Precedent Behavioral Extraction System For Personalization Recommendation

Mahima

Abstract: Hosting a compilation of billions of videos, YouTube presents one of the leading scale and most precious videos personalization recommendation system in existence. The recommendation system works on to personalized set of videos to users based on their past actions on the website. In this paper, we highlight the some of the major challenges that the system faces and how to address them. To tackle these issues, we have proposed a Precedent Behavioral Extraction Module (PBEM), which also deals with large-scale heterogeneous information to fulfill the requirements of the potential users. PBEM approach especially focus on the remarkable performance enhancements brought by machine learning. PBEM is a new approach as it works on discovering the precise web browsing behavior from uncertain keywords and defines the semantic measurement with user recommendation of keywords within the user query

Index Terms: Recommendation, Machine Learning, Personalization Behavior.

1. INTRODUCTION

The term information surplus is approximately equal with the Internet, referring to the huge amount of data that available in electronic form on the Internet and the incapability of users to utilize it. The sovereignty to convey oneself with circulating information to the internet and Web has a series of benefits, however, the job of the user of this context is formulate more complicated not only due to the necessitate to retrieve the desire of the information to the job at hand but also due to the necessitate to necessitate the consistency and trustworthiness of the information obtainable. The necessitate to give users with information adapted to their requirements led to the expansion of several information filtering methods that create summaries of users and effort to filter huge information tributary, providing the user with only those information that it assumes to be of interest to the user. The objective of personalization recommendation system is to present users with what they desire or require without obliging them to request for it explicitly [1]. However, this does not in direct approach entail a completely mechanized system, instead it covers mechanism where the user is not capable to completely express precisely what are looking for but in act together with an automated system can guide them to items of interest. It means that a personalization recommendation system must somehow suppose what the user desired depend on either earlier or present interactions with the user. This in itself suppose that the system somehow attains information on the user and concludes what user requires are depend on this information. In these circumstances personalized recommendation system plays a vital role for providing desired information explicitly. Personalized recommendation is about influencing all obtainable information about users of the Web to convey a personal experience. The "recommendation" of these methods is at several stages staring from the availability of functional, useable information through to the inferences crafted utilizing this information and accessible domain knowledge with ontology at the time of providing the personalized recommendation for the user.

- *Mahima, Assistant Professor, Gurugram University, Gurugram, Haryana, India Mainag1310@gmail.com*

This paper has been formalized into five sections. Section 2 presents the work of eminent researchers. Section 3 defines the proposed work and Section 4 discusses the results. Section 5 finally concludes.

2. RELATED WORK

The section highlights the work of researchers and describes the unfolded issues which further needed kind of attentions. Mukamakuza et.al [1] describing the results of social media in determining the user web browsing behavior with social recommender system. It also identified the connection between vastness and identification in social recommendation system. Social recommendation approach formulates utilization of the obtainable information about social relations among users to develop the eminence of the recommendations. This work summarized the present status of recommendation system with their rating and accessing behavior and gives precious feedback on various methods. Gorgoglionea et.al [2] proposed recommender systems (RSes) to reduce the information overload for web users and give the required information retrieved more acceptable. The main issue highlighted in this paper is to confirm and convey personalized recommendations to their users by choosing among various preferences. The defined system also worked on recommendation techniques of e-commerce companies to generate and allocate recommendations to customers. Chun-Hua et.al [3] have formalized hybrid social recommender system to support the precision of recommendations by enhancing the performance of hybrid recommender interface with several types of user queries. It is knowledge base interface that has been examined as an approach to improve straightforwardness and expand the user satisfaction experience. The work summarized the user web interaction behavior patterns and individual reaction by utilizing various set of arithmetical methods and gives results on the basis of recommended enlightenment tools. Jiahui et.al [4] presented personalized news recommendation approach in Google News and utilized Bayesian algorithm for detecting users' present news interactions from the actions of that specific user and the news styles described in the action of each users. The authors have merged the information filtering technique by analyzed user profiles with a current collaborative filtering technique to create personalized news recommendation system. The results on the reside traffic of Google News website shown that the merged technique enhances the eminence of news recommendation system and boosts traffic to the website. Paul et.al [5] proposed deep neural network model for recommending YouTube videos and describe two different issues: candidate generation and ranking. Ranking is a standard machine learning issue yet this deep learning method outperform earlier linear and tree-based techniques for watching time calculation. The defined approach described the precedent user behavior with queries and quintile simplifications. The investigational consequences described improved on watching time subjective ranking estimation calculations evaluated to calculating click- behavior directly. Huan et.al [6] intended a generative approach of Multi-site Probabilistic Factorization (MPF) to confine both the cross as well as site-explicit predilections. The main attribute can be extract with matrix factorization by using multi-site user video utilization information. The described MPF approach assists to address the information sparsely issue, data unity issue, and personalization of user behaviors.

