Personality-Aware Product Recommendation System Based on User Interests Mining and Metapath Discovery

Sahraoui Dhelim, *Member, IEEE*, Huansheng Ning, *Senior Member, IEEE*, Nyothiri Aung, Runhe Huang, *Senior Member, IEEE*, and Jianhua Ma, *Member, IEEE*

***Abstract*—A recommendation system is an integral part of any modern online shopping or social network platform. The product recommendation system as a typical example of the legacy recommendation systems suffers from two major drawbacks: recommendation redundancy and unpredictability concerning new items (cold start). These limitations take place because the legacy recommendation systems rely only on the user's previous buying behavior to recommend new items. Incorporating the user's social features, such as personality traits and topical interest, might help alleviate the cold start and remove recommendation redundancy. Therefore, in this article, we propose Meta-Interest, a personality-aware product recommendation system based on user interest mining and metapath discovery. Meta-Interest predicts the user's interest and the items associated with these interests, even if the user's history does not contain these items or similar ones. This is done by analyzing the user's topical interests and, eventually, recommending the items associated with the user's interest. The proposed system is personality-aware from two aspects; it incorporates the user's personality traits to predict his/her topics of interest and to match the user's personality facets with the associated items. The proposed system was compared against recent recommendation methods, such as deep-learning-based recommendation system and session-based recommendation systems. Experimental results show that the proposed method can increase the precision and recall of the recommendation system, especially in cold-start settings.**

***Index Terms*—Big-five model, personality computing, product recommendation, recommendation system, social networks, social computing, user interest mining, user modeling.**

I. Introduction

WITH the widespread of personal mobile devices and the ubiquitous access to the internet, the global number of digital buyers is expected to reach 2.14 billion people within the next few years, which accounts for one-fourth of the world population. With such a huge number of buyers and the wide variety of available products, the efficiency of an online store is measured by their ability to match the right user with the right product; here comes the usefulness of product recommendation systems. Generally speaking, product recommendation systems are divided into two main classes.

1) *Collaborative Filtering (CF):* CF systems recommend new products to a given user based on his/her previous (rating/viewing/buying) history and his/her neighbors (similar users). For example, as shown in Fig. 1(a), most of the people previously bought a football jersey, and they have also bought a football; thus, the system predicates that the user might be interested in buying a football.
2) *Content Filtering or Content-Based Filtering (CBF):* CBF systems recommend new items by measuring their similarity with the previously (rated/viewed/bought) products. For example, as shown in Fig. 1(b), football is recommended because it is semantically similar to the football jersey.

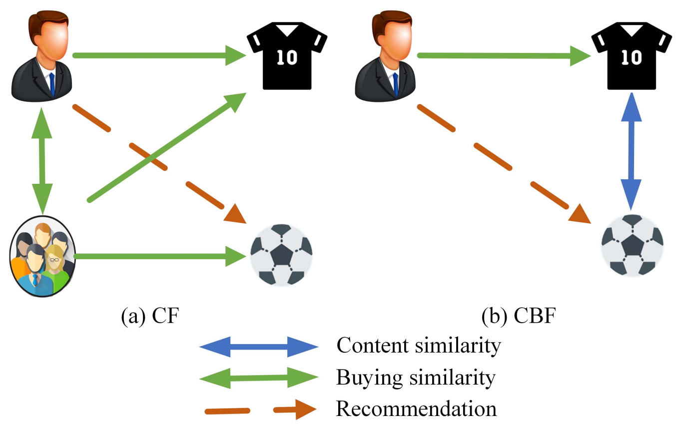


Fig. 1. Collabortive filtering and content filtering.

Far from that, with the popularity of online social networks, such as Facebook, Twitter, and Instagram, many users use social media to express their feeling or opinions about different topics or even explicitly expressing their desire to buy a specific product in some cases, which made social media content a rich resource to understand the users' needs and interests [1]. On the other hand, the emerging of personality computing [2] has offered new opportunities to improve the efficiency of user modeling in general and particularly recommendation systems by incorporating the user's personality traits in the recommendation process. In this work, we propose a product recommen-

Manuscript received September 24, 2019; revised September 5, 2020 and October 22, 2020; accepted November 7, 2020. Date of publication November 24, 2020; date of current version January 29, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61872038, and in part by the Fundamental Research Funds for the Central Universities under Grant FRF-BD-18-016A. *(Corresponding author: Huansheng Ning.)*

Sahraoui Dhelim, Huansheng Ning, and Nyothiri Aung are with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: ninghuansheng@ustb.edu.cn).

