Recommendation Systems: a deep survey for new insights and directions

Sunit Kumar Jaiswal
Department of Computer Science and Engineering
BIT, Gorakhpur-273209
India
sunit007mmmec@gmail.com

Sudhir Agarwal
Department of Computer Science and Engineering
BIT, Gorakhpur-273209
India
sagarwal22@bit.ac.in

***Abstract*— We are living in the digital world where artificial intelligence(AI) and machine learning(ML) has changed the way of living and working. Recommendation Systems(RS) are one of the major applications of this domain which has impact on our day to day life. This paper is not only covering the approaches of RS but also an overview of issues ,shortcomings and future advancements possible. In the survey we found that the major approaches for recommender systems are content-based-filtering, collaborative filtering and hybrid approach. While reviewing we came to know that most of the RS are using content filtering (almost 60%) and rest are using collaborative filtering and hybrid. We came to know that several short coming in results are also there which is needed to be overcome while building any RS. For example it is very difficult to tell which approach is better than other because in some cases content filtering overshadows the collaborative filtering while in other cases reverse is true. while reviewing in this paper we are not only discussing the problems but also common framework which will be helpful in developing the better.**

Keywords*— *Recommendation Systems, Content Filtering, Collaborative Filtering, Hybrid Approach

I. Introduction

Have you ever wonder how the movie app knows in which type movie you are interested in or e-commerce app knows which type of clothes you are going to buy. The social networking sites shows the potential friends to whom you are interested in to be connected. So there are countless applications available which we use in our day to day life in the present digital world. All these applications uses recommender systems(RS) which allows us to choose the best option available.

A recommendation system[1] is a computer based technique which makes prediction based on users usage history. Many RS algorithm are proposed to get more accurate and precise suggestions. RS now also exploring of social sites and contextual information to generate dynamic recommendation. of RS, shortcomings and issues and potential areas to be addressed.

II. Recommendation Systems:

The popular recommendation/filtering approaches in Recommendation Systems are listed as follows:

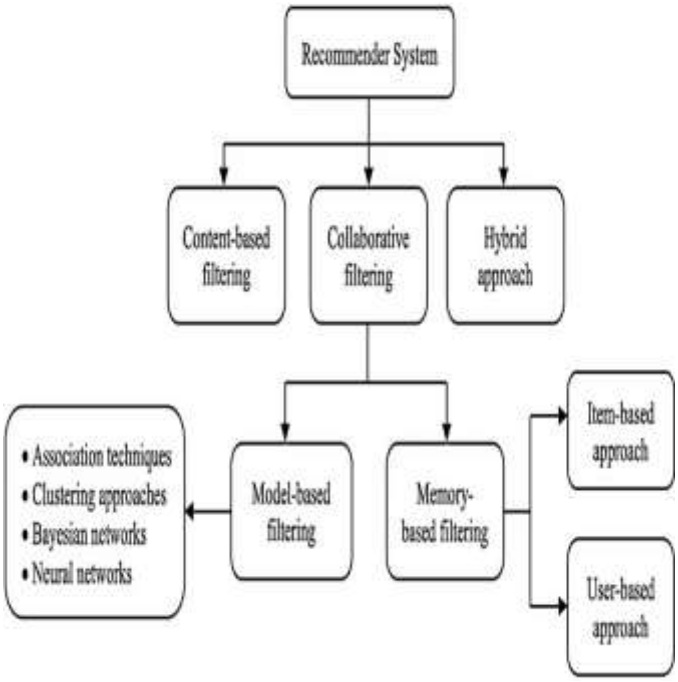

Figure 1: Different Approaches of Recommender System

A. Content Based Recommender Systems(CBRS)

In CBRS content based filtering (CBF) is utilized to recommend items with the help of user profile and item description matching. Users previous search and purchase history are the basis of recommendation. The system learns to recommend similar items that the users liked in their past.

Calculation of similarity is done considering features related with the comparable items.