3. PERSONALIZATION RECOMMENDATION SYSTEM

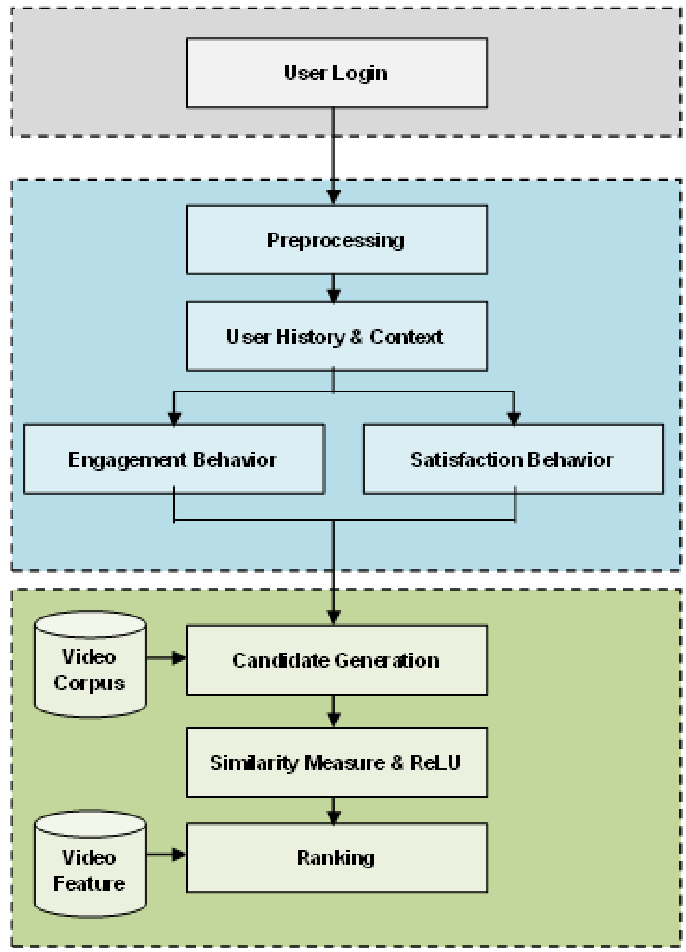
The personalization recommendation system plays a vital role in finding significant content to its users that works with machine learning algorithms. Basically, the machine learning algorithms consists of two types of algorithms. The first algorithm works on candidate generation that uses the users' video watching history and utilizes the technique of collaborative filtering technique to recommend parallel videos based on watching history. It based on user-item interactions such as ratings, number of purchases, likes, etc. The second algorithm assists evaluate videos employing logistic regression, and predicting what the user desires. This method is called Ranking. The ones which align to the regression analysis become the 'viral videos' or 'trending videos'. This algorithm based on content filtering technique to provides information about items characteristic information, which uses characteristic information such as keywords, categories and users (preferences, profiles, etc.). Precedent Behavioral Extraction Module (PBEM) system works with *Hybrid systems* that merge both types of above algorithm with intend of avoiding issues, which are formulated when working with just one kind. Presently, YouTube remains the most watched website for video contents. It presents videos wrap a list of subjective right from educational to renowned music/movies/stories/talks/games videos by artistes and famous personalities and has even extended out to gaming and most videos are obtainable for free.