Runhe Huang and Jianhua Ma are with the Faculty of Computer and Information Sciences, Hosei University, Tokyo 184-8584, Japan.

Digital Object Identifier 10.1109/TCSS.2020.3037040

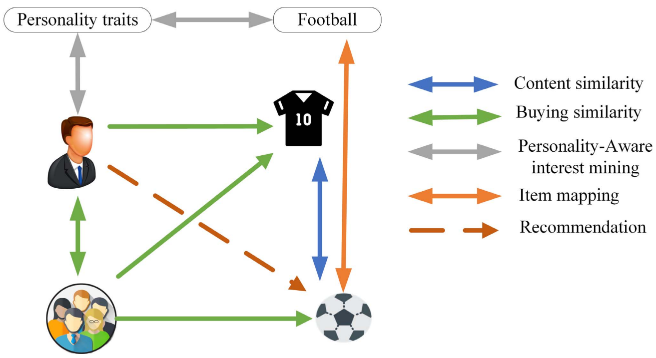


Fig. 2. Interest mining-based product recommendations.

dation system that predicts the user's needs and the associated items, even if his/her history does not contain these items or similar ones. This is done by analyzing the user's topical interest and, eventually, recommending the items associated with the theses interest. The proposed system is personality-aware from two aspects; it incorporates the user's personality traits to predict his/her topics of interest and to match the user's personality facets with the associated items. As shown in Fig. 2, the proposed system is based on a hybrid filtering approach (CF and CBF) and personality-aware interest mining.

Since we have multiple types of nodes (users, items, and topics), the system is modeled as a heterogeneous information network (HIN), which includes multiple types of nodes and links. In our case, product recommendation could be formulated as link prediction in HIN [3]. For example, in Fig. 2, given the user's previous rating and topical interest represented in an HIN, the problem is to predict whether or not a link exists between the user and the product (the ball). One of the main challenges of link prediction in HIN is how to maintain a reasonable balance between the size of information considered to make the prediction and the algorithm complexity of the techniques required to collect that information. Since, in practice, the networks are usually composed out of hundreds of thousands or even millions of nodes, the method used to perform link prediction in HIN must be highly efficient. However, computing only local information could lead to poor predictions, especially in very sparse networks. Therefore, in our approach, we make use of metapaths that start from user nodes and end up in the predicted node (product nodes in our case), and we try to fuse the information from these metapaths to make the prediction.

The contributions of this work are summarized as follows.

1) Propose a product recommendation system that infers the user's needs based on his/her topical interests.
2) The proposed system incorporates the user's big-five personality traits to enhance the interest mining process and perform personality-aware product filtering.
3) The relationship between the users and products is predicted using a graph-based metapath discovery; therefore, the system can predict implicit and explicit interests.

The remainder of this article is organized as follows. In Section II, we review the related works. In Section III, the system design of the proposed system is presented. In Section IV, we evaluate the proposed system. Finally, in Section V, we conclude the work and state some of the future directions.

II. Related Works

In this section, we review the recent advances of personality-aware recommendation system and interest mining schemes as well.

A. Personality and Recommendation Systems

Many works have discussed the importance of incorporating the user's personality traits in the recommendation systems. Yang *et al.* [4] proposed a recommendation system of computer games to players based on their personality traits. They have applied text mining techniques to measure the players' Big-five personality traits and classified a list of games according to their matching with each dominant trait. They have tested their proposed system on 2050 games and 63 players from the Steam gaming network. While Wu *et al.* [5] presented a personality-based greedy reranking algorithm that generates the recommended list, where the personality is used to estimate the users' diversity preferences, Ning *et al.* [6] proposed a friend recommendation system that incorporates the big-five personality traits model and hybrid filtering, where the friend recommended process is based on personality traits and the users' harmony rating. Ferwerda *et al.* [7] studied the relationship between the user's personality traits and music genre preferences; they have analyzed a data set that contains personality test scores and music listening histories of 1415 Last.fm users. Similarly, in [8], they conducted an online user survey where the participants were asked to interact with an application named Tune-A-Find and measured taxonomy choice (i.e., activity, mood, or genre), individual differences (e.g., music expertise factors and personality traits), and different user experience factors. Similarly, Hafshejani *et al.* [9] proposed a CF system that clusters the users based on their big-five personality traits using the K-means algorithm. Following that, the unknown ratings of the sparse user–item matrix are estimated based on the clustered users. Dhelim *et al.* [10] discussed the benefits of capturing the user's social feature, such as personality traits that are represented as cyberentities in cyberspace. Similarly, Khelloufi *et al.* [11] showed the advantages of leveraging the user's social features in the context of service recommendation in the Social Internet of Things (SIoT).

B. Interest Mining

Far from personality, many previous works have discussed user interest mining from social media content. Piao *et al.* [1] surveyed the literature of user interest mining from social networks, and the authors reviewed all the previous works by emphasizing the following on four aspects: 1) data collection; 2) representation of user interest profiles; 3) construction and refinement of user interest profiles; and 4) the evaluation measures of the constructed profiles. Zarrinkalam *et al.* [12] presented a graph-based link prediction scheme that operates over a representation model built from three categories of information: user explicit and implicit contributions to topics, relationships between users, and the similarity among topics.

TABLE I
COMPARISON WITH RELATED WORKS

Recommendation system	Recommended content	Personality model	User interest	Representational model	Recommendation technique
Meta-Interest	products	Big-Five	Yes	HIN	personality-aware meta-paths filtering
metapath2vec [20], Shi *et al.* [21]	generic	No	No	HIN	meta-paths embedding
GNN-SEAL[22]	generic	No	No	graph neural network	heuristics from local subgraphs
Song *et al.* [23]	social	No	Yes	graph-attention neural network	session-based social recommendation
PersoNet[6]	friends	Big-Five	No	homogeneous network	collaborative filtering
Yang *et al.* [4]	games	Big-Five	No	homogeneous network	content filtering
Hafshejani *et al.* [9]	products	Big-Five	No	homogeneous network	K-means clustering

Trikha *et al.* [13] investigated the possibility of predicting the users' implicit interests based on only topic matching using frequent pattern mining without considering the semantic similarities of the topics. While Wang *et al.* [14] proposed a regularization framework based on the relation bipartite graph that can be constructed from any kind of relationship, they evaluated the proposed system from social networks that were built from retweeting relationships. Dhelim *et al.* [15] discussed the usage of the user's interests to customize the services offered by a cyber-enabled smart home. Faralli *et al.* [16] proposed Twixonomy, a method for modeling of Twitter users by a hierarchical representation based on their interests. Twiconomy is built by identifying topical friends (a friend represents an interest instead of a social relationship) and associate each of these users with a page on Wikipedia. Dhelim *et al.* [17] used social media analysis to extract the user's topical interest. Kang *et al.* [18] proposed a user modeling framework that maps the user' posted content in social media into the associated category in the news media platforms, and based on that, they used Wikipedia as a knowledge base to construct a rich user profile that represents the user' interests. Liu *et al.* [19] introduced iExpand, a new CF recommendation system based on user interest expansion via personalized ranking. iExpand uses a three-layer, user–interest–item, representation scheme, which makes the recommendation more accurate and with less computation cost and helps the understanding of the interactions among users, items, and user interests.

Table I shows a comparison between the proposed system and some of the related works presented above. For the convenient manipulation of heterogeneous graphs, some works have used metapaths embedding to represent the network information in lower dimensions, such as metapath2vec [20] and Shi *et al.* [21]. However, in highly dynamic graphs, such as the user–topic–product graph in our case, where the graph update happens very frequently, computing the metapath embedding all over again is very expensive in terms of computation. As we will discuss in the experimental section, our method requires more computational power to compute the initial metapaths compared with the metapath embedding methods but required less computing power for the update operation, which makes it favorable for highly dynamic graphs.

III. SYSTEM DESIGN

In this section, we will present the theoretical framework of the proposed system.

TABLE II
BIG-FIVE TRAITS AND ASSOCIATED CHARACTERS

Personality Trait	Related Characters
Openness to Experience	Artistic, Curious, Imaginative, Insightful, Original, Wide interests
Agreeableness	Trusting, Generous, Appreciative, Kind, Sympathetic, Forgiving
Conscientiousness	Efficient, Organized, Planful, Reliable, Responsible, Thorough
Extraversion	Energetic, Outgoing, Active, Assertive, Talkative
Neuroticism	Anxious, Unstable , Tense, Touchy, Worrying, Self-pitying

A. Big-Five Traits

There are many personality theories that have tried to explain human personality. The most prominent personality theory is known as the five-factor model (FFM) or big-five personality traits. The FFM is based on a common language description of personality, which makes it a compatible model for computing tasks, such as machine learning personality recognition, natural language analysis, and semantic technologies, to name a few. FFM is widely used for different purposes, such as mental disorders diagnosis or job recruitment. The model defines the following five factors: neuroticism, openness to experience, extraversion, agreeableness, and conscientiousness, often denoted by the acronyms OCEAN or CANOE. The big-five factors are shown in Table II along with their related personality facets. Many previous psychological studies have proven the relationship between user's interests and personality traits, such as the relationship between personality traits and Holland's big-six domains of vocational interest (RIASEC) [24] and the relationship between hobby interests and personality traits [25].

The purpose of Meta-Interest is to recommend the most relevant items by detecting the user's topical interests from its social networking data. Fig. 3 shows the general system framework of Meta-Interest. The recommendation process includes five steps. Step 1 is the personality traits' measurement, which can be obtained by asking the user to take a personality measurement questionnaire or using automatic personality recognition by analyzing the subject's social network data. The personality measurement phase is the only static part of the system, which is because personality traits have been proven to be relatively stable over time. Step 2 is mining the user's topical interests, including explicit and implicit interest minings. Explicit interest mining is performed by analyzing the text shared by the user in social networks in order to detect

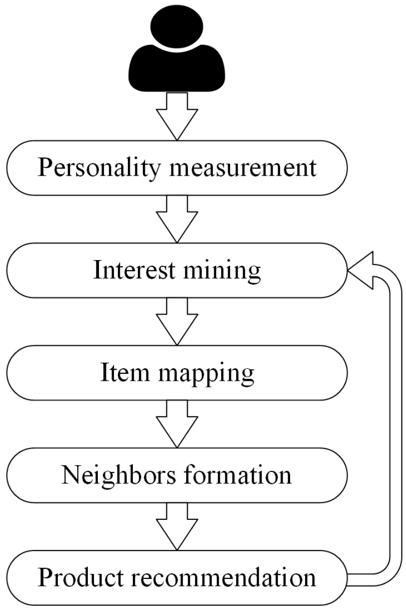


Fig. 3. Meta-Interest recommendations process.

keywords that reflect its topical interests. Implicit interest mining involves a more complex analysis of the social network structure and other latent factors that may influence the user's topical interests. In Step 3, Meta-Interest matches the items with the corresponding topics. The matching is in the form of a many-to-many relationship that is to say that a topic might be related to many items. Similarly, an item might be related to more than one topic. In Step 4, the set of most similar users (neighbors) to the subject user is determined. In this context, Meta-Interest uses three similarity measures, personality similarity, viewing/buying/rating similarity, and common interest similarity. Finally, Step 5 is the item recommendation phase, and the recommendation is refined by updating the neighbors' set and the user's topical interest profile and topics–items matching.

B. Notations

The notations and symbols used in the current work are explained in Table III.

C. Representational Model

Let $U = \{u_1, u_2, \ldots, u_n\}$ be the set of users, $T = \{t_1, t_2, \ldots, t_m\}$ the set of topics, and $P = \{p_1, p_2, \ldots, p_k\}$ the set of all items. The system is modeled as a heterogenous graph that consists of three subgraphs $G = (G_U, G_T, G_P)$, as shown in Fig. 4. $G_U = (V_u, E_u)$ is undirected graph where its node set V_u is the users set U, and the edges set E_u represents the similarity relationship between users. In addition to online behaviors similarity, such as posting and follower/followee similarities, the personality traits' similarity between users is also considered to compute the overall similarity between users. Similarly, the graphs $G_T = (V_t, E_t)$ and $G_P = (V_p, E_p)$ represent the nodes and relationship between topics and items, respectively.

TABLE III
NOTATIONS AND SYMBOLS

Symbol	Meaning
U	The set of all users
u_x	The user x
T	The set of all topics
t_y	The topic t
$\varphi(u_x, u_y)$	The similarity measure between users x and y
$\vartheta(P_x, P_y)$	The similarity measure between item P_x and item P_y
$\overrightarrow{P_x}$	User u_x's personality traits vector
α	User similarity weight parameter
β	Item relatedness weight parameter
Γv	Denotes the set of neighbors of node v
P_l	Meta-path length
w_p	The weight of meta-path P
l_{max}	The maximum length of a meta-path
$\delta^l_{i,j}$	The score between user u_i and item p_j with the meta-path maximum link constrain as $l_{max} = l$
ε	Link prediction score threshold

1) Users' Representation: As mention earlier, one of the most important aspects of the proposed system is that it incorporates the user's personality traits and their related facets to detect the user's interest and eventually in product recommendations. The users' graph $G_U = (V_u, E_u)$ is constructed by measuring the similarity between its vertices. In this regard, we consider three types of similarities: topic interest similarity, product interest similarity, and personality traits' similarity, which we denote as SimT, SimI, and SimP, respectively. Formally, let $U = \{u_1, u_2, \ldots, u_n\}$ be the set of all users and $P_i = \{P_O, P_C, P_E, P_A, P_N\}$ be the big-five personality trait vector of the user u_i; $T_i = \{t_1, t_2, \ldots, t_m\}$ is the set of topical interest of u_i, and $I_i = \{i_1, i_2, \ldots, i_k\}$ is the set of items that were previously viewed by u_i

$$\varphi(u_x, u_y) = \alpha \frac{\sum_i \left(p_x^i - \overline{p_x}\right)\left(p_y^i - \overline{p_y}\right)}{\sqrt{\sum_i \left(p_x^i - \overline{p_x}\right)^2 \sum_i \left(p_y^i - \overline{p_y}\right)^2}} + (1-\alpha)\left(\left\| \frac{2|T_x \bigcap T_y|}{|T_x| + |T_y|} \right\| \left\| \frac{2|I_x \bigcap I_y|}{|I_x| + |I_y|} \right\|\right) \quad (1)$$

where $\overline{p_x}$ and $\overline{p_y}$ is the average value of the personality traits vector for user u_x and u_y, respectively, and p_x^i and p_y^i are the ith trait in the personality traits vector of user u_x and u_y respectively. α is the user similarity weight parameter that tunes the contribution of item-topic similarity and personality similarity in the total similarity measure.

2) Topics Representation: The interests of a given user are represented in the form of a set of topics. The topic space is represented by the graph $G_T = (V_t, E_t)$, where the vertices represent the topics and the edges represent the semantic similarity relationship between these topics. To associate these topics with items graph nodes, each topic node is associated with a category of open directory project (ODP) [26] (see Fig. 5). ODP is a public open directory for web sites' classifications. Currently, it contains 3.8 million websites that have been categorized into 1 031 722 categories by 91 929 human editors. We have used the four-level subcategories to construct the topics graph; these categories are used to match the interest topics with the related items from the item graph.

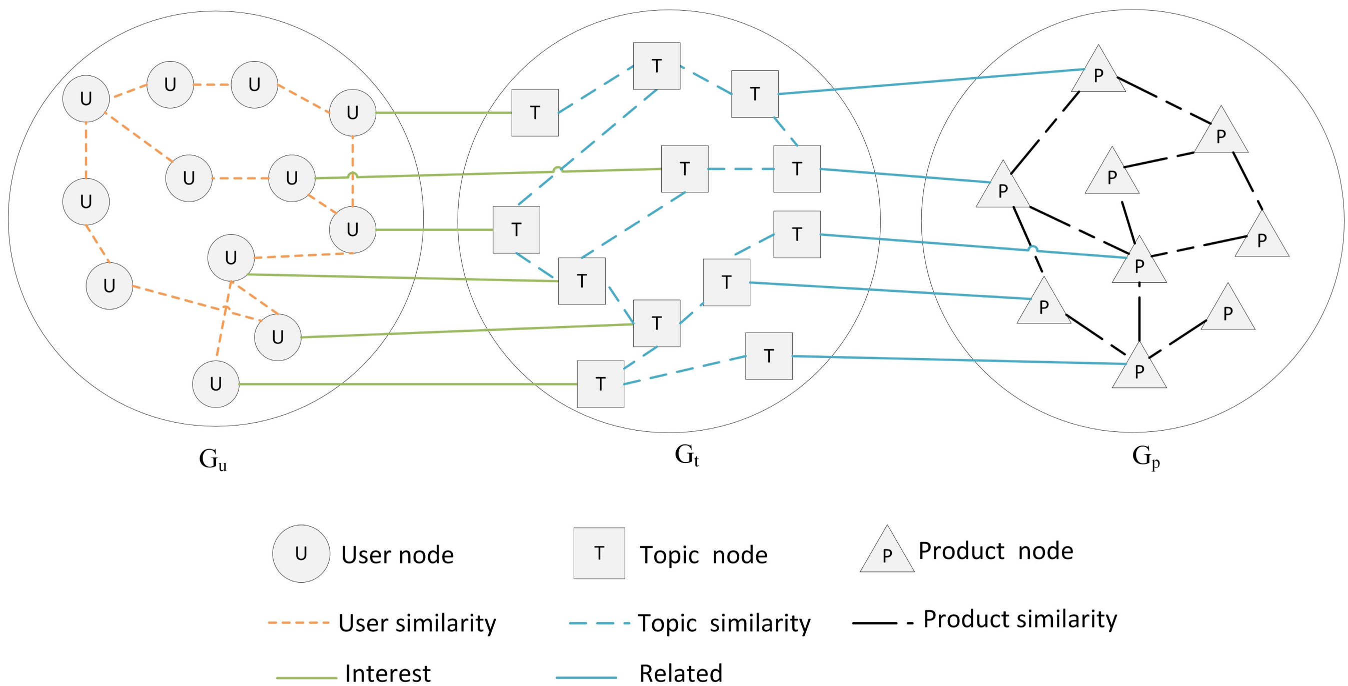


Fig. 4. User–topic–item heterogeneous information network.

Fig. 5. OPD root categories.

3) Item Representation: Similar to the users and interest topics, the items are represented as a graph data structure $G_P = (V_p,\ E_p)$, where the nodes represent the items and the edges represent the similarity between the items. The similarity between items is computed from two similarity measures, content similarity and collaborative similarity. The content similarity is measured by common item's metadata tags, while the collaborative similarity is calculated by measuring the ratio of common buyers/viewers between the two items to the total buyers/viewers of each item. Formally, let $C_x : \{c_0, c_1, \ldots, c_n\}$ and $C_y : \{c_0, c_1, \ldots, c_m\}$ denote the content tags of item P_x and P_y, respectively, and V_x and V_y represent the sets of their viewing/buying users. The similarly between P_x and P_y is computed using the function ϑ, as shown in (2), where β is the item similarity threshold, and it is used to tune the contribution of content similarity and collaborative similarity to the overall similarity measure, $\beta = 0$, when the item has no views and never been bought before (item cold start)

$$\vartheta\left(P_x,\ P_y\right) = \beta\left\|\frac{2|C_x \bigcap C_y|}{|C_x| + |C_y|}\right\| + (1-\beta)\left(\left\|\frac{2|V_x \bigcap V_y|}{|V_x| + |V_y|}\right\|\right). \tag{2}$$

Algorithm 1 Interest_mining

Input u_x, s_x, F_x **Output** I_x

```
if (s_x > CS) then
    Semantic_Annotation(s_x)
    Topics_Extraction(s_x)
else
    for f ∈ F_x do
        I_x ← I_x ∪ {Personality_facet_topics(f)}
    end for
end if
```

D. Interest Mining

The main advantage of our approach is that the proposed system makes use of the user's interests along with the user's personality information to optimize the accuracy of system recommendations and alleviate the cold-start effects. By analyzing the user's social network posted data, we can infer his/her topical interests. The task can be achieved by applying automatic topic extraction techniques, such as latent Dirichlet allocation (LDA) [27] or frequency-inverse category frequency (TFICF) [28]. However, such techniques are supposed to be applied to long articles, and they do not yield good results if applied on the user's short sparse noisy posts, such as tweets [29]. Therefore, to overcome this problem, we have enriched each post from the user's data using semantic annotators, which could help to reduce the noise and alleviate

Algorithm 2 Item_mapping

Input p_z, U_{p_z}
Output I_{p_z}
if $(views(p_z) > CS)$ **then**
  $I_{p_z} \leftarrow OPD\_Topics(p_z)$
else
  for $f \in F_x$ and $u_x \in U_{p_z}$ **do**
    if $(|u_y, f \in F_y| > \frac{|U_{p_z}|}{2})$ **then**
      $I_{p_z} \leftarrow I_{p_z} \cup \{Personality\_facet\_topics(f)\}$
    end if
  end for
end if

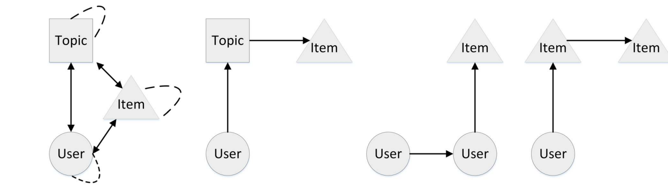
ambiguity of the post and increase the topic detection accuracy, as shown in the proposed framework in [18]. Algorithm 1 shows the pseudocode of interest mining steps. When the user is during the cold-start phase or completely did not view any articles (lines 1–4), Meta-Interest estimates the topical interest based on the interests of users with similar personality facets. Otherwise, it crawls the viewed news articles and extracts the labels of each news article to serve as the topical interest of the user, as we will see in the experimental section.

E. Item Mapping

After populating the topics public space using ODP ontology categories, the items are matched with these topics. Each item is associated with one or more topics and, subsequently, recommended for users that have these topics within their topical interests. Algorithm 2 shows the pseudocode of the item interest mapping process. With newly added items that have not been viewed by any user, the item is directly associated with the corresponding topic category in ODP ontology, whereas items that have passed the cold-start phase are associated with the interest of those that are related to the personality facets that are shared among the users who bought this item.

F. Metapath Discovery

After building the users–topics–items heterogeneous graph $G = (G_U, G_T, G_P)$ that incorporates the users, topics, and items subgraphs and their interrelationships. At this stage, the objective is to predict for a given user the N-most recommended items that match his/her topical interests and previous buying/viewing behaviors. Predicting the users' recommended items is formulated as a graph-based link prediction problem. Link prediction problem has been investigated in many works before, and many schemes have been proven to achieve high accuracy in their predictions, such as Adamic/Adar [30], Katz [31], and Jaccard [32]. However, these schemes are supposed to work on homogeneous graphs where all nodes represent the same type of entities and all the edges are connecting these entities, which is not the case with our heterogeneous graph. Since, in our representation model $G = (G_U, G_T, G_P)$, nodes can represent different entities (users, topics, and items) and the links can connect different nodes (user–user, user–topic, user–item, topic–item, item–item, and topic–topic). We use



(a) Network Scheme (b) Meta-Path U-T-P (c) Meta-Path U-U-P (d) Meta-Path U-P-P

Fig. 6. Network scheme and length 2 metapath samples.

metapaths [21] to predict the matching score between a given user node in G_U and an item node in G_P.

A metapath is a sequence of relations between nodes defined over a heterogeneous network, which can be used to define a topological structure with various semantics. In our case, we investigate the metapaths that start from a user node and end with an item node $P : \{u \rightarrow x \rightarrow \cdots \rightarrow x \rightarrow i\}$. Each metapath is characterized by the number of links between the source and destination nodes, and it is called the path length P_l. For example, the possible metapath with path length P_2 from a user node to an item node is presented in Fig. 6. For a given metapath $P : \{s \rightarrow x \rightarrow \cdots \rightarrow x \rightarrow d\}$, any path in the network that connects nodes s and d following the same intermediate node types as defined by P is called a path instance of P. For a given metapath P, the path count is the number of all path instances $P_c = |\{p : p \in P\}|$. In our case, we consider all metapaths that start with a user node and end with an item node with maximum metapath length to $l_{\max} = 2$. We have made the maximum length to 3 because short metapaths are semantically more important than long ones, and they are good enough for capturing the structure of the network. Besides that, it is computationally expensive to explore longer metapath because the path count increases exponentially with the increase in the path length P_l [33].

By exploring all metapaths with length constraint, we could holistically extract all relationships between nodes with different filtering combinations. We are interested in three types of metapaths: first, the interest metapaths (IP) of the format ⟨U-T-P⟩ [see Fig. 6(b)] that represents metapaths that are based on interest mining and item matching; second, the friendship metapaths (FP) of the format ⟨U-U-P⟩ [see Fig. 6(c)] that represents metapaