International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Target User Specific Q-Learning (TUQL) Personalized Product Recommendation

R.V. Karthik*[1], Sannasi Ganapathy[2]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract: Recommendation system plays important role to predict the relevant product from large number of available products. The conventional recommendation systems focus on the item user interaction and predict the products based on the similar user interest or item purchased history. Later recommendation is further enhanced by context aware and sequential based recommendation; which predicts based on current search and browse session information respectively. In the recommendation; current important challenge is addressing both accuracy and Serendipity; predicting the interested unknown products. In this paper, we propose hybrid recommendation framework to overcome this challenge, target user specific Q learning reinforcement (TUQL) approach, predicts Top N recommendation effectively based on the current context, user past purchase behavior, temporal data and consider real target end user. The experimental results of the proposed recommendation system show better performance than the existing product recommendation systems in terms of prediction accuracy on relevant products for the target users and lesser computation time.

Keywords: *Product recommendation, recurrent neural network, sentiment analysis*

1. Introduction

Social networking and ecommerce are very popular now due to high penetration of internet. Especially the growth of smart mobile phones increased the online usage. Online ecommerce playing significant role in today's commodity business helps lot of small vendors extend their business by collaborating with online ecommerce organization. Social network applications like Twitter, LinkedIn, YouTube, micro blogs, online review sites makes easy to connect and exchange the information. With the huge volume of products in the online market, recommendation system plays important role in reducing the dimension and provide the personalized recommendation to the users. This improves the customer satisfaction and revenue which are key factors for successful business.

In general, recommendation systems are classified into two basic approaches: collaborative filtering (CF) and content-based (CB) approaches. In collaborative filtering approaches, recommendation is done with the user item matrix methodology by considering the similar product purchased users and items. Thanks to latest technologies AI and Machine learning which further helps to improve the recommendation system by applying hybrid approaches and machine learning techniques like Neural network. It helps to addresses the important challenges in the recommendation system like cold start problem etc. [12].

Artificial intelligence (AI) is the field of development of machines with intelligence that work and reacts like humans. AI is now widely used for Vision recognition, speech recognition, automatic message conversation (ChatBot), decision taking scenarios in the field of banking, automobile, finance, ecommerce.

Recommendation system [1] should avoid the false positive characteristic; products are recommended but those are not like buy the customers. Recommendation should consider multi factors for quality personalized recommendation like past purchase product time, purchase interval, any influence on the session, and relationship between the purchased products. Recommendation system should consider two aspects [2] for the recommendation; one is static past purchase behaviour information and other one is dynamic prediction of user current context by observing the session or products interacted etc.

Reinforcement learning plays significant role in the second dynamic aspect, which continuously check the current context or environment and dynamically predicts the recommendation and learn on its own from the recommendation outcome. Reinforcement learning comes under machine learning; it learns on its own within certain environment by continuously interacting and tends to maximize its reward by acting based on the action feedback. Advantage of the reinforcement learning is no need train any model or not require huge data for prediction. It will predict the product and will observe the response for the prediction in the environment, customer interest. Approach in taking decision making for prediction is said to be policy. If user shows interest in the recommended list; it takes as reward and adjust the prediction decision or policy accordingly. If the product is not liked; then consider as punishment and adjust the policy for the next recommendation. Short messages generated from social networking or real time websites like twitter are also used as the knowledge source for the recommendation system.

Recommendation should consider both long term preferences and short-term individual customer preferences or contexts for the personalized recommendation. Reinforcement learning is applied various application in financial Forex trading [12].

[1] *School of Computer science, Vellore institute of technology, Chennai – 600127, India*
ORCID ID: 0000-0003-1579-5750
[2] *Centre for Cyber Physical Systems, School of Computer Science and Engineering, Vellore institute of technology, Chennai – 600127, India*
ORCID ID: 0000-0001-9177-5378
* *Corresponding Author Email: karthikr.v2016@vitstudent.ac.in*

In the recommendation system, for the quality recommendation it is require considering all the important influence factors like context, temporal, Sequence, and demographic attributes [17-22]. In the proposed recommendation system, all the influence factors are considered to improve the recommendation; it contains below three important modules:

- Products are recommended based on the target user – to whom exactly user want to purchase the product. Customer reviews are processed here to get the real target user and rating is done based on opinion or sentiment in the review.
- Context and temporal information are extracted from the past purchase history. This reduces the dimension from the wide range of products and improves the precision of the recommendation.
- Reinforcement learning – Q approach policy is developed to decide the recommendation and continuously learn based on the interaction and response received for each recommendation action.

This research article is described with the below sections. Related research articles in recommendation system and reinforcement learning are described in Section 2. Section 3 describes the Scope of improvement and motivation to the proposed target user specific reinforcement learning approach. System architecture and the related modules are explained in Section 4. Section 5 explains the complete proposed work with all the necessary modules and algorithm. Results and evaluation are described in the Section 6 with standard metrics. Section 7 describes the conclusion and future work.

2. Literature Survey

2.1. Personalized Recommendation System

Fatima Rodrigues et al [1] proposed two step approaches; clustering the similar customer by considering the additional attributes purchase time and then derived the relationship between the purchased products for better recommendation. Considering the time of purchase attributes helped to reduce the dimension of large items and improved the quality of the recommendation. Sydney Chinchanachokchai et al [7] research outcome shows that more widely used Collaborative filtering recommendation system shows preferrable approach for expert user community. Ming-Chuan Chiu et al [8] used unsupervised natural language processing for personalized recommendation in a smart product service system. In this approach, both user data and the current context generated from the system are used.

Sang Hyun Choi et al [9] proposed the model to do personalized product recommendation by considering the multiple features even when there are minimal customer preference inputs. Sandra Garcia Esparz et al [11] used the real time websites like twitter messages for product recommendation by indexing and retrieving the information from it. Chin-Hui Lai [13] proposed product recommendation method in the social networking websites. Method used the interaction based on the product rating given by the user and review comments. Shan Liu et al [16] developed the deep inverse reinforcement algorithm (IRL) to predict the preferred route for food delivery person based on the historical GPS data.

2.2. Reinforcement Learning recommendation System

Elena Krasheninnikova et al [5] used Markov decision process (MDP) reinforcement approach for predicting the insurance renewal price for the customer. It determines policy based on the outcome of proposed amount in sequence decision approach. Stefano Ferretti et al [10] applied reinforcement learning technique to automatically adapt the web contents based on the user profile.

Dezhuang Miaoa [2] proposed humming query reinforcement learning approach recommendation system for Music recommendation. It is integrated framework of using the static past information and predicts the current interest of the user by learning with real time continuous interactions. Reinforcement learning helps here to identify the short-term preference and improves the quality of the recommendation. Michael Reisener et al [3] used reinforcement learning and Machine learning techniques for predicting the product portfolio management based on the company goals by align with the company goal and driving parameters. This is data-based decision methodology which helps to achieve the company long term goals.

Shenggong Ji et al [6] proposed the novel reinforcement learning approach for dynamic taxi route recommendation to enable the vacant taxis to find the best route to pick the passengers. Reinforcement learning are applied by utilizing the real time temporal features. Feng Liu et al [15] used reinforcement learning to resolve the Top aware issue by applying the supervised learning to address the short term and long-term user preferences.

Gang Ke et al [12] proposed dynamic recommendation system for cross platform using reinforcement technique by considering the variable interest of user. Proposed method also improved the accuracy of the prediction. Karthik et al [21] proposed the Feature based recommendation system by rating the product based on the interested feature and used variable rating approach instead of common rating scale for each product. The product score is calculated by applying sentiment analysis on each customer review for the respective product. Karthik et al [23] proposed multi demographic recommendation approach for predicting the interested product by considering the features, demographic information, and the target user context. Demographic information such as location helps to reduce the dimension from the huge product list. Karthik et al [22] proposed the Fuzzy recommendation system by considering the target user context and temporal information. It predicts the product with different buckets highly recommended product, likely to recommend products and not recommended products.

3. Scope of Improvement

There are lot of hybrid recommendation system are developed but could see the scope of improvement in the below aspect. Especially considering the multi factors: temporal and context influence and should dynamically adjust the recommendation based on the change in user interest. User interest various over time, so it is important to change the recommendation based on the current context and interest. Could see the improvement required in the recommendation system with the below aspects.

- **Target user specific recommendation**: All the recommendation systems are considering the user past purchase details or products are predicted based on the overall rating. But there are chances, user purchasing the product for others, friend, or family members. So, it is required to this factor and recommendation system must predict and recommended the products by considering this target end user, Product to whom intended to buy.
- **Quick adapt and discover new interest with less**

cost: For every search; recommendation system requires to adapt the new context or user interest and need to adjust the recommendation list accordingly with less computation cost. It should not predict very new set of products every time; this will lose the user credibility. Meanwhile recommending the same list again could also bore the user. It is requiring the balanced one to predict the recommendation with user interest and adapt the recommendation list such way that it makes user interest. Reinforcement learning is very good approach here which helps to adapt and discover new based on the current context.

- **Influence of context and temporal**: Context and temporal details plays significant role in the recommendation system. From the past purchase details, recommendation system must learn any temporal details available, where user purchasing the product with any timeline and sequence. For e.g., user purchase baby soap for every month in last three months, would also interest to buy the soap in upcoming months.

4. System Architecture

To dynamically adjust the recommendation list with respect to new user context or interest, reinforcement learning approach is used in the proposed framework. System interacts with the user interfaces. It gets the context information from the user or customer. Recommendation systems are built to understand the current context and determine the interest products based on context and temporal details. Past purchase and wish list details are used to determine any temporal influence and any past user interest and purchase behavior. Recommendation system contains various modules: i) User interface module: which accepts the user inputs for the recommendation prediction and will display the Top recommendation to the user. ii) Target user specific information are obtained from the customer review and stored in the ontological database. iii) Product rating is calculated for each product by applying sentiment analysis technique in the customer review. iv) Get temporal information available from the past purchase or wish list details. v) Finally reinforcement learning module predicts the Top N recommendation from the available candidate list and provide to the user. it will also check the response of the user for the provided recommendation. Based on the feedback, the action takes as reward or punished. If the user clicks any of the recommended product and adding it to wish list or adding it to the cart would be reward for the recommendation. If it is not referred and further explored, then it is consider as punished. Based on this output, proposed recommendation system adapts and discovers new things based on the user interest and current new context. Figure 1 shows the system architecture of proposed approach.

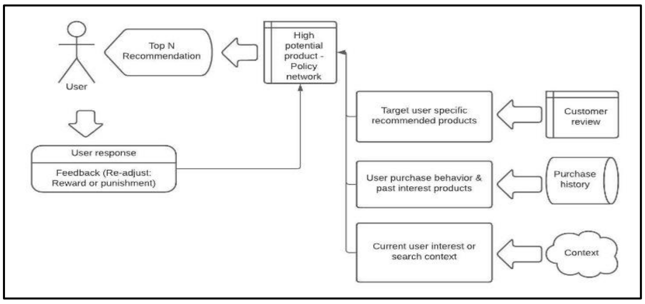

Figure 1. Proposed approach system architecture

5. Proposed work

Target user specific product rating and the dynamic adjustment are the key significant improvements focused on the proposed Target user and reinforcement learning (TURL) approach. The proposed work contains three modules: Target user specific rating determines context & temporal influence and obtains the Top N recommendation.

5.1. Product rating

Users are sharing the feedback or opinion about the product in the reviews. In most of the reviews, users mention the target user to whom it is intended to purchase. For e.g.: In the review it is given as "I have bought this product for my 8 months kid; he is happy and playing with this toy". Here target user is 8 months baby. This information helps for the recommendation system to recommend the product when other users looking for some toy for 8 months baby. Most of the recommendation system takes the overall rating of the product for the recommendation. This has some challenge as the product may not be liked by different user category. Target user specific rating would be the effective way of calculation and helps when predict based on the context. The recommendation would be more precise and personalized.

Table 1. Algorithm1: TUR approach algorithm

Target user specific product rating by TUR approach algorithm	
Input	Product meta data {p1, p2… pn} User reviews {u1,u2…..un} Rating {r1,r2…..rn}
Output	Obtain the rating of the product with each category
Step 1	For each product, do the preprocessing for each review
Step 2	Get the respective target user category from each review comments using Target user-based Sentiment computation algorithm (TUSSCA)
Step 3	Store the computed the Sentiment score and associated target users details in the Ontological database $pri_u = \sum_{i=0}^{n} (sc_i) \,/\, Overall\ rating\ of\ p_i$
Step 4	Store the computed product rating for each target user

Sentiment analysis technique is used to calculate the sentiment score for each review. Product rating is calculated for each target user. One product has variable rating based on the user category. Table 1 shows the outcome of product rating from the TUR approach. Each product has different rating for each user category. For example, Remote car toy has high rating for boys, whereas same product has low rating for girl user category.

Table 2. Product rating for each user category

Product	Target users rating		
	Boys	Girls	Kids
P_1	3	4	2
P_2	3	4	2
P_3	3	4	3

5.2. Influence of Context and temporal information

Considering the past purchase details are that much important as like understating the current context. It is must for recommendation system to utilize the past purchase information as much as possible for upcoming prediction. It helps to determine the potential target user and temporal information.

Temporal influence case 1: User purchasing the product at regular interval. For eg: user purchasing the wash detergent for every 2 months. This gives the prediction timeline close to 2 months after the last purchase; user most likely to purchase the product.

$p_{(t+1)} = \mathrm{p}_{t1}, p_{t2}..$

Temporal influence case 2: User purchasing the relevant product with some interval. For eg: User purchase "NB" baby diaper from the brand B and after 2 months purchased "S" size baby diaper. This gives the prediction that within next 3-month timeline, user most likely interest to purchase "M" size baby diaper. Here the product gets vary with the timeline. Context information also observed from the past purchase details. From the above examples: target user kid is noted to do the personalized recommendation. User has more chances to buy the product with the past purchase context.

Table 3. Algorithm 2: Context and detection

Context and temporal detection algorithm	
Input	Purchase Product meta data {p1, p2… pn} Time of purchase {t1,t2…..tn}
Output	Obtain the target user context and temporal preference of the user
Step 1	For each product
Step 2	Get the respective target user category from the ontological database
Step 3	Check the similarity score $SimScore_u = \sum_{i=0}^{n} Cos(prodi\ \&\ prod(i+1))$
Step 4	Store the relevant product in the list
Step 5	Check the temporal data $prod_{ti} = Diff(pt_i - pt_{i+1)}$
Step 6	Store the product and timeline obtained from the temporal details

5.3. Reinforcement learning

Recommendation system can predict the products based on the past purchase behavior, but it is hard to meet when user interest varies. We can't expect the same purchase behavior for every time user purchase. So, it is must for recommendation system to dynamically understand the current interest and predict the recommendation. Reinforcement learning plays vital role in the current proposed work, where it will learn the user behavior dynamically and adjust the recommendation candidate list accordingly.

Markov decision process (MDP) is typical reinforcement learning technique where ideal scenario to apply in continuous state and actions [4]. It is represented with the tuple (S, A, r, p) where S stands for the current state, A represents the action taken in the environment, r gives the reward or punishment, and p represents the probability of changing the current state S1 to another state S2. Goal of this approach to determine the policy which determine the action for each respective state with maximize the return reward RR. MDP is obtained by the Eq. (1).

$$RB(\pi) = \sum_{i=0}^{n} r^i \ \mu^i \tag{1}$$

Where, r^i represents the intermediate reward received in each state and μ^i gives the respective discount.

In the online recommendation system, Agent represents the recommendation framework or approach. Environment will be online product purchase search or shopping behavior, whereas state would be current context or new search are states in the reinforcement learning. Recommending the top N recommendation list is the action and decision approach used to predict the top N recommendation is said to be policy. User refers the recommended product, add to the wish list or cart is considered as reward and if user not referring the recommendation list considered as punishment.

Q learning approach is another more widely used reinforcement learning technique in the recommendation system. This approach computes action-value stream. Q learning reinforcement learning is represented with the tuple ($\mathrm{S_c}$, $\mathrm{A_c}$, $\mathrm{S_n}$ $\mathrm{r_n}$) Q- function, where $\mathrm{S_c}$, represents the current state, $\mathrm{A_c}$ describes the action applied for the current, $\mathrm{S_n}$ is the next state it moves when current action is applied, $\mathrm{r_n}$ is the immediate reward received. Q-learning approach is computed by the below Eq. (2).

$$Q^{\pi} < -Q^{\pi}(S_C, A_c) + \alpha(r_n + \mu\, max\, Q^{\pi}(S_{C+1}, A) - Q^{\pi}(S_c, A_c)) \tag{2}$$

Where μ is the discount rate and α is the learning rate.

5.4. Mapping product prediction with reinforcement learning

In the proposed work Q-learning approach is used called Target user specific reinforcement learning (TURL) which considers both the current context and the past purchase details to come up effective decision of predicting the candidate recommendation list. In the proposed framework considered 5 different cluster product groups for personalized recommendation evaluation. Initially the Q learning map contains the setup of products from the different clusters. Q function map is developed with the possible state as rows and the possible action in the column. When the policy predicts the initial recommendation, it act to current state S1in the environment and it evolves and moves to the next state S2. Based on the action feedback; user interested or not, policy calculates the immediate reward and adjust the recommendation policy to predict the next set of recommendation list. Table 2 shows the different cluster formed in the proposed work.

Table 2. Cluster categorization for personalized recommendation

Cluster	Description	Representation
Cluster 1	Product referred from the current context (Current user interest)	C1 $= \{c_1p_1, c_1p_2 \ldots.\}$
Cluster 2	Products referred from the past purchase history (Relevant products)	C2 $= \{c_2p_1, c_2p_2 \ldots.\}$
Cluster 3	Products referred from the temporal information in the past purchase or search history (Repeated purchased products)	C3 $= \{c_3p_1, c_3p_2 \ldots.\}$
Cluster 4	Products referred from the sequence of session or from the past purchase (Sequence)	C4 $= \{c_4p_1, c_4p_2 \ldots.\}$
Cluster 5	Trending relevant products in the current timeline (Most selling product at current time t)	C5 $= \{c_5p_1, c_5p_2 \ldots.\}$

In the proposed approach, Policy randomly takes the product from each cluster randomly which secured top rank and wait for the outcome. If the user interested and interacted with products, then policy will reward the cluster where the interested products belong to. This makes decision to consider the products more from those clusters and gives more reward. In case in sub sequent interaction, if the user is not interested; then it is punished, and value is reduced accordingly. Table 3 shows the Q learning policy value item matrix which helps to manage and take decision for product recommendation.