4. CHALLENGES

With the rapid growth of information on internet, it is very difficult to recommender systems to extract desired and accurate information. There are several parameters are utilized to evaluate the efficiency of recommender systems such as "conversion rate" and "click through rate", however, these parameters are inefficient, due to immense video content uploaded on e-commerce website coupled with the influx of duplicated and mis-titled videos. Several times recommendation systems do best work with earlier user information such as watch history. However it moves towards to its accomplishment in real-time, systems require continuous and regulated data. Besides the above defined issues for creating real time large-scale video recommendation problems, we require to address the following additional factors: Scale: Due to the enormous sparsity of these parameters, it's complicated for measure earlier matrix factorization methods to extent between the full characteristic spaces. Additionally, earlier matrix factorization techniques have a complicated time handling a permutation of definite and permanent variables. Consistency: several content and collaboration-based algorithm works on consistency but fail to provide consistent aspiration results. Here needs to add intelligent algorithm that able to adequately simple to train, test, and set up deep neural networks in a distributed network.

5. PRECEDENT BEHAVIORAL EXTRACTION MODULE (PBEM)

The overall architecture and flow of the proposed recommendation system are illustrated in Fig. 1. PBEM is composed of two main constituents: one for learning and adapting user past search behavior and another for generating user desirable contents trends.



5.1 Preprocessing

A set of key words is given as an input to PBEM, which describes the user information needs. It performs a set of jobs to accomplish the preprocessing tasks such as Tokenization (splits the words into token), Grammatical feature (add several grammar rules), Parsing is used to examine a set of words, either in general language or in scripts based or fitting in proper formal syntax. Stop word is a process of eliminating the unwanted query words, those have contains miner value in the sentence

.

5.2 User History & Context
The second phase of PBEM is to takes the user's activity history (IDs of videos being watched, search history, and user-level demographics) and outputs a several hundred videos which might generally be appropriate to the user. The basic thought is that this module should optimize for precision and every occurrence should be extremely relevant, even if it needs forgoing several documents that can be broadly accepted but irrelevant.

5.3 Engagement Process
Engagement process is used to evaluate user behaviors such as clicks and watches. To determine a set of engagement metrics for online videos, like average watch time, average watch percentage, and a novel metric, qualified engagement that is standardized with video length measure, stable over time and interconnected with video quality features. This paper describes engagement as containing behavioral features or click-based interfaces as well as straightforward content viewing and reading behavior. Moreover, viewing videos and reading comments is also part of engagement process. Users can select to stay passive by basically consuming content, or play an active role by contributing in several interactions and even repurpose content to fit their requirements.

Step 1: *Import WordNet*

Step 2: *User login & input query*

Step 3: *Preprocessing*

$Tokenization\ words = w.split(Query)$

Part of speech, Grammer Rules & Remove Stop Words

Step 4: *To Explore user history words with engangement & other behaviour*

Step 5: *Candidate Generation*

Step 6: *Similarity index*

$$cos\emptyset = \frac{a * b}{||a||b||}$$

5.4 Candidate Generation
This takes inputs from the user's history and context module as input and extracts/presents a small subset (hundreds) of videos from a large corpus (millions) as per user curiosity. This work used hybrid filtering techniques for candidate generation. Hybrid filtering techniques is a combination of content based filtering and collaborative filtering. Content-based filtering contains a recommending item based on the attributes of the items themselves. It recommends items similar that user has liked in the earlier time/search. It utilized similarity among items to recommend items parallel to what the user likes such as If user watches two cute cat videos, then the system can recommend cute animal videos to that user. On other side, Collaborative filtering works on user item interaction behavior and works on the concept, which similar users like similar things eg users who bought this product also bought product such as if user X is similar to user Y, and user Y likes video 1, then the system can recommend video 1 to user X (even if user X hasn't seen any videos parallel to video 1).

