A Personalized Recommendation for Web API Discovery in Social Web of Things

Marwa Meissa¹, Saber Benharzallah², Laid Kahloul¹, and Okba Kazar¹

¹LINFI Laboratory, Biskra University, Algeria
²Department of Computer Science, LAMIE Laboratory Batna 2 University, Algeria

Abstract: With the explosive growth of Web of Things (WoT) and social web, it is becoming hard for device owners and users to find suitable web Application Programming Interface (API) that meet their needs among a large amount of web APIs. Social-aware and collaborative filtering-based recommender systems are widely applied to recommend personalized web APIs to users and to face the problem of information overload. However, most of the current solutions suffer from the dilemma of accuracy-diversity where the prediction accuracy gains are typically accompanied by losses in the diversity of the recommended APIs due to the influence of popularity factor on the final score of APIs (e.g., high rated or high-invoked APIs). To address this problem, the purpose of this paper is developing an improved recommendation model called (Personalized Web API Recommendation) PWR, which enables to discover APIs and provide personalized suggestions for users without sacrificing the recommendation accuracy. To validate the performance of our model, seven variant algorithms of different approaches (popularity-based, user-based and item-based) are compared using MovieLens 20M dataset. The experiments show that our model improves the recommendation accuracy by 12% increase with the highest score among compared methods. Additionally it outperforms the compared models in diversity over all lengths of recommendation lists. It is envisaged that the proposed model is useful to accurately recommend personalized web API for users.

Keywords: Web of Things, recommender system, web API, collaborative filtering, rating prediction, social networks, IoT.

Received February 20, 2021; accepted March 7, 2021
https://doi.org/10.34028/iajit/18/3A/7

1. Introduction

In recent years, the Internet has been enriched by a huge amount of data. Especially, with the growing mass movement of connected devices in what is known as Internet of Things (IoT). The emergence of IoT contributes to massively inflated data, where, more than 38 billion connected devices in 2025 and is estimated to reach 50 billion by 2030¹, which each person in the world projected to have six connected devices. This transformation from real-world functionalities of devices to the digital world has essentially led up to the emergence of a new generation of software and applications, which is developed to be consumed by these things and allows them to integrate and to communicate with various other entities on the web. Hence, the application programming interface presentation is widely used to expose the different functionalities of IoT devices and enables them to exchange data. Moreover, it allows users to control and monitoring their devices. In this context, intuitively, users and device owners connect those devices via their smartphones and search for suitable web Application Programming Interface APIs to be consumed.

Recently, the rapid spread of Application Programming Interface (APIs) due their benefits for web of things made users confused to choose the most appropriate web API among a huge number of APIs. Furthermore, it is becoming more and more challenging for users to discover web APIs in this big data environment. In front of this massive growth of web APIs, Recommender Systems (RS) are one of the most trends that are built to face data proliferation challenge, which enables to filter the pool of web APIs and help users to find suitable ones by giving them personalized and relevant suggestions. Despite this urgent need to develop recommendation engine for APIs in WoT, there is very few publications in the literature that have addressed the API recommendation issues in IoT environment. Goa et al. [7] provide a recommendation model that based on collaborative learning for IoT API in industrial systems. Similarly, Cao et al. [4] propose a web API recommendation method based on Quality of Service (QoS) for IoT mashup applications. However, the above approaches are oriented developers, which the crux of recommendation engine is mainly recommending the most appropriate web APIs among the huge number of functionally equivalent APIs. Therefore, the recommendation engine neglects

personalization and it offers the same recommended APIs for all users. So far, personalized recommendation is promising way to handle the problem of information overload and especially to meet individualities of device owners.

Collaborative Filtering (CF) approaches are widely used to generate classical personalized recommender systems [12, 15, 14]. In addition, the social knowledge that offered by social networks are also exploited to improve the performance of personalized recommender systems [2]. In our previous work [17], we propose a social-aware recommendation approach for web APIs discovery and selection based on rating prediction. The evaluation metrics of the prediction accuracy typically is divided into two main classes [6, 9]: decision-support metrics and statistical accuracy metrics. Decision-support accuracy evaluates how effective a prediction model is at helping a user select high-quality items. Statistical accuracy metrics measures the accuracy of a prediction engine by comparing predicted values with real values as we have employed in our previous work [17]. In this paper, we extend the experiments of the previous paper by employing decision-support measures to evaluate the prediction accuracy. Furthermore, we measure the diversity of recommendation results in order to address the problem of accuracy-diversity [8, 29]. Therefore, our proposed model PWR recommends accurately personalized web APIs for users in WoT environment. In summary, the major contributions of this paper are:

- Presenting a novel model for personalized web API recommendation in social WoT.
- Proposing an enhanced user nearest neighbour selection method that combined explicit and implicit feedbacks of users.
- Providing an enhanced method for rating prediction that have been experimented on MovieLens dataset2.

The remainder of this paper is structured as follows: the second section presents a motivated scenario. Then the third section define preliminary. Section 4 presents the proposed recommendation model in detail. In section 5, we demonstrates the experiments followed by the results and discussion in section 6. In section 7, we overview the related literature to our work. Finally, the paper in concluded in section 8.

2. A Motivated Scenario

To motivate our approach, we present running scenario in E-Health-based IoT ecosystem as shown in Figure 1. The scenario is as follow: “Alice has diabetes; she got an electronic Glucose meter. She is looking for an APIs that can be used for her device (e.g., API that enables her to send her Glucose measurement to her doctor). Firstly, she connect her device via mobile phone, and then send a request to the recommender system. Here, recommendation engine searches for similar users to Alice (e.g. are also diabetic) who have been used the same kind of device before. Then, it chooses the APIs that have been invoked by those users (Sophia, Emma, James). Typically, there are a huge amount of APIs, thus, recommender engine uses rating values of APIs to select the highest ones and recommend the Top-K APIs to Alice.”

Figure 1. Running scenario in E-health-based IoT system.

3. Preliminary

Figure 2 illustrates the social WoT by proposing a model that based on graph representation. We denote a tripartite Graph $G=(U, W, O)$, where $U$ is the set of user $U=\{u_1, u_2, … u_m\}$, $W$ is the set of web APIs $W=\{a_1, a_2, … a_n\}$ and $O=\{d_1, d_2, … d_l\}$ is the set of IoT devices. The model is based on two basic components: nodes, and social links. The nodes are users, web APIs and IoT devices, and social links are multi-relationships between them. The social links expressed by four relationships: R1, R2, R3, R4. R1 refers to user-user relation, it is employed to select user neighbours’ based on his explicit feedback (i.e., ratings) and implicit (co-used devices and preference similarities). R2 denotes the device-to-device similarity. R3 presents the user-web API rating relation, it is expressed by rating matrix $R$ where $R$ is of dimension $m \times n$ and the entry $r(u, a)$ is the rating value given by user $u$ to web API $a$. Finally, R4 defines the relation of invocation between web API and device, this social link is represented by a matrix $D=\{l_{u, d}\}_{m \times l}$. In this matrix, $l_{u, d}$ denotes the user $u$ invocation of an API for device $d$.

---

2https://grouplens.org/datasets/movielens/20m/
4. The Proposed Recommendation Model

In this section, we present our recommendation approach. The core idea of the proposed model is to provide relevant and personalized suggestions of APIs for WoT users. The structure of recommendation model is illustrated in Figure 3. The recommendation process is divided into two phases: online and offline. In the offline phase, there are two main stages, which are device similarity computation and k-nearest neighbours’ selection for user. The online phase includes two main tasks: rating prediction and API ranking.

4.1. Offline Phase

This phase consists of two models: user similarity and device similarity. The aim of user similarity model is to select neighbors of the target user in social WoT. The purpose of device similarity model is to find similar devices of the target device. These models are adopted in offline to reduce the time cost of recommendation due the requirement of IoT environment such as resource constraint of IoT device.