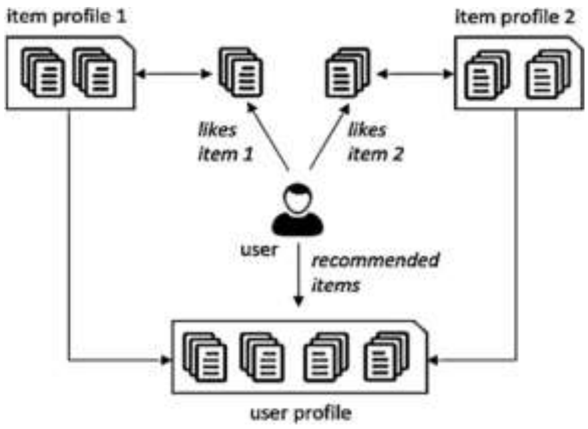

Figure 2: Content Based Recommendation Systems

B. Collaborative Filtering Recommendation Systems (CFRS)

"A man's reputation is formed by the company he keeps." CFRS is based on the idea that if two or more users have demonstrated similar interests shown in their past so the interests shown are possible to match in future as well. "Collaborative Filtering is a method of keeping note of a user's previous evaluations and ratings on things in order to suggest similar items in future. Though the user has never dealt with the specific item, if his peers have, it will be recommended to him. It is self-evident that a large number of users group must be evaluated in order to attain adequate suggestion accuracy. For a reliable suggestion, trust is critical. The following are some CFRS classifications:

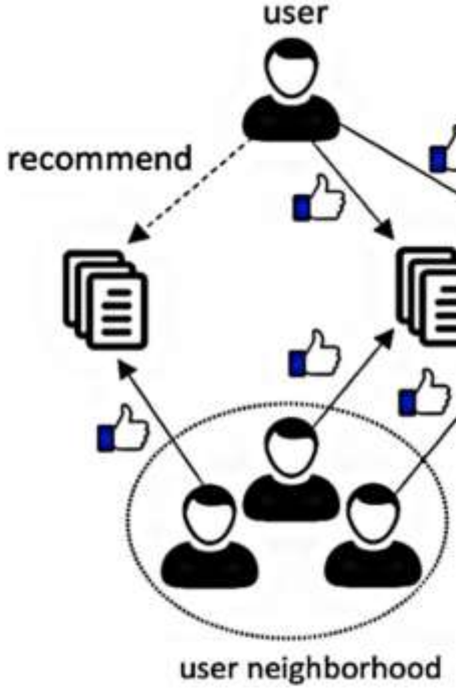

Figure 3: Collaborative Filtering Recommendation Systems

a. Memory-based collaborative recommendation systems (CRS):

Two essential steps in the memory-based CRS are the prediction computation and similarity measure Memory-based CRS are classified as follows based on the similarity computation method:

1. Item-based CRS: On the set of items similarity computation is performed.

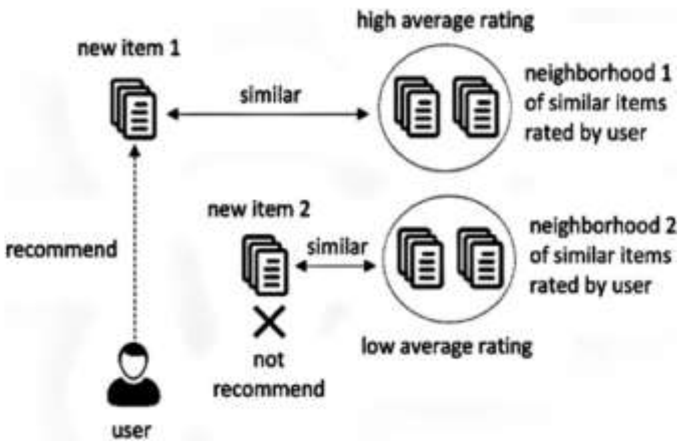

Figure 4: Item-based CRSS

2.User-based CRS: User similarity values are used to compute Similarity.

b. Model-based CRS:

In Model-based CRS machine learning algorithms can be used for implementation like as Clustering, Bayesian networks, Sparse Factor Analysis ,Markov Decision Processes, and Dimensionality Reduction .

C. Demographic Recommender Systems (DRS)

User's demographic profile used in DRS, which includes things like age, gender, education, occupation, and location. Clustering algorithms are used to categorise target customers based on demographic data. If the demographic characteristics remain constant, the same collection of things may be recommended to users. As a result, some fresh and superior recommendations may be overlooked by users.

D. Hybrid Recommendation Systems (HRS)

In Hybrid2 RS multiple filtering techniques are used. One of

popular combination in HRS is CFRS ,Hybrid approach and CBS and objective is to improve the accuracy and limitations of individual filtering.