5.5 Similarity Measure
Semantic similarity is a measure described a set of terms, where the idea of distance among them is depend on the similarity of their sense/meaning. It used to calculate the amount of the semantic similarity among sense of language and concepts through a cosine similarity measure according to the evaluation of data relating with their meaning or defining their domain. The cosine similarity measure finds similarity among two vectors (two documents on the Vector Space) and measures the cosine of the angle among them. The cosine similarity measure is to resolve the equation of the dot product for the Ø :

$$a, b = ||a||b|| \cos . \theta \quad (1)$$

$$cos\emptyset = \frac{a * b}{||a||b||} \quad (2)$$

Cosine Similarity measure generates a metric that find the related between two documents by defining at the angle instead of magnitude.

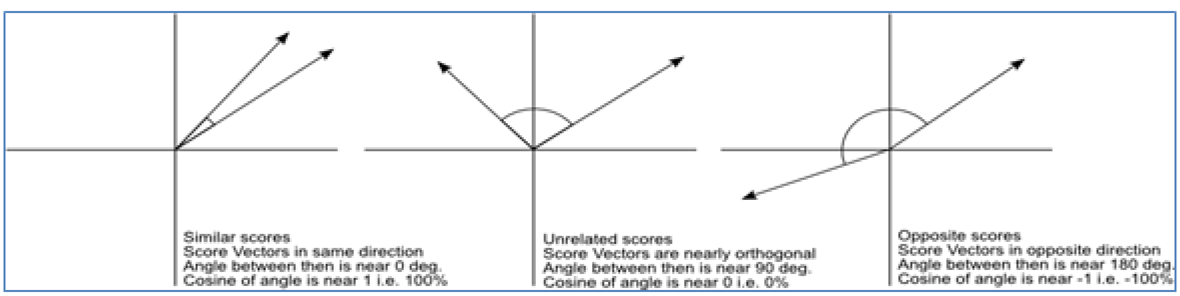


Figure 2: *The Cosine Similarity values for different documents*

PBEM initiate the utilization of rectified linear units (ReLU) as the classification method in a deep neural network (DNN). It also used as an activation function in DNNs, with Softmax function as their classification method. The methods returns 0 if it receives any negative input, but for any positive value xx it returns that value back. So it can be written as f(x)=max(0,x)f(x)=max(0,x).

5.6 Ranking
The main objective of ranking module is to utilize impression information to concentrate and standardize candidate calculations for the particular user interface.

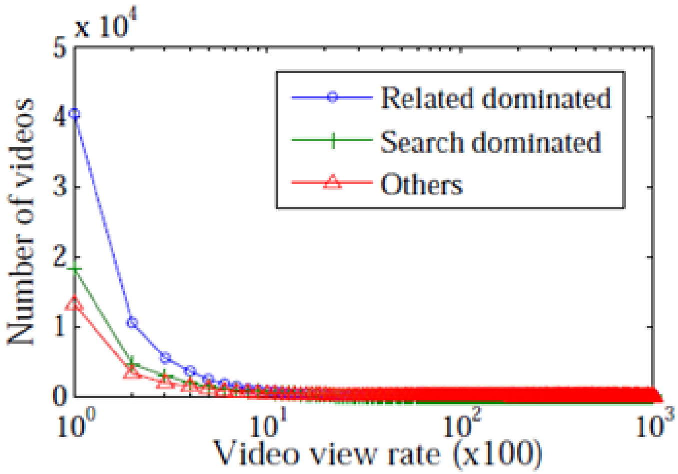


Figure 3: *Number of videos for each type after aggregation with constant view rate*

It describing more characteristics of video and the user's correlation to the video because only a few hundred videos are being ranked rather than the millions ranked in candidate generation. In this work to define ranking system utilized a deep neural network with parallel manner as candidate generation to allocate an independent score to each video sense through logistic regression. The set of videos is then sorted by ranking score and returned to the user.

