International Journal of Engineering Trends and Technology (IJETT) – Special Issue – April 2017

The Survey of Recommender Systems

Soanpet . Sree Lakshmi
Asst.Prof (SS)
Department of CA
Vasavi College of Engineering
HYD-31

Dr.T .Adi Lakshmi
Professor and Head
Department of CSE
Vasavi College of Engineering
HYD-31

Abstract:

This paper is the survey of Recommender systems. Recommender systems (RS) are useful to make suggestions to the users. The RS vary largely according to the domain in which they are applicable. The broad categories of the RS are discussed along with their limitations. The various issues of RS that can be addressed are identified and tabularised briefly for the crisp understanding. It also discusses certain aspects of Long Tail issues and context aware systems in specific.

Keywords: Recommendation systems, Issues, Long tail, Context aware Systems.

Introduction:

Recommendation systems (RS) serve the right item to the user in an automated fashion to satisfy user and to improve businesses..

Most commercial RS are the collaborative, query less discovery engines. They have become important area of research. There are many practical applications in this area. Book recommendations, Music Recommendations, Movie recommendations, Tourism, Medical applications to name a few. Each practical problem needs a different approach to cater to its specific dimension.

Major task of the recommender system is to present recommendations to users. The task is usually conducted by first predicting a user's ratings for each item and then ranking all items in descending order.

Outline of Recommendation systems:

Categories:

Recommender systems have their relevance to information retrieval in different areas.[1]. They have been continuously researched upon and have vast potential of improving the business. The engine in such software gives advice about what we might enjoy listening to or watching or reading next, based on what you/ others like you have just listened to or watched or read. Most RS depend upon the user rating structure.

RS are usually classified into 4 categories, based on how recommendations are made:

- Content-based recommendations: User will be recommended items similar to the ones preferred in the past. Personalized Recommendations. [1]
- Collaborative recommendations: User will be recommended items that people with similar tastes and preferences liked in the past. [2]

  (a) User similarity is considered

  (b) Item similarity is considered

- Knowledge based: Use a Knowledge component model and recommend user what he needs. [14]
  Eg: PCA, Association rule mining, Clustering models, multilayer perceptron, Latent semantic analysis
- Hybrid Approaches: These approaches combine various methods[14]
  1) Monolithic hybridization (ensemble)
  2) Parallelized hybridization
  3) Pipelined hybridization

Challenges / limitations with these systems :

(a) With Content -based recommendations:

1. Over specialization : System does not recommend the items that are different from anything that the user has seen before. Sometimes this might become problem because the user might want to try something new and the system would never make it happen. Serendipities (variety in recommendations ...) are ignored.

So for this the user must be presented with range of options and not only few selected alternatives be made available.

2. Limited Content analysis problem;

- In this we might represent 2 different items with same set of attributes and they hence cannot be differentiated i.e., the data captured may not be sufficient enough to distinguish two similar but different items.

- Content may not be automatically extractable.

3. New User Problem: New users don't have sufficient ratings before so he would not be able to get accurate recommendations.

(b)With Collaboration based recommendations:

1. New user problem
2. New Item Problem
3. Sparsity in ratings
4. Impact of power users lead to power user attack.
5. Assumption: Users having similar preferences previously will continue to have same preferences.
6. Measure of similarity between users.
   - Mostly used Pearson's correlation to find user similarity and can hence predict current users preference.
7. How do we measure a similarity between items.
   -Mostly used cosine similarity to find user similarity and can hence predict current user's preference.
   - Also pre-compute the item similarity **(considered more stable compared to user similarity measure)**

(c) With Knowledge based approaches:

1. Model needs to be trained-training effort.
2. Only the trained model is allowed to make recommendation.
3. Models have to be updated and re-trained – hence expensive to build and maintain.

(d) With Hybrid Approaches:

1. Conflict resolution is the major issue with Hybrid approaches.
2. They suffer from the limitations of both approaches while enjoying the advantages of both approaches.

Context Aware Recommendation Systems:

How Contextual Factors Change	**Knowledge of the RS about the Contextual Factors**		
	Fully Observable	**Partially Observable**	**Unobservable**
Static	Everything known about context	**Partial and static context knowledge**	**Latent Knowledge of context**
Dynamic	**Context Relevance is dynamic**	**Partial and Dynamic context knowledge**	Nothing Is Known about the context

Approaches:

Contextual pre filtering (or contextualization of recommendation input). In this paradigm information about the current context is used for selecting only the relevant set of data, and ratings are predicted using any traditional 2D recommender system on the selected data.