- Hybrid Strategies for Recommendation Systems

a. Weighted

The findings of all recommendation algorithms are combined in weighted-hybrid recommendation systems, which then determine the recommended item/score. value's All recommendation methods are given identical weight at first, and then the weight is gradually adjusted if user ratings forecasts are confirmed or not. However, this approach implicitly presupposes that the relative value of individual strategies is uniform across all conceivable objects, which is not necessarily the case.

b. Switching

One recommender is chosen from among the constituents using a switching approach. For different users/profiles, a separate system may be utilised. If the CBF is unable to offer a good recommendation with high confidence, an other approach, such as the CF, is used.

c. Mixed

When a large number of recommendations are required at the same time, this form of recommendation system is necessary. In a consolidated list, the mixed hybrid technique displays recommendations from each of its components side by side. Combining numerous separate lists is a difficult task. Merging can be done in one of two ways: based on expected rating or depending on the recommender.

d. Feature Combination

Feature combination allows the complimentary features of one technique to be combined. For instance, incorporating a CF recommendation into an algorithm designed to process data using a different way (for example, CBF). Dealing with CF as an additional feature data related to each model and using CBF over this built-up dataset is how content-collaborative merging is done. These systems take into account collective data but do not totally rely on it. As a result, the system's sensitivity to the number of users who rated an item is lowered, boosting the system's accuracy.

e. Feature Augmentation

Item's rating are generated using feature augmentation approach and use those information in the upcoming suggestion technique's processing. Using the participating domain's recommendation logic, feature augmentation can build a new feature for each item. When the core recommendation component is well-developed but requires extra knowledge characteristics, this is employed. The output features of one recommender are included in the features used by the second, but in cascaded, this is not the case.

f. Cascaded

This is a method for forming a hierarchical hybrid such that a weaker technique with a lower priority cannot override the decisions taken by stronger technique or higher priority . If there is a tie in the score of higher priority and the stronger recommenders, it is broken with the lower priority recommender. The lower priority strategy is not employed on items that have already been well differentiated by the first or on goods that are poorly regarded.

- Meta-Level

This is a hybrid technique uses one output model learned by RS and then uses as an input by another. This is not like feature enhancement. All general features of the learned model are used as an input for a next one in a feature augmentation hybrid, while the full learned model is utilised as an input in a meta level hybrid. The RS cannot function with raw profile data. Using some of the provided RS pairs, a meta-level hybrid is challenging to construct.

E. Knowledge-based Recommendation system (KBRS)

Knowledge-based RS rely their recommendations on more information regarding the current user's interactions with various objects. The case-based reasoning technique, which divides the user's demand into multiple situations based on various criteria and offers recommendations that match the user's probably preferred option, is a feature of KBRSs. Constraint-based RS is another variety of KBRS that adapts to user preferences and recommends products that fit those choices. In the absence of such a product, a list of comparable alternatives to the chosen item is recommended. But the expense of KBRS is high.

F. Context-Aware Recommendation Systems (CARS)

A context can be defined by a number of factors such as time, location, partner, mood, and so on. A user's demographic traits do not change over time, however contextual information varies as the user's surroundings change. In CARS, mobile applications that collect contextual data are critical. By capturing the user's emotional context, better recommendations can be made.

III. Techniques For Information Retrieval And Processing:

There is a lot of data in the digital world, and it is boundless. People's interactive participation may magnify the situation. The RS must analyse all possible areas to extract informative data in order to understand users' preferences and tastes in order to make an efficient and profitable recommendation. Every RS uses information retrieval and processing strategies to complete this task. The following are the most commonly utilised information retrieval and processing techniques in RS.

Machine learning[16]

Machine learning gives machines the ability to learn on their own, without having to be explicitly programmed. Many techniques are available, including logistic regression, association rule learning decision trees, clustering, support vector machines and Bayesian networks, among others.

a. Logistic regression

Using both continuous and discrete data, logistic regression[3] is used to predict discrete variables.

b. Decision tree

The decision tree is a strong tool for selecting an option from a list of possibilities.

c. Association rule learning

For recommendations, association rule learning is utilised to get common pattern, relationship, correlation, or causal structure from user and item datasets.

d. Cluster analysis

Cluster analysis, which is unsupervised learning, is used in RS to organise a large number of items based on pattern ,similarity and structure.

e. Bayesian network

In large networks such as social networks, probabilistic model Bayesian network classifier is used to address problems related to classification.

f. SVM

Support vector machines, which are supervised learning algorithms, are used to analyse data combining associated learning method with regression and classification (linear and nonlinear).

g. LDA

Topic modelling is the process of extracting a common theme from a collection of texts. A topic is identified in a document using a different combination of words. In RS, topic modelling is done using LDA (a probabilistic model of a corpus).

h. Deep learning

This is critical for uncovering patterns which are hidden in data and has unleashed in a new era of research in data mining. In RS, it can be utilised to create effective and dynamic behaviour modelling.