6. RESULTS AND EXPERIMENTS

A user of the PBEM just requires inputting searching keywords, what types of video they are searching for (the query). The PBEM expects the video, which the user will like greatest and the video obtains conveyed to the user's login version for existing and future (the item). The input metric is utilization ratio, if a video was extremely correlated to user desire, the score is 1; otherwise it's 0. In this experiment, we collected data from the watch history of a huge number of users for a period of four weeks (28 days) and processed data of each individual user. The recommended videos are generated by the PBEM method and all result analyzed by the feedback of the users. The PBEM manually generate result for each individual users and request users to which video they may click if the video emerged in the recommendation factor of PBEM home page. Based on these users's feedback, the results are shown in.

Table 1: *Recommendation set of PBEM*

User No	Recommended	Videos Watched	Videos Ignored	Success Rate
User 1	43	28	15	65.11%
User 2	52	40	12	76.92%
User 3	33	25	8	76.76%
User 4	55	43	12	78.18%

The PBEM approach has been evaluated on approximately 200 users. According to their response they have clicked around 74% of the recommended video. The maximum results of PBEM are determined manually considering user's feedback and their feedback on current recommendation system. It is very complicated to identify user's desires as billions of user's do not sense the similar way. The results of PBEM approach show that it is accepted by most of the users.

Experimental Settings

In this section, we confer the efficiency of Precedent Behavioral Extraction Module (PBEM) by performing broad experiments on real data assembling from Video database, and express its advanced presentation in creation.

Parameters Setting: Parameters utilized in this work are identified by utilizing grid research to attain the most favorable values, as shown in Table 2.

Table 2: *Parameter Settings*

	f	λ	A	b	$\eta 0$	A	β	Ξ
Values	75	0.03	2.7	2	0.00002	2	0.02	11

Offline Experiments

To evaluate the performance of PBEM approach, we used video database and perform experiments to compare the result of the factors in approach such as efficiency of running training and the efficiency of adaptable updating strategy. This section briefly summarized the investigational settings for the assessment, including dataset and analyzing factors.

Data Sets

To conduct the experiments with proposed approach, data collected from Tencent Video database with one week process rate and preserve users who have more than 65 actions and videos with more than 70 connected actions. The system attains the data of first seven days as training data set and the eighth day as test data set. The Statistics summaries of the dataset after cleaning are describes in Table 3.

Table 3: *Data Set Statistics*

	Users	Videos	Actions	Test Actions
Counts	846929	18197	74503809	13838599

Evaluation Criterion

To evaluate the performance of PBEM, we have utilized two evaluation parameters to determine the performance. The first one is to measure performance metric for rating accuracy with recall metric and the top-N recommendation performance instead of rating prediction. PBEM is determine the quality of video personalization recommendation model with recall metric, also known as hit rate that is mostly utilized for analyzing top-N recommendation systems.
The recall metric is defined as follows:

$$\mathrm{recall} = \frac{\sum_{c \in C_{\mathrm{test}}} \sum_{ic} \mathrm{l}(\mathrm{v_c} \in \mathrm{top} - \mathrm{N_c})/\mathrm{N}}{|\mathrm{C_{test}}|} \qquad (3)$$

where i_c indicates that video v is liked by customer c in summarized data and $l(z)$ is an indicator method, which revisits 1 if state z contains and 0 otherwise. A better recall rate indicates which the PBEM is capable to commend further acceptable videos and providing to enhanced results. The other analyzing factor, PBEM utilized is average rank. We define $rank_{ci}$ as the percentile ranking of video *v* with the standard set of each videos recommendation for customer *c* and $rank_{ci}$ are defined as the percentile ranking of video *v* with the set of desired video set of customer *c* in test data. The $rank_{ci}$ = 0% defines that video *v* is determined to be the best satisfactory results for customer *c* while $rank_{ci}$ = 100% defines that video *c* is not recommended to customer *c*, as the similar container for $rank_{ci}^{t}\, s$. The average rank factor is described as the average percentile ranking of personalization recommended videos that is:

$$\mathrm{rank} = \frac{\sum_{\mathrm{c,v}} \mathrm{rank}_{\mathrm{cv}}^{\mathrm{t}}\,(1 - \mathrm{rank}_{\mathrm{cv}})}{\sum_{\mathrm{u,v}}\,(1 - \mathrm{rank}_{\mathrm{cv}})} \qquad (4)$$

where $1 - rank_{cv}$ identifies the comparative rating determined by the approach and videos not recommended, $1 - rank_{cv}$, thus $1 - rank_{cv} = 0$. The minor value of *rank* is more satisfactory, defining the recommended videos have an average minor rank in testing data.