Table 3. Value item matrix for each cluster

Action-state	**Cluster Reward points**				
	C_1R_n	C_2R_n	C_3R_n	C_4R_n	C_5R_n
$\boldsymbol{A_1S_1}$	0	5	10	2	20
$\boldsymbol{A_2S_2}$	5	5	10	3	20
$\boldsymbol{A_3S_3}$	10	10	10	3	25

In the proposed work, there is possibility of below two scenarios for each action in the state. Scenario 1: User shows interest in the one or more recommended products. Let P_i be the product is interested; either user browse the product description or add to cart or add to the wish list. Let C_j is the cluster from the 5 cluster which contains the productP_i. Then add the reward to C_jR_n and give more priority for this cluster when compared to other cluster for next recommendation predication.

Scenario 2: User not shown any interest in the recommendation. Policy will punish the reward point and change the ratio of the products taken from the Cluster. Thus on each iteration it learns and tries to maximize the reward to get the optimal policy.

Table 4. Algorithm 3 : Q learning approach

Q-learning approach	
Input	Product from each cluster {C1, C2… C5} with products ordered with rank Reward value Q function Q(s,a)
Output	Policy π - Predicting recommended products with the given context
Step 1	Set reward r=0, state s=0
Step 2	Take the initial action randomly (s_1, a_1) θ ← random initialization
Step 3	Observe the outcome of the action in the environment $SimScore_u = \sum_{i=0}^{n} Cos(prodi \,\&\, prod(i+1))$
Step 4	Obtain the reward r_i
Step 5	Determine the cluster of the interested product and add the reward $Q^\pi < -Q^\pi(S_1, a_1) + \alpha\big(r_i + \mu max Q^\pi(S_2, a) - Q^\pi(S_1, a_1)\big)$
Step 6	Repeat step 2 to Step 4 for each state – update the policy π
Step 7	return the derived policy π

Q learning approach is used to pick the products from different cluster and display the recommendation to the user. Based on the feedback policy dynamically adjusts the ratio and value item matrix and predict the next recommendation. This is the continuous "action state" approach of providing the recommendation as action and check the response of the users as State.

Finally, the overall TUL reinforcement approach algorithm is explained below. Algorithm 1 is used for extracting target user specific category and variable rating. Algorithm 2 is used for extracting the temporal information and user long term preferences from past purchase history. Algorithm 3 is applied by getting the products from the clusters with the current context. SoftMax technique is used to predict the Top N recommendation. In the smart phone and tablets Top 10 recommendation is widely used. So SoftMax is used to pick the highly interested and top-rated products from the candidate recommendation list.

Table 5. Algorithm 4 : Target user specific reinforcement

Target User specific Q-Learning reinforcement	
Input	Product meta data {p1, p2… pn}, User reviews {u1,u2…..un}, User context search information {CS: cs1, cs2,cs3…} User demographic
Output	Predicting Top N recommendation with the given context
Step 1	For each product user review, use algorithm 1 Target user-based Sentiment computation algorithm (TUSSCA) and compute the Sentiment score and associated target users details in the Ontological database
Step 2	Apply algorithm 2 for obtain the sequence and temporal information from the past purchase history {P1,P2…Pn} -> {p1,t1} {(P1,P2),t2}
Step 3	Sort the obtain products based on rank and categorize into each cluster
Step 4	Apply algorithm 3; determine the reinforcement policy and predict the recommendation $Q^{\pi} < -Q^{\pi}(S_1, a_1) + \alpha\big(r_i + \mu max Q^{\pi}(S_2, a) - Q^{\pi}(S_1, a_1)\big)$
Step 5	Apply SoftMax to display Top N recommendation to the user from the candidate recommendation list

6. Results and discussion

To evaluate the proposed target user specific reinforcement learning approach, widely used ecommerce Amazon dataset is considered. To examine the new approach in broader level, different product datasets are considered to check the influence and applicability of different user categories.

Dataset 1 is considered by combining the Digital music, Musical instruments, and CDs category datasets. This dataset covers the music recommendation aspects. Dataset 2 is considered for Video games and toys recommendation products. Home appliance recommendation is considered in Dataset 3 where Home and Kitchen, appliances and Grocery products are combined. Dataset 4 contains the sports and outdoor products. These dataset helps to cover all the important recommendation aspect and different product types. Table 4 shows the complete dataset details with the number of products and number of user review information.