IV. Problems and challenges associated with RS:

Due to changing objectives of Organization's for using and deploying RS, it is difficult to measure their performance. User's happiness is the most telling indicator in general. Though a heuristic method cannot be utilised to measure happiness of user, still evaluation of RS is based on how well they can solve common interests. Here explanations are given for the metrics used to test the performance of RS and various major difficulties for example cold-start, accuracy, scalability diversity and data sparsity.

a. Limited content analysis

Accuracy of RS is determined by the input provided by the user. Lower information equals lower performance; larger information equals higher RS performance. The limited content analysis problem is the name for this issue. The whole domain information is required to create an appropriate recommendation.

b. Over-specialisation

The goal of RS is to help people to find out new items. The good RS must diversity at many aspects. Though, it is possible that recommendation systems are giving the opposite result. The good RS recommends popular and highly rated goods that a user may enjoy. As RS not proposes goods from a non-homogeneous set of items, accuracy may suffer. To address this issue, new hybrid ways must be developed that will improve the efficiency of the suggestion process. Based on the user's previous behaviour, the learning methods used strive to discover the most relevant documents. In this method the user's are restricted to documents that are similar which are already explored. It is referred to as the over-specialization issue.

c. Cold start

Whenever a new user or item is introduced to RS, the system has no previous data on which to make a recommendation (ratings, preferences, search history, etc.). This problem is known as a "cold start." It's also known as the new user/item problem. To address this issue, it is possible to leverage the user's profile and the demographic data of the user. This method, however, falls short since consumers with comparable demographic traits could have varying degrees of interest in the same product.

d. Scalability

The RSs' complexity4 increases when they work on massive datasets with millions of individual things set and a large number of users. The systems must be able to respond quickly to online requests and provide suggestions to all users based on their purchase and rating histories, necessitating the usage of scalable solutions.

e. Synonymy

The problem of several terms with similar meanings is referred to as synonymy. The majority of RSs are unable to locate identical or comparable things with different names (synonyms). As a result of this inability, a number of issues arise. For example, while the terms 'children film' and 'children movie' essentially mean the same, memory-based CF system[6,7,8,9]would not be able to compute similarity between them.

f. Abbreviation

The RS unfamiliar with the terminologies of users regularly use throughout online discussions, may not recognise the thing that the user is seeking for. As a result, incorrect recommendation is generated. Here main idea is to keep both lists intact and group the reduced words with their full forms.

g. Long tail

Those items in an RS originally rated negatively or not at all that follows a top-N suggestion will be eventually removed from the catalogue. This issue is inextricably linked to diversity. It emphasises the importance of recommending a variety of things to consumers, as well as how varied the items are from one another. However, RSs do not comply with this feature, resulting in the long tail problem. Because he did not rate or did not have access to many important goods, a user may miss suggestions for many of them. This frequently results in long tail problem (LT) occurring while a large number of goods remain unrated or have low ratings.

h. Black-box problem

The RS's efficiency is improved by increasing the clarity of recommendation. The user's pleasure with the advice is linked to the level of faith he or she has in the recommendation's goals. When a system is opaque to the user, the black boxproblem occurs, resulting in a loss of trust in the system As a result of black box issues, which cause consumers to lose faith in the system, the potential of recommendation is hindered, resulting in an unsuccessful suggestion attempt. The problem can be overcome by information of user about the factors used to make recommendations, such as other people' interest behaviors, the profile created through their own interests, and so on. This encourages users to purchase the item because their choices are impacted to some extent by the choices of their peers who purchased the same item. People are compelled to place their faith and trust in the recommendations when transparent reasons are presented to them, for example, similar users' evaluations on an item . Netflix, for example, makes movie suggestions based on massive background algorithms, but it is made sure that the recommendations are always interesting to the people by disclosing the foundation of their recommendations.