Efficiency of Demographic analysis

In this section, we evaluate the PBEM utilizing demographic analysis and universal methods. Customers in Video database are grouped into dozens of clusters between that are selected three major demographic groups.

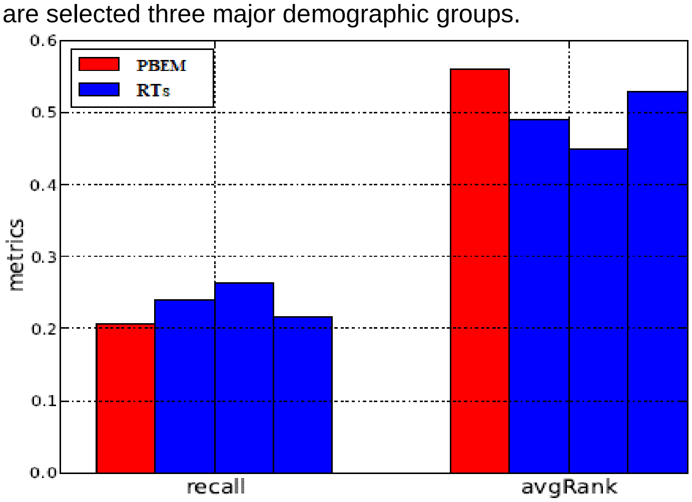


Figure 4: *Comparison of PBEM vs RTs*

The performance of PBEM and recommendation techniques (RTs) is shown in Figure 4, where the *recall* factor is shown by real ratio instead of fraction to formulate the values of *rank* (avgRank) and *recall* factor may be shown in the image in a chic way. The standard enhancement is greater than 10%, while the highest progress is equal to 20%.

Efficiency of Regulating Dynamic Technique

PBEM calculate the efficiency of online regulating dynamic techniques by using several different approaches.

Table 4: *Dynamic Techniques*

	Customers	Videos	Actions	Sparsity(%)
Technique 1	166925	6598	14990037	1.37
Technique 2	152939	7163	13102178	1.21
Technique 3	95104	5011	8480327	1.79

First of all, we describe the alternative models as comparative methods.

- Binary Approach (BnA): utilizes the binary ratings and ignores the confidence levels of customers' actions.
- Conf Approach (CfA): obtains the confidence levels as ratings and the learning rate is constant for each customer action.
- Combine Approach (CmA): takes the binary ratings to train system and obtain benifit of confidence levels of customer actions to execute an adaptable updating.

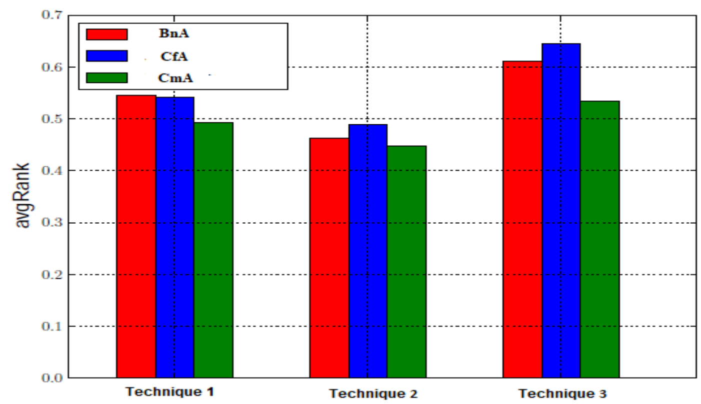
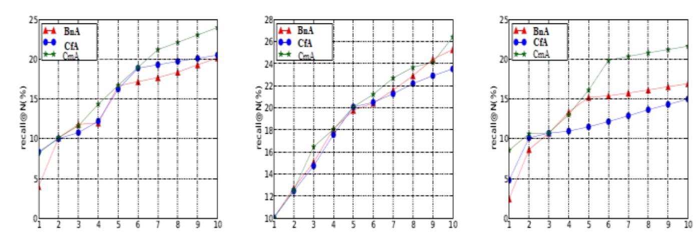