Table 5. Datasets used for experiments

Dataset	Type	Product category	Data set details
Dataset 1	Music recommendation	Digital music + Musical instruments + CDs	1,100,200 product and 7,500,000 user reviews
Dataset 2	Toys & Games recommendation	Video Games + Toys and Games	715,200 product and 10,700,000 user reviews
Dataset 3	Home appliance recommendation	Home and Kitchen + Appliances + Grocery and Gourmet Food	1,590,000 products and 10,700,200 user reviews
Dataset 4	Sports product recommendation	Sports and outdoors	962,800 product and 12,980,800 user reviews

6.1. Evaluation method

To evaluate the proposed work, the experiments are conducted and compared with the similar existing recommendation approach. Traditional collaborative filtering approach is considered which the general recommendation methodology is in the recommendation system. To consider the feature aspect, feature based recommendation approach is considered for the evaluation. Multi factor influence (MDH) approach is taken for comparison which includes the demographic aspects. Fuzzy recommendation system is also included in the evaluation study to check the improvement of the proposed work with the previous research work. Recurrent neural network GRU approach is also taken into evaluation and compared against the proposed reinforcement learning approach.

6.2. Evaluation method

To evaluate, used both standard evaluation metrics Recall and precision and recommendation specific metrics. This proposed recommendation system has been evaluated based on the standard evaluation metrics such as Precision and Recall defined in the Eq. (3) and (4).

Recall is calculated by ratio of total number of recommended products are purchased by the customer to the total number of products that are purchased.

$$Recall = \frac{\text{No of Recommended Products purchased}}{\text{Total no.of products purchased}} \quad (3)$$

Precision is obtained by ratio of total of recommended products interested by customer to the total number of purchased products.

$$Precision = \frac{\text{No of products intersted by user in the recomended list}}{\text{Total no.of Products recommended}} \quad (4)$$

In addition to standard metrics, Recommendation specific metrics namely Diversity and Serendipity are also considered for evaluation. Diversity gives the similarity of the recommended products. For the quality recommendation, it is good to give some different products which user is interested instead of proposing all the products with similar type. This will make user more interesting and improve the customer satisfaction. Diversity is calculated by the below Eq. (5)

$$Diversity = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Sim}(x_i, y_j) \quad (5)$$

Serendipity metrics describes the interesting surprise products that user get it in the recommendation which user is interested, but not known to users or not explored by the user in the past. Serendipity is obtained by using the Eq. (6) and (7).

$$Serendipty_{customer} = \frac{1}{n}\sum_{x=1}^{n} \max(P_x(Customer) - P_x(allcustomers), 0) * rel_x\,(customer) \quad (6)$$

$$P_x = \frac{no. - rank_x}{no. - 1} \quad (7)$$

6.3. Target user specific categorization and ranking

User review processing and target user extraction helps to bucket the products with each user category and product is rated based on the sentiment expressed in the review. This clusters the products with each user category and ordered with respective rating. This helps the reinforcement policy to pick the highly recommended products for the given search context. For example, if user looks for the water bottle for 1 year old child. The recommendation system looks for the water bottle product with child user category and picks top rated products. Algorithm 1 helps to obtain this important information. Figure 2 shows how the products are categorized with specific user category.

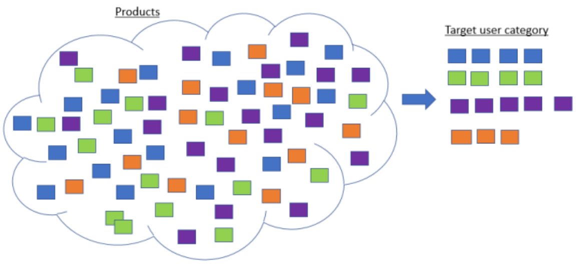

Figure 2. Product - target user category clustering

6.4. Extraction on temporal information

From user profile, referring the past purchase and browse history provides the temporal interested product details. Specific products are purchased at regular intervals. Long the product is purchased, and then likely user now interests to purchase that product. For example, if user purchased product P1 for every 2 months in last 6 months, then this product is likely to recommend if the interval of last purchase is close to 2 months. Algorithm 2 helps to extract this important information and helps the reinforcement policy to decide the top recommended products at certain time context. Figure 3 Shows the extraction of repeated, and sequence interested products from the past purchase and browse history.

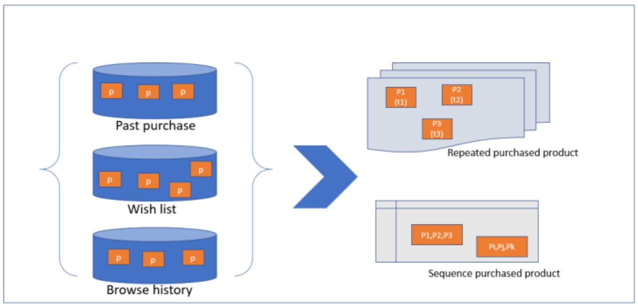

Figure 3. Temporal interest product extraction

6.5. Experimental results

Experiments are conducted by considering all the 4 datasets. Each recommendation methodology CF, Feature based, MDH and Fuzzy recommendation system are evaluated with the same datasets and results are compared. Figure 5 & Figure 6 shows the precision & recall outcome for all the techniques for each dataset.