Contextual post filtering (or contextualization of recommendation output). In this paradigm con-textual information is initially ignored, and the ratings are predicted using any traditional 2D recommender system on the entire data. Then, the resulting set of recommendations is adjusted (contextualized) for each user using the contextual information.

Contextual modelling (or contextualization of recommendation function). In this paradigm contextual information is used directly in the modelling technique as part of the rating estimation.

Motivation :

Identified Issues which needs to be addressed in recommendation systems through the literature survey [1] to [14]

Sl.no	**Issue**	**Description**
1.	Cold start problem	New User(force user to rate few items) ,New Items
2.	Scalability of the approach	More users than items
3.	Recommending the items Long tail	Niche Items
4.	Accuracy of the prediction	General items Vs. Controversial items accuracy is hard to determine
5.	Sparse / missing data	- means few common ratings – hard to use common missing value corrections
6.	Erroneous and malicious data	Identification issues are difficult in large scale datasets
7.	Hybrid approaches	Conflict resolution while using ensemble/ hybrid approaches
8.	Ranking of the recommendations	-based on the predicted ratings – -further alter ranks based on user context - more similar -more rank , less similar- less rank .
9.	Impact of temporality	Need various approaches like recency determination, validity measures – like in news recommendations
10.	Impact of context-awareness	various approaches to Model context Context modelling varies largely depending on specific domain.
11.	Mobility and Pervasiveness	Provide services based on location.
12.	Text recommendations	use of TF-IDF (term Frequency and Inverse Document Frequency)- Length of the document makes impact , hence has to be normalized
13.	Pre-processing	Dimensionality reduction –Singular value decomposition (SVD), Data compression techniques, Latent semantic Indexing.
14.	Model Maintenance	- Incremental updates possible
15.	Ratings	– granularity (likert response scales 1-7),multidimensional ratings , -not all users provide ratings, implicitly calculate ratings(click through data), Other sources to learn user preferences.
16.	Interactive models	Use of query based model to tune recommendations
17.	Novelty and diversity of recommendation	Issues to be worked on are providing reasons for recommending. -Graph based approaches.
18.	Privacy Preservation.	To Address the privacy of the data captured for recommendation
19.	Big-data	Velocity, volume and variety of the huge data generated. Various methods for distributed processing of big data analytics are being studied.
20.	Tag based recommendations	Genres, tags – if not meaningful, or are many, then becomes complex – simple cases matrix factorization methods are useful.

Related work:

(a) The Long tail Phenomenon :

Recommendation in the physical world is fairly simple. It is not possible to tailor the store to each individual customer. Thus, the choice of what is made available is done only by the aggregate numbers.

The difference between the physical and on-line worlds has been called the long tail phenomenon, the long-tail phenomenon forces on-line institutions to recommend items to individual users.

Anderson in his book [3] coined a term-"The Long Tail"- to describe the phenomenon that niche products can grow to become a large share of total sales. In the book, he claimed that Internet technologies have made it easier for consumers to find and buy niche products, which renders a shift from the hit market into the niche market.

In [2]authors proposed a novel suite of graph-based algorithms for the long tail recommendation. Using user-item information with undirected edge-weighted graph for long tail item recommendation. To improve recommendation diversity and accuracy to help users find their favorite long tail items. Matrix factorization models may not directly applicable to items in the long tail as the preserve only the principal component factors ignoring the latent factors.
The authors propose suite of novel algorithms to capture the latent information about items in long tail to improve recommendation diversity and accuracy.

In paper [5] ,author discussed the adaptive clustering algorithm to improve recommendations for items in the long tail in which , depending upon the data the items in the long tail are clustered and a recommendation is set out for the items in the entire cluster. a study of the Long tail Problem of Recommendation Systems when many items in the long tail have only a few ratings, thus making it hard to use them in Recommendation Systems.

The approach presented in this paper clusters items according to their popularities, so that the Recommendations for the tail items are based on the ratings in more intensively clustered groups and for the head items are based on the ratings of individual items or groups, clustered to a lesser extent. such an adaptive clustering if done efficiently improves the Recommendation accuracy for the tail items while maintaining reasonable computational performance.

In paper [15] the authors discuss the impact of RS when Recommendations are forced to omit popular items(short head) and to use niche products only(long tail). The authors affirm empirically that the effects resulting from Item consumption may increase the utility of personalised recommendations when compared to popular recommendations. The authors demonstrated their idea based on the tourism data during the off- season Vs on- demand season.