5. Conclusion

Making a choice among many options and based on a vast amount of web info is typically a challenging and complicated task. By using online RS, we can get beyond this. RSs employ efficient information retrieval and filtering methods to carry out their duties skillfully and precisely. To accomplish these objectives, a significant amount of research has been conducted recently, and numerous methodology and technique recommendations have been made.This paper describes the many recommendation strategies used in RS, such as collaborative, hybrid, content-based, knowledge-based, demographic, and context-aware recommendation. Cold start[9], limited content analysis, sparsity, over-specialisation, scalability, abbreviation, synonymy, black box and long tail issues are all explored briefly in the design and implementation of RS systems. Information retrieval techniques such as Bayesian network classifier, support vector machine, cluster analysis, logistic regression, decision trees, association rule learning, TF-IDF, machine learning, LDA [10], and deep learning are briefly reviewed.. This paper's main goal and purpose is to track down the latest RS research developments. There have been some fascinating statistics discovered. The majority of research in RS, for example, is focused on CF and knowledge-based approaches. China is the leading contributor to RS. IEEE publishes the vast bulk of the articles. It has also been observed that RS research peaked in the years 2013-2014. Following that, the popularity of study in this topic has gradually diminished, most likely due to saturation. However, we feel that RS research is not yet dead. AI[11,12,13] cognitive computing[14] and IoT, technologies have given a new lease on life. We believe that in the near future, RS research will take on various new and interesting directions.

REFERENCES

[1] Shaina Raza et al. News recommender system: a review of recent progress, challenges, and opportunities
https://link.springer.com/article/10.1007/s10462-021-10043-x(2021)

[2] Deepjyoti Roy & Mala Dutta.A systematic review and research perspective on recommender systems
https://rdcu.be/cWSC6

[3] Gozes, O. et al. Rapid AI development cycle for the coronavirus (COVID-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv e-prints 2003, arXiv:2003.05037 (2020).

[4] Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(20)30120-1 (2020).

[5] Wang, S. et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). Springer, https://link.springer.com/article/10.1007/s00330-021-07715-1 (2021).

[6] Song, Y. et al. Deep learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT images. IEEE, https://ieeexplore.ieee.org/document/9376253 (2021).

[7] Jin, C. et al. Development and evaluation of an AI system for COVID-19 diagnosis.
medRxiv, https://doi.org/10.1101/2020.03.20.20039834 (2020).

[8] Punn, N. S. & Agarwal, S. Automated diagnosis of COVID-19 with limited pos- teroanterior chest X-ray images using fine-tuned deep neuralnetworks.Springer https://link.springer.com/article/10.1007/s10489-020-01900-3(2021)

[9] Tostmann, A. et al. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Neth- erlands, March 2020. Eurosurveillance 25, 2000508 (2020).

[10] Feng, C. et al. A novel triage tool of artificial intelligence assisted diagnosis aid Mei, X. et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. Nat. Med. 26, 1224–1228 (2020).

[11] Jin, J.-M. et al. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. Front. Public Health 8 (2020).

[12] BMJ GH Blogs. Sex, gender and COVID-19: Disaggregated data and health dis- parities. BMJ Global Health blog https://blogs.bmj.com/bmjgh/2020/03/24/sex- gender-and-covid-19-disaggregated-data-and-health-disparities/ (2020).

[13] Whittington, A. M. et al. Coronavirus: rolling out community testing for COVID-19 in the NHS. BMJ Opinion https://blogs.bmj.com/bmj/2020/02/17/coronavirus- rolling-out-community-testing-for-covid-19-in-the-nhs/ (2020).

[14] Menni, C. et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat. Med. 26, 1037–1040 (2020).

[15] Hastie, T., Tibshirani, R. & Friedman, J. In The Elements of Statistical Learning: Data Mining, Inference, and Prediction (eds. Hastie, T., Tibshirani, R. & Friedman, J.) 337–387 (Springer, 2009).

[16] Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. Do we need hun- dreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181 (2014).

[17] Omar, K. B. A. XGBoost and LGBM for Porto Seguro's Kaggle challenge: A compar- ison Semester Project (ETH Zurich, 2018).

[18] Josse, J., Prost, N., Scornet, E. & Varoquaux, G. On the consistency of supervised learning with missing values. arXiv:1902.06931 [cs, math, stat] (2019).

[19] Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794 (Association for Computing Machinery, 2016).

[20] Cong Feng et. al. A novel artificial intelligence-assisted triage tool to aid in the diagnosis of suspected COVID-19 pneumonia cases in fever clinics
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7940949/(2021)

[21] ACRnews (2017) [online] [online] http://www.acr-news.com/ (accessed 6 April 2017).