4.2. Online Phase

This section presents the online phase of our recommendation model. Two stages are proposed: rating prediction and API ranking.

a. User Similarity Model

In order to measure the similarity level between users, we propose a similarity module that calculates the similarity degree. This module is based on three factors as shown in the next Equations:

- Preference similarity refers the similarity of user preferences. It is calculate by Jaccard distance as follows:
  \[ \text{sim}_p(a, b) = \frac{|P_a \cap P_b|}{|P_a| \cup |P_b|} \]  
  \[ (1) \]
  Where \( \text{sim}_p(a, b) \) denotes the degree of similarity between user a and user b, \( P_a \) and \( P_b \) refers to the set of preferences of users a and b, respectively.

- Co-used devices similarity is computed the similarity degree between user a and b, where \( D_a \) and \( D_b \) refers to the set of used devices by user a and b, respectively.
  \[ \text{sim}_d(a, b) = \frac{|D_a \cap D_b|}{|D_a| \cup |D_b|} \]  
  \[ (2) \]
  \( \text{sim}_d(a, b) \) is the similarity degree between user a and b, where \( D_a \) and \( D_b \) refers to the set of used devices by user a and b, respectively.

- Invocation similarity represents the similarity of invocation behavior among two users a and b, it is computed by Pearson Correlation Coefficient (PCC) of both user ratings \( r_{a,x} \) and \( r_{b,y} \) on API s as follows:
  \[ \text{sim}_i(a, b) = \frac{\sum_{s \in S} (r_{a,s} - \bar{r}_a)(r_{b,s} - \bar{r}_b)}{\sqrt{\sum_{s \in S} (r_{a,s} - \bar{r}_a)^2 \sum_{s \in S} (r_{b,s} - \bar{r}_b)^2}} \]  
  \[ (3) \]
  The set of similar users is defined as neighbors \( (a) = \{ b | \text{sim}(a, b) > 0 \} \), where the final similarity degree \( \text{sim}(a, b) \) is given by the following formula:
  \[ \text{sim}(a, b) = \frac{\text{sim}_p(a, b) + \text{sim}_d(a, b) + \text{sim}_i(a, b)}{3} \]  
  \[ (4) \]

b. Device Similarity

The similarity \( \text{Dis} (x, y) \) between two devices x and y is computed by the Euclidean distance as in Equation (5), where the \( F_i \) represent the features of device such as device type, availability time, location, mobility…etc. Here, we used device profile model that proposed in [10]:
  \[ \text{Dis} (x, y) = \sqrt{\sum_{i=0}^{N}[F_i(x) - F_i(y)]^2} \]  
  \[ (5) \]
by similar users of the target user and have been consumed by similar devices to target device are selected as candidate APIs.

The aim of this stage of recommendation is to predicting rating values of the candidate APIs. The proposed method is combined user-based CF and item-based CF technique to predict the missing ratings and to score the APIs via employing user similarity. Then, it is polymerizing the results returned from every iterations. Finally, the K-largest score of APIs are selected for the next stage. We have used the following equation to predict the rating value.

\[
\text{Pred}(u, a) = \frac{\bar{R}(a)}{\bar{R}(u, a) - \bar{R}(a)} \times \frac{\text{sim}(u, v)}{\max_{u, u'} \text{sim}(u, u')}
\]

Where \(\text{Pred}(u, a)\) is the predicted rating of API \(a\) by the user \(u\), \(\bar{R}(a)\) is the average ratings of API \(a\) and finally, \(\text{sim}(u, v)\) is the similarity degree among user \(u\) and the target user \(v\).

b. Web API Ranking

From the previous stage, for each API \(a\) in the set of candidate APIs, Equation (6) is applied to calculate the relevancy degree. Where \(N_u\) is the number of similar users whom invoked API, \(N_v\) is the number of similar devices that have consumed this API. \(U; O\) are cardinalities of total users and devices, respectively.

\[
\text{Relevancy}(a) = \frac{N_u}{U} + \frac{N_v}{O}
\]

The predicted rating score that given for each API in prediction stage is adjusted by relevancy degree. Then the APIs have been ranked according to their final adjusted score as show in the following algorithm:

**Algorithm API recommendation in PWR model**

**Input**

\(M\) User-web API rating matrix

\(R\) Relevancy vector

neighbors Set of user neighbors

**Method**

For each API \(a\) \(\epsilon\) set of candidate APIs do

For each user \(u\) \(\epsilon\) neighbors do

\[P[i][j] \leftarrow \text{Pred}(u, j, a, i)\]

End for

\[S[a, i] \leftarrow 1/N \sum_{U \epsilon U} \sum_{j \epsilon U} \max(P[i][j] \times \text{Relevancy}(a, i))\]

End for

\([\text{recommended}]_l\text{ist} \leftarrow a, j \text{ Top-k} \{S[a, i]\}\]

Output \([\text{recommended}]_l\text{ist}\)

### 5. Experimental Evaluation

For experimental evaluation, due the lack of real-world dataset that meets our benchmark to validate the performance of our proposed approach, we select MovieLens 20M dataset from Group lens Research Project. This data set consists of 138,000 users, 27,000 different movies and 465,000 tag applications; the total number of rating is 20 million ratings. In this paper, we consider the movies as APIs and tags as IoT devices. We filtered the dataset so that only the users who had tags were selected, meaning that the ratings for users who own devices was taken only. We define two matrices \(R\) and \(D\): \(R\) is the user-web API rating matrix and \(D\) is the user-device-web API invocation matrix. The filtered dataset statistics shown in the Table 1.

<table>
<thead>
<tr>
<th>#Ratings</th>
<th>4601</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Users</td>
<td>321</td>
</tr>
<tr>
<td>#APIs</td>
<td>2307</td>
</tr>
<tr>
<td>#Devices</td>
<td>747</td>
</tr>
<tr>
<td>Range of Rating</td>
<td>[0.5,5]</td>
</tr>
<tr>
<td>K-density</td>
<td>0.62 %</td>
</tr>
<tr>
<td>D-density</td>
<td>1.17</td>
</tr>
</tbody>
</table>

### 5.1. Compared Methods

To verify the performance of our PWR model in this paper, we selected seven baselines to compare with the proposed approach. These are:

1. **UPCC** is the traditional user-based collaborative filtering method, which exploiting the historical behavior of users to compute users similarity by Pearson Correlation Coefficient for making prediction [17, 20, 22, 28].
2. **IPCC** is adopting Pearson Correlation Coefficient, which gets the predicted rating based only on the similarity between items [13, 19, 23].
3. **UC-KNN** employs similar user attribute information for web API recommendation based on Cosine similarity and k-nearest neighbor algorithm [18].
4. **IC-KNN** is the item-based CF model for rating perdition using k-nearest neighbor algorithm based on the Cosine similarity measure.
5. **PHR** is a popularity-based recommendation baseline. We define popularity of an API as the high rated APIs in rating matrix (i.e., the average rating of API).
6. **PMI** refers to the recommendation of APIs based on their invocation frequency by IoT devices.
7. **PHS** is a variation of the popularity-based model that defines the popularity of API by the high-scored APIs.

Among the above, 1) and 3) are user-based CF methods, 2) and 4) are item-based CF methods, and 5), 6) and 7) are popularity-based models. For the experiments, the dataset is divided into two parts: 20% were randomly selected to represent the test data and 80% constitutes the training set. Our experiment is implemented on a PC with Intel Core i5 CPU and 4 GB RAM under windows 7 using Python 3.7.

### 5.2. Evaluation Metrics

To evaluate the quality of recommendation in our PWR model, firstly, we use the personalization metric that assess if a recommendation model suggests the same items to different users. It is employed to measure the personalization degree of our recommendation model...
in comparison with other models. Personalization (PER) is defined by the dissimilarity between user’s lists of recommendation as following:

\[
PER@N = 1 - \cos(L_u, L_v)
\]

(8)

Where \( N \) the number of top-N recommended APIs. \( \cos(L_u, L_v) \) denotes the Cosine similarity between the recommended API list of user \( u \) and the recommended list of user \( v \).