(b) Related work in context aware RS:

User behaviour in social network is studied in [6]where the authors demonstrated the effectiveness of context-aware review helpfulness rating prediction frame work(CAP) in solving the rating prediction problem.

The paper[7] discusses the learning of the context information by explicit querying user and also from implicit learning. It also deals with one such probabilistic model that integrates user profiles , item representation and contextual information by computing the conditional probability of each item given the user profile and additional context.

A spatial topic model that captures the correlation between users movements and between user interests and the function of location to improve precision of recommendations is proposed in [8].

The importance of freshness of data is deemed more important than relevancy or context in certain recommendation systems is demonstrated in [12].

The various approaches of implementation of CARS and specifically modelling of context is discussed in[13].Context-Aware Recommendation systems (CARS) generate more relevant recommendations by adapting them to specific contextual situation of the user . This paper explores how contextual information can be used to create intelligent and useful Recommendation Systems. It provides an overview of the multi faceted notion of context, discusses several approaches for incorporating contextual information in the Recommendation process, and also illustrates the use of such approaches in several application areas where different types of contexts are exploited.

Conclusions :

The Recommender systems are of more practical interest in businesses. Hence in the survey the various issues that direct possible research are identified.

From [1] possible extension suggested is knowing more about user can improve the quality of recommendation.

From [6][7][8][12][13] we know that context-aware systems help us in improved understanding of user behaviour and hence improves the recommendation. The techniques used are pre-filtering approach, post-filtering approach and context modelling .

In addition study of items in the long tail [2][3][4][5][10] [13] [15]and their behaviour clubbed with context aware techniques improves the diversity and accuracy of recommendation for the long tail items Hence the study of "the impact of context awareness in the long tail items" is proposed.

Bibliography:

[1] *G.Adomavicius and A.Tuzhilin*, "Toward the Next Generation of Recommender Systems: A Survey of the State -of-the Art and Possible Extensions, "*IEEE Trans . Knowledge and Data Eng., vol.17, no.6,pp.734-749,June 2005.*

[2] *Hongzhi Yin Bin Cui Jing Li Junjie Yao Chen Chen* , "Challenging the Long Tail Recommendation"**,** Proceedings of the VLDB Endowment, Vol. 5, No. 9.

[3] *C. Anderson*, The Long tail, *Hyperion Press,2006.*

[4] *Y.J.Park and A.Tuzhilin*, "The Long Tail of the Recommender Systems and how to leverage it," *Proc.ACM conf. Recommender Systems, pp . 11-18, 2008.*

[5] *Y.J.Park* ,"The Adaptive Clustering Method for the Long Tail Problem of Recommender Systems*", IEEE Trans . Knowledge and Data Eng., vol.25, no.8,August 2013.*

[6] Jiliang Tang et al.,,"Context Aware Review Helpfulness Rating Prediction "ACM recsys'2013 .October 12-16.

[7] *Negar Hariri etal.*, "Query Driven Context Aware Recommendation " *ACM recsys'2013 .October 12-16.*

[8] Bo Hu , *Martin Ester ,"Spatial Topic* Modelling in online Social Media for Location Recommendation " *ACM recsys'2013 .October 12-16.*

[9]. *Erheng Zhong, Wei Fan and Qiang Yang*. 2014. user behaviour learning and transfer in composite social networks. *ACM Trans. Know.Discov . Data 8,1, Article 6 (February 2014),32 pages.*

[10] *Idan Szpektor , Aristides Gionis and Yoelle Maarek* , "Improving Recommendation for Long-Tail Queries via Templates" WWW 2011,march 28-April 1, 2011.

[11] Proceedings of ACM RecSys 2012, Dublin.

[12] *Saranya K.G and G.S Sadasivan* , "A Personalized Online News Recommendation System" International Journal of Computer Applications, 57(18): 6-14 , Nov'12

[13] *Gediminas Adomavicius , Bamshad Mobasher, Francesco Ricci and Alex Tuzhilin*. "Context-Aware Recommender Systems" , FALL 2011 AAAI .

[14] *Robert Bell Yehuda Koren*. "Recommender Systems New Approaches with Netflix Dataset" . AT&T Labs ICDM 2007

[15] *Paolo cremonesi , Franca Garzotto Roberto Pagano, Massimo quadrana,''* Recommending without Short Head *'',* www '14 companion, April 7-11, 2014