[22] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M. and Ayyash, M. (2015) 'Internet of things: a survey on enabling technologies, protocols, and applications', IEEE Communications Surveys & Tutorials, Vol. 17, No. 4, pp.2347–2376.

[23] Badrul, S., George, K., Joseph, K. and John, R. (2001) 'Item-based collaborative filtering recommendation algorithms', in 10th International Conference on World Wide Web.

[24] Celma, Ò. and Serra, X. (2008) 'FOAFing the music: bridging the semantic gap in music recommendation', Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 6, No. 4, pp.927–934.

[25] Cobos, C., Rodriguez, O., Rivera, J., Betancourt, J., Mendoza, M., León, E. and Herrera-Viedma, E. (2013) 'A hybrid system of pedagogical pattern recommendations based on singular value decomposition and variable data attributes', Information Processing and Management: An International Journal, Vol. 49, No. 3, pp.607–625.

[26] Fesenmaier, D.R., Ricci, F., Schaumlechner, E., Wöber, K. and Zanella, C. (2003) 'DIETORECS: travel advisory for multiple decision styles', in Proceedings of the International Conference on Information and Communication Technologies in Tourism, Wien, Austria.

[27] Hayes, C. and Cunningham, P. (2001) 'Smart radio – community based music radio', Knowledge Based Systems, Vol. 14, Nos. 3–4, pp.197–201.

[28] Hotho, A., Jäschke, R., Schmitz, C. and Stumme, G. (2006) 'Information retrieval in folksonomies: search and ranking', in The Semantic Web: Research and Applications: 3rd European Semantic Web Conference, Budva, Montenegro

[29] Hu, Y., Yang, Y. and Huang, B. (2015) 'A comprehensive survey of recommendation system based on taxi GPS trajectory', in International Conference on Service Science.

[30] Khan, M.M., Ibrahim, R. and Ghani, I. (2017) 'Cross domain recommender systems: a systematic literature review', ACM Computing Surveys (CSUR), Vol. 50, No. 3, pp.36:1–36:34.

[31] Knees, P. and Schedl, M. (2013) 'A survey of music similarity and recommendation from music context data', ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 10, No. 1, pp.2:1–2:21.

[32] Lawrence, R., Almasi, G., Kotlyar, V., Viveros, M. and Duri, S. (2001) 'Personalization of supermarket product recommendations', Data Mining and Knowledge Discovery, Vol. 5, No. 1, pp.11–32.

[33] Lee, S.K., Cho, Y.H. and Kim, S.H. (2010) 'Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations', Information Sciences, Vol. 180, No. 11, pp.2142–2155.

[34] Lu, J., Shambour, Q., Xu, Y., Lin, Q. and Zhang, G. (2010) 'BizSeeker: a hybrid semantic recommendation system for personalized government-to-business e-services', Internet Research, Vol. 20, No. 3, pp.342–365.

[35] Mayeku, B., Edelev, S., Prasad, S., Karnal, H. and Hogrefe, D. (2015) 'PECALE: an environment for enhancing personalization and learner engagement in an online learning platform', in Proceedings of the 2015 IEEE 15th International Conference on Advanced Learning Technologies.

[36] Meymandpour, R. and Davis, J. (2015) 'Enhancing recommender systems using linked open data-based semantic analysis of items', in 3rd Australasian Web Conference, Sydney, Australia.

[37] Miller, B.N., Konstan, J.A. and Riedl, J. (2004) 'PocketLens: toward a personal recommender system', ACM Transactions on Information Systems, Vol. 22, No. 3, pp.437–476.

[38] Moghaddam, M.G., Mustapha, N., Mustapha, A., Sharef, N.M. and Elahian, A. (2014) 'AgeTrust: a new temporal trust-based collaborative filtering approach', in International Conference on Information Science Applications (ICISA).

[39] Pereira, N. and Varma, S.L. (2019) 'Financial planning recommendation system using content-based collaborative and demographic filtering', in Smart Innovations in Communication and Computational Sciences, pp.141–151.

[40] Porcel, C., Herrera-Viedma, E. and Moreno, J.M. (2009a) 'A multi-disciplinar recommender system to advice research resources in university digital libraries', Expert Systems with Applications, Vol. 36, No. 10, pp.12520–12528.