To evaluate the prediction accuracy of the proposed model, we use the Receiver Operating Characteristic (ROC) metric, which indicates to the quality of recommended web APIs. The ROC metric is defined as following:

\[
ROC = \frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} a_i}
\]

(9)

Where \( d_i \) and \( a_i \) refer to the probability of detection and the probability of false alarm respectively. In our experiments, for the predicted ratings \( \{p_1, p_2, p_3, \ldots, p_n\} \) and real ratings \( \{r_1, r_2, r_3, \ldots, r_n\} \), we use two thresholds \( T_1, T_2 \), and we also define \( d_i \) and \( a_i \) as following:

\[
d_i = \begin{cases} 1 & \text{if } p_i \geq T_1 \text{ and } r_i \geq T_2 \\ 0 & \text{else} \end{cases}
\]

(10)

\[
a_i = \begin{cases} 1 & \text{if } p_i \geq T_1 \\ 0 & \text{else} \end{cases}
\]

(11)

6. Results and Discussion

In the following, we present the results of the experiments in order to highlight how the proposed model PWR outperforms the other compared models by the prediction accuracy and personalization performance.

6.1. Prediction Accuracy Performance

We compare our prediction model PWR with other compared methods in ROC metric. We set the thresholds for real ratings. We set \( T_1 \) equals to the median \( T_2 \geq 4 \) between the highest predicted score and the lower predicted score in each API for all models. Figure 4 shows the results of the comparison.

![Figure 4. The comparison of ROC metric between the compared models.](image)

The results show that under all ROC values of the compared algorithms our model PWR achieves the highest value. That means that the prediction in our model is more accurate than other models. Additionally, we observe that the lower score in ROC is obtained in item-based algorithms, we also observe that the popularity-based models (PMI, PHS and PHR) achieve a higher ROC values than the baseline user-based and item-based approaches. The superiority our PWR model over all the compared algorithms is confirmed by a 12% increase in the recommendation quality with popularity-based approaches.

6.2. Personalization Evaluation

Table 2 shows the results of obtained personalization degree over different Top-N recommendation \( (3, 5, 7, 10, 12, 15, \text{and } 20) \) in order to see how our model improves the diversity of recommendation results.

<table>
<thead>
<tr>
<th>Model</th>
<th>Top-N</th>
<th>Per@3</th>
<th>Per@5</th>
<th>Per@10</th>
<th>Per@15</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-based models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPCC</td>
<td>0.0009</td>
<td>0.24</td>
<td>0.05</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>UC-KNN</td>
<td>0.0002</td>
<td>0.12</td>
<td>0.003</td>
<td>0.00001</td>
<td></td>
</tr>
<tr>
<td>Popularity-based alg.</td>
<td>0.0</td>
<td>0.00</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Item-based alg.</td>
<td>0.0</td>
<td>0.00</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Our Model PWR</td>
<td>0.29</td>
<td>0.86</td>
<td>0.38</td>
<td>0.39</td>
<td></td>
</tr>
</tbody>
</table>

From the above results in Table 2. We can make the following observations:

- Under all personalization values, PWR achieved the largest values even by varying the length N of recommended lists. It is a significant that our model PWR recommends higher personalized results than the other baselines.
- User-based models outperformed the other models (popularity and item-based algorithms). That indicates that employing user similarity is beneficial in personalized recommendation.
- Popularity-based and item-based models achieved the lower score. That means that these models recommend the same web APIs for different users which is the opposite of what should be in personalized recommendation oriented users in IoT environment.

7. Related Works

In this section, we present the related literature as three aspects: Recommendations systems in IoT environment, social network-based recommendation and rating prediction-based approaches.