Figure5: *rank Comparison of Alternative Techniques*

The performance of above mentioned techniques is summarized in Figure 5 and obtaining $recall@R$ as factor, where R ranges from 1 to 10. As per results shown in figure 4, *Combine approach* increasingly executes superior than the other two techniques, with a regular enhancement equal to 10%.



recall@R of Binary Approach *recall@R of Conf Approach* *recall@R of Combine Approach*

7. CONCLUSION AND FUTURE DIRECTION

The PBEM presents the approach to discover the embedded user desires from the billions of videos. This work may be utilized handle the scalability and consistency issues of existing recommender systems. The PBEM approach is generic and may be utilized with any existing recommender systems. The accuracy and efficiency of the existing recommender systems executed exploiting the PBEM approach is satisfactory. In future we try to calculate the performance of PBEM approach using automated system with better metrics of recommender systems.

REFERENCES

[1] Mukamakuza Carine, Sacharidis D, Werthner H, Mining User Behavior in Social Recommender Systems, ACM, pp 11-15, 2018.

[2] Gorgoglionea Michele, Pannielloa U, Tuzhilinb Alexander, Recommendation strategies in personalization applications, Information and Managemnet-Elsevier, pp 1-12, 2019.

[3] Chun-Hua Tsai, Peter Brusilovsky, Explaining Recommendations in an Interactive Hybrid Social Recommender, ACM, pp 391-396, 2019.

[4] Jiahui Liu, Peter Dolan, Elin Rønby Pedersen, Personalized News Recommendation Based on Click Behavior, ACM, pp 1-4. 2019.

[5] Paul Covington, Jay Adams, Emre Sargin, Deep Neural Networks for YouTube Recommendations, ACM, pp 9-13, 2016.

[6] Huan Yan, Chunfeng Yang, Donghan Yu, Yong Li, Depeng Jin, Dah Ming Chiu, Multi-site User Behavior Modeling and Its Application in Video Recommendation, JOURNAL OF IEEE TRANSACTION ON KNOWLEDGE AND DATA ENGINEERING, pp 1-14, 2019.

[7] Sunny Thukral, Rana V, Versatility of Fuzzy Logic in Chronic Diseases: A Review, Journal of Medical Hypotheses, Elsevier, Vol 122, pp 150-156, 2018.

[8] Vijay Rana, Singh G, "An Analysis of Semantic Heterogeneity Issues and their Countermeasures Prevailing in Semantic Web", ICROIT 2014, IEEE Xplore, pp 16-22, 2014.

[9] Sunny Sharma, Sunita, Vijay Rana, A semantic framework for ecommerce search engine optimization, International Journal of Information Technology-Springer, pp 1-6, 2018.

[10] Siddharth Patwardhan, Satanjeev Banerjee and Ted Pedersen, SenseRelate::TargetWord - A Generalized Framework forWord Sense Disambiguation, Association fo Computational Linguistics-ACM, pp 73-76, 2005.

[11] Simon Scheider and Werner Kuhn, How to Talk to Each Other via Computers: Semantic Interoperability as Conceptual Imitation, Applications of Conceptual Spaces Volume 359 of the series Synthese Library, pp 97-122, 2015.

[12] Gorgoglione Michele, Panniello, Tuzhilin, Recommendation strategies in personalization applications, Information & Management, pp 12-18, 2019.

[13] Shanahan T, Trang, Erik, Getting to know you: Social media personalization as a means of enhancing brand loyalty and perceived quality, Journal of Retailing and Consumer Services Vol 47, pp 57-65, 2019.

[14] Resnik, P, Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*, 1995.

[15] Sunita., & Rana, V, An Effective Preprocessing Algorithm for Information Retrieval System, International Journal of Recent Technology and Engineering (IJRTE). 8(3),pp 6371-6375, 2019.