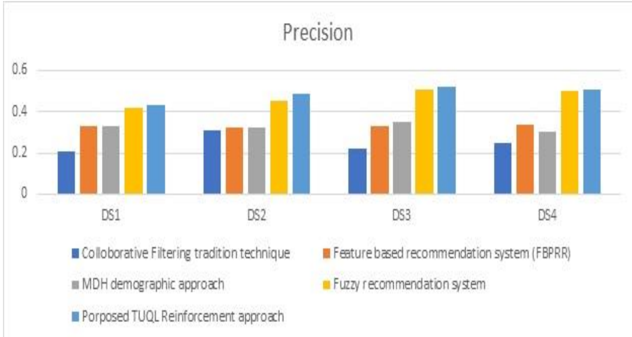

Figure 4. Precision Top 10 recommendation

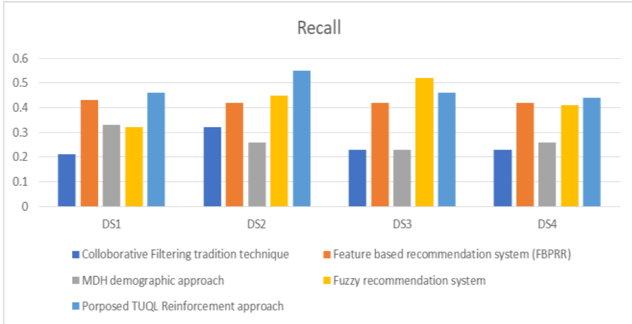

Figure 5. Recall Top 10 recommendation

Recommendation system specific serendipity shows better results when compared to the existing approaches. Especially when the sub sequent search, reinforcement policy able to adjust the recommendation and discard already recommended products when it is not interested by the user. Punished value is updated in the Q value item matrix. This gives greater possibility to show the interested unknown products to the user, meeting the short-term preference details. Figure 7 shows the Serendipity outcome for all the compared recommendation approaches.

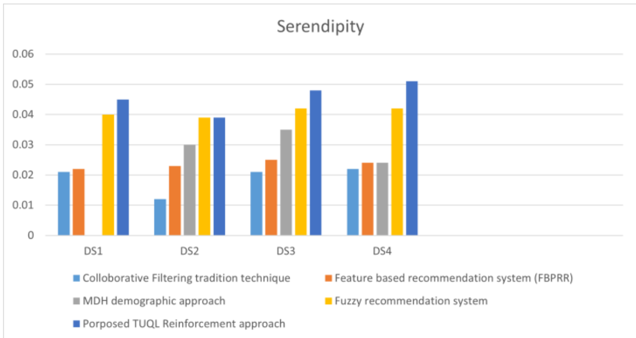

Figure 6. Serendipity Top 10 recommendations

7. Conclusion

In the proposed reinforcement approach, all the important influence factors are considered for both long term preferences and short-term user interest. Sentiment analysis technique is used to rate the products based on target user specific helps to reduce the dimension of the huge volume of products and improve the recommendation quality. Considering the temporal information of repeated and sequence interested products used to predict the recommendation with specific time context. This helps to avoid the false positive aspect of recommendation characteristic. Both target specific and temporal module greatly helps to improve decision making process in the reinforcement learning.

Construction of different clusters helps to give the surprise products to customer in each recommendation which makes user interesting; this gives good serendipity value when compared to existing similar approach. Especially the policy able to learn from the responses and update the value item matrix with positive value when recommendation is interested, else it dynamically adjusts by applying the negative value as punishment and use it for next recommendation decision. Experimental results show better results and improve the quality of recommendation. In Further work, proposed methodology can be extended by handling the inputs or feedback from the social media short messages or review blogs. This will improve the recommendation system further.

Author contributions

Karthik R.V: Conceptualization, Methodology, Software, Field study, Writing-Original draft preparation, Software
Ganapathy Sannasi: Validation, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Fatima Rodrigues and Bruno Ferreria,"Product recommendation based on shared customer behavior",Conference on enterprise information system,Procedia computer science, 2016, pp 136-146.

[2] Dezhuang Miaoa, Xuesong Lua, Qiwen Donga, Daocheng Honga," Humming-Query and Reinforcement-Learning based Modeling Approach for Personalized Music Recommendation", Procedia Computer Science 176 ,pp 2154–2163, 2020.

[3] Michael Reisener, Christian Dolle, Christopher Dierkes and Merle Hendrikje Jank,"Applying Supervised and Reinforcement Learning to Design Product Portfolios in Accordance with Corporate Goals", Procedia CIRP 91, 30th CIRP Design 2020, pp. 127–133.

[4] Joao Carapuc, Rui Neves, Nuno Horta,"Reinforcement Learning Applied to Forex Trading",Applied Soft Computing Journal,2018.