7.1. Recommendation Systems in IoT

Recommender systems have emerged as powerful tool that have been built to help users on the Internet and guide them to select the most appropriate items from a massive amount of data. Additionally, the prevalence of IoT led up to a tremendous growth of data. Therefore, a numerous of RS oriented IoT ecosystems have been recently proposed in literature. In this works
thing recommendation systems are developed. On the other side, the convergence of IoT and the World Wide Web (WWW) in what is known as WoT gave rise to a new generation of RS. So far, Recommendation oriented web is an increasingly important area in research. Cao et al. [4] have been proposed a QoS-aware web APIs recommendation for mashup creation. Similarly, Mashal et al. [16] proposed a recommendation mechanism for IoT services recommendation, which various relationships among user, services and objects are analysed and exploited. Very recently, Tang et al. [22] provided a novel method for APIs recommendation in IoT environment, the proposed method is based on historical data on APIs and keywords that describe the functionalities of APIs.

7.2. Web API Recommendation-based on Social Networks

With the emergence of social web, social information are widely employed to enhance recommendation performance. This work [27] proposed a platform for service recommendation, which explored user preference and tagging relation among users and service to predict missing values of QoS. The authors in [11] provided a recommendation mechanism for web service discovery that based on social network, which they integrated friendships relations among users and services to improve recommendation performance. In order to predict missing values of rating, the social information are also employed in service recommendation in [5]. The recommendation in [3] is based on social network and user interest, which the method aims to provide a mashup service for the users, based on two defined social relationships: calling relation between mashups services and web APIs and marking relation among web APIs/mashups services and tags.

7.3. Rating Prediction-based Recommendation

In recommendation process, filtering stage is a main task. Two classical methods are applied to filter items and users: CF-based filtering [1] and content-based filtering. In the context of web-based recommendation such as web services, APIs and apps recommendation, CF is widely applied for predicting missing values, (e.g., QoS values Rating, feedbacks...etc.), CF-based filtering is applied in [26] for rating prediction; which the authors combined user-based CF and social network for similarity computation in order to enhance neighbour selection phase. Similarly, the authors in [5] proposed a recommendation mechanism that based on rating prediction, which enables providing personalized results for users in social network-based environment. In this paper, we employed CF-based filtering to filter users and improving neighbours’ selection by using historical information and social information. Then we applied content-based filtering for device similarity computation.

8. Conclusions

This paper presented a personalized API recommendation model in social WoT environment. The basic idea is to predict rating values and recommend the top-k APIs, based on their adjusted rating score that is computed based on their relevancy degree and their predicted rating that is weighted by similarity values between users and the target user. Additionally, a similarity measurement model is proposed that based on three factors, which are co-invocation, co-used and preference. The rating prediction method is combined item-based and user-based CF techniques in order to enhance prediction accuracy. The experimental results demonstrate that performance of our recommendation system outperforms the other compared methods in both prediction accuracy and diversity of recommend lists. In the future work, we will further explore more the social knowledge to enhance recommendation performance. We also plane to investigate the utilization of trust concept of users, devices and web APIs to enhance recommendation credibility.

References


**Marwa Meissa** is a PhD student in intelligent Computer Science Laboratory University Mohamed Khider Biskra, Algeria. Her interests include service composition, recommender system, social computing and Internet of Things.

**Saber Benharzallah** is a professor and researcher in the computer science department of Batna 2 University (Algeria). Received his Ph. D degree in 2010 from the Biskra University (Algeria). Prof. Benharzallah is currently director of laboratory LAMIE (Batna 2 University). His research interests include Internet of things, service-oriented architecture, context aware systems, Social IoT.

**Laid Kahloul** received the Ph.D. degree in computer software and theory from the Computer Science Department, Biskra University, Biskra, Algeria, in 2012. He is currently a professor with the Computer Science Department, Biskra University, Algeria.

**Okba Kazar** professor in the Computer Science Department of Biskra. He is a member of international conference program committees and the "editorial board" for various magazines. His research interests are artificial intelligence, multi-agent systems, web applications and information systems.