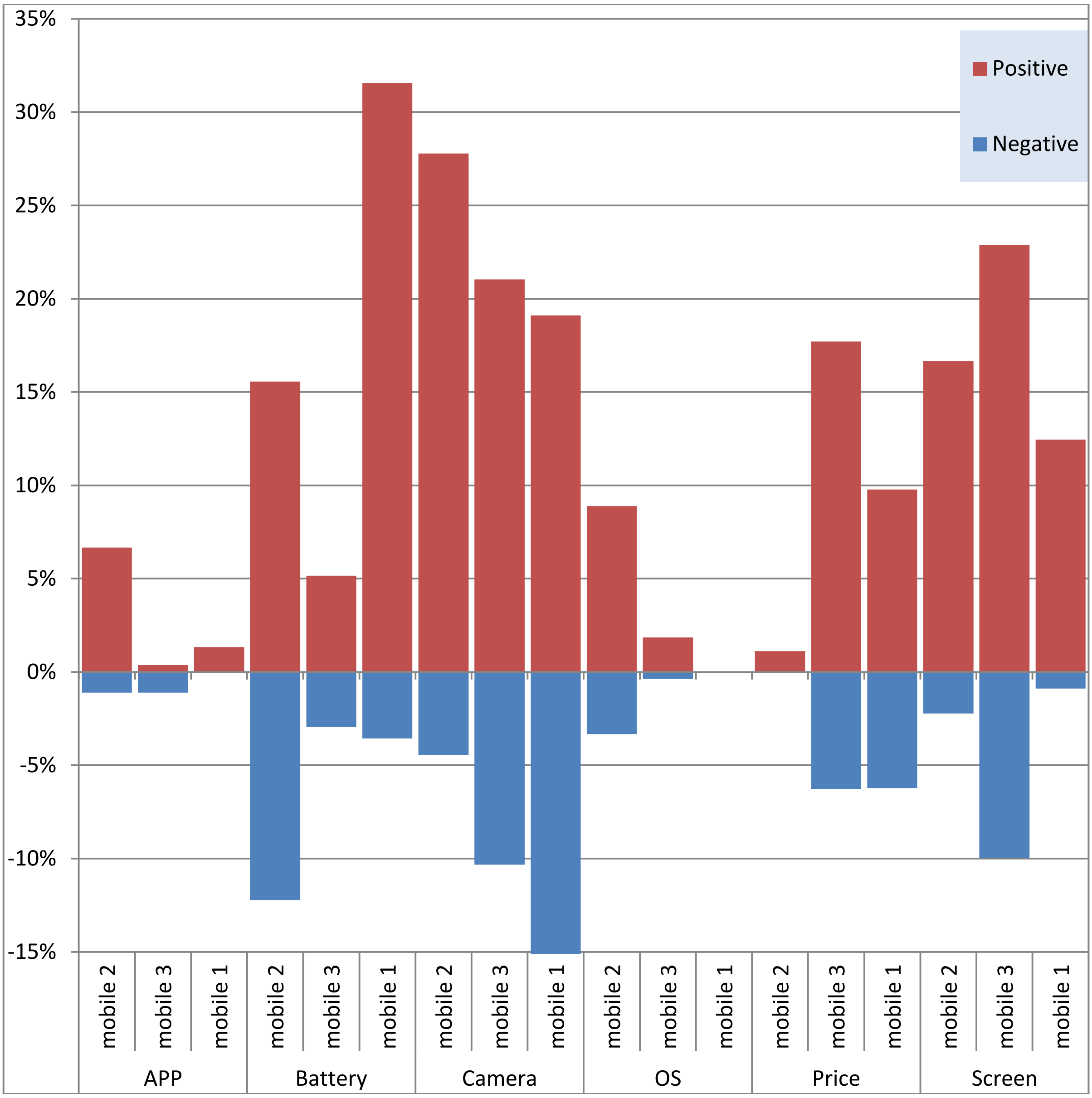

Fig.3. existing consumer opinion on product features [24]

4. CONCLUSION

In this paper, we have presented an item-oriented algorithm on cold-start problem in recommendation system that identifies the new item specification from the existing or related item. To identify user opinion on new item specification we used our previous work [24].

The proposed work reduce the item latency to Consumers and retailer, in which the item reached in audience quicker because the consumer will be familiar with the new item quality before purchasing.

5. REFERENCE

[1] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. Methods and metrics for cold-start recommendations. In SIGIR '02:Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pages 253–260, New York,NY, USA, 2002. ACM.

[2] X. N. Lam, T. Vu, T. D. Le, A. D. Duong.Addressing Cold-Start Problem in Recommenda-tion Systems. Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication (ICUIMC'08), New York, USA, 2008: 208-211.

[3] A Method to Solve Cold-Start Problem in Recommendation System based on Social Network Sub-community and Ontology Decision Modelm

[4] Recommender Systems, Prem Melville and Vikas Sindhwani IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

[5] Daniel Billsus and Michael J. Pazzani. 1999. A hybrid user model for news story classification. In Proceedings of the seventh international conference on User modeling. 99–108.

[6] Manuel de Buenaga Rodr´ıguez, Manuel J. Ma˜na L´opez, Alberto D´ıaz Esteban, and Pablo Gerv´as G´omezNavarro. 2001. A User Model Based on Content Analysis for the Intelligent Personalization of a News Service. In Proceedings of the 8th International Conference on User Modeling 2001 (UM '01). SpringerVerlag, London, UK, UK, 216–218. http://dl.acm.org/citation.cfm?id=647664.733412

[7] E. Banos, I. Katakis, N. Bassiliades, G. Tsoumakas, and I. Vlahavas. 2006. PersoNews: a personalized news reader enhanced by machine learning and semantic filtering. In Proceedings of the 2006 Confederated international conference on On the Move to Meaningful Internet Systems: CoopIS, DOA, GADA, and ODBASE - Volume Part I (ODBASE'06/OTM'06). Springer-Verlag, Berlin, Heidelberg, 975–982. DOI:http://dx.doi.org/10.1007/11914853 62

[8] Deepak Agarwal and Bee-Chung Chen. 2009. Regression-based latent factor models. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09). ACM,19–28.

[9] Liang Zhang, Deepak Agarwal, and Bee-Chung Chen. 2011. Generalizing matrix factorization through flexible regression priors. In Proceedings of the fifth ACM conference on Recommender systems (RecSys '11). ACM, New York, NY, USA, 13–20. DOI:http://dx.doi.org/10.1145/2043932.2043940

[10] Zeno Gantner, Lucas Drumond, Christoph Freudenthaler, Steffen Rendle, and Schmidt-Thie Lars. 2010. Learning Attribute-to-Feature Mappings for Cold-Start Recommendations. In Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10). IEEE Computer Society, Washington, DC, USA, 176–185.

[11] W. W. Cohen, R. E. Schapire, and Y. Singer, "Learning to order things," Journal of Artificial Intelligence Research, 243{270, 1999.

[12] A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence, "Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments," Proceedings of the Seventeenth Conference on Uncertainty in Ariti¯cial Intelligence, 2001.

[13] R. Jin, L. Si, and C. Zhai, "A Study of Mixture Models for Collaborative Filtering," Journal of Information Retrieval, 2006.

[14] Park and Tuzhilin 2008] Park, Y.-J. and Tuzhilin, A. 2008. The long tail of recommender systems and how to leverage it. In Proc. of the 2008 ACM Conf. on Recommender systems. 11-18.

[15] S.-T. Park and W. Chu. Pairwise preference regression for cold-start recommendation. In RecSys'09, pages 21–28, 2009.

[16] K. Zhou, S.-H. Yang, and H. Zha. Functional matrix factorizations for cold-start recommendation. In *SIGIR'11*, pages 315–324, 2011.

[17] M. Sun, F. Li, J. Lee, K. Zhou, G. Lebanon, and H. Zha. Learning multiple-question decision trees for cold-start recommendation. In *WSDM'13*, pages 445–454, 2013.

[18] J. Lin, K. Sugiyama, M.-Y. Kan, and T.-S. Chua. Addressing cold-start in app recommendation: Latent user models constructed from twitter followers. In SIGIR'13, pages 283–293, 2013.

[19] Li, Q., & Kim, B. M. (2003, October). Clustering approach for hybrid recommender system. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on* (pp. 33-38). IEEE.

[20] J. B. Schafer, K. J., and J. Riedl. "Recommender systems in ecommerce," In Proceedings of the ACM Conference on Electronic Commerce, 1999.

[21] B. Sarwar, G. Karypis, J.Konstan, and J.riedl, "Item-based collaborative filtering recommendation algorithm," WWW10, ACM, Hong Kong, May 1-5, 2001, pp. 285-295.

[22] M. Deshpande and G. Karypis, "Item-based top-N recom-mendations algorithms," ACM Transactions on Information Systems, vol. 22, no. 1,pp. 143–177, 2004.

[23] X.Y Su and T. M. Khoshgoftaar "A Survey of Collaborative Filtering Techniques," Advances in Artificial Intelligence, pp. 1-19, August 2009.

[24] Daoud, M., Naqvi, S. K., & Ahmad, A. (2014). Opinion Observer: Recommendation System on ECommerce Website. *International Journal of Computer Applications*, *105*.

[25] Daoud, M., Naqvi, S. K., & Jha, A. N. Semantic Analysis of Context Aware Recommendation techniques.

[26] Daoud, M., Naqvi, S. K. (2015). Recommendation System Techniques in Ecommerce System