[5] Elena Krasheninnikova, Javier Garcia,Roberto Maestre and Fernando Fernandez," Reinforcement learning for pricing strategy optimization in the insurance industry", Engineering Applications of Artificial Intelligence 80,2019, pp. 8-19.

[6] Shenggong Ji, Zhaoyuan Wang, Tianrui Li and Yu Zheng," Spatio-temporal feature fusion for dynamic taxi route recommendation via deep reinforcement learning", Knowledge-Based Systems 205, 2020.

[7] Sydney Chinchanachokchai, Pipat Thontirawong and Punjaporn Chinchanachokchai,"A tale of two recommender systems: The moderating role of consumer expertise on artificial intelligence based product recommendations", Journal of Retailing and Consumer Services 61,2021.

[8] Ming-Chuan Chiu, Jih-Hung Huang, Saraj Gupta and Gulsen Akman,"Developing a personalized recommendation system in a smart product service system based on unsupervised learning model", Computers in Industry 128, 2021.

[9] Sang Hyun Choi and Byeong Seok Ahn,"Rank order-based recommendation approach for multiple featured products",Expert Systems with Applications 38,2011, pp. 7081-7087.

[10] Stefano Ferretti , Silvia Mirri , Catia Prandi , Paola Salomoni , Automatic Web Content Personalization Through Reinforcement Learning, The Journal of Systems & Software, 2016.

[11] Sandra Garcia Esparza, Michael P. O'Mahony and Barry Smyth, "Mining the real-time web: A novel approach to product recommendation", Knowledge-Based Systems 29, 2012, pp. 3–11.

[12] Gang Ke, Hong-Le Du and Yeh-Cheng Chen,"Cross-platform dynamic goods recommendation system based on reinforcement learning and social networks",Applied Soft Computing Journal 104,2021.

[13] Chin-Hui Lai , Shin-Jye Lee and Hung-Ling Huang ,"A Social Recommendation Method Based on the Integration of Social Relationship and Product Popularity", International Journal of Human-Computer Studies,2018.

[14] Gyudong Lee, Jaeeun Lee and Clive Sanford,"The roles of self-concept clarity and psychological reactance in compliance with product and service recommendations", Computers in Human Behavior 26, 2010, pp. 1481–1487.

[15] Feng Liu, Ruiming Tang, Huifeng Guo, Xutao Li, Yunming Ye and Xiuqiang He,"Top-aware reinforcement learning based recommendation",Neurocomputing 417, 2020, pp. 255–269.

[16] Shan Liu, Hai Jiang, Shuiping Chen, Jing Ye, Renqing He and Zhizhao Sun,"Integrating Dijkstra's algorithm into deep inverse reinforcement learning for food delivery route planning",Transportation Research Part E 142, 2020.

[17] Andre Marchand and Paul Marx,"Automated Product Recommendations with Preference-BasedExplanations",Journal of Retailing 96, 2020, pp. 328–343.

[18] JaeKwan Park, TaekKyu Kim and SeungHwan Seong,"Providing support to operators for monitoring safety functions using reinforcement learning",Progress in Nuclear Energy 118, 2020.

[19] Lingyun Qiu and IzakBenbasat,"A study of demographic embodiments of product recommendation agents in electronic commerce",Int. J.Human-ComputerStudies 68, 2010, pp. 669–688.

[20] Manuela Ruiz-Montiel,Javier Boned, Juan Gavilanes, Eduardo Jimenez, Lawrence Mandow and Jose-Luis Perez-de-la-Cruz," Design with shape grammars and reinforcement learning",Advanced Engineering Informatics 27, 2013, pp. 230–245.

[21] Karthik R.V, Sannasi Ganapathy and Arputharaj Kannan, "A Recommendation System for Online Purchase Using Feature and Product Ranking", Eleventh International Conference on Contemporary Computing (IC3), pp. 1-6, 2018.

[22] R. V. Karthik and Sannasi Ganapathy,"A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce", Applied Soft Computing Journal, Vol. 108, pp.1-18, 2021.

[23] R. V. Karthik and Sannasi Ganapathy, "Online Product Recommendation System Using Multi Scenario Demographic Hybrid (MDH) Approach", IFIP International Federation for Information Processing, ICCIDS 2020, IFIP AICT, Vol. 578, pp. 1–13, 2020.

[24] Yang Zhang, Daniel (Yue) Zhang, Nathan Vance, Dong Wang, "An online reinforcement learning approach to quality-cost-aware task allocation for multi-attribute social sensing",Pervasive and Mobile Computing 60, 2019.

[25] Jianmo Ni, Jiacheng Li, Julian McAuley, "Empirical Methods in Natural Language Processing (EMNLP)", 2019. https://nijianmo.github.io/amazon/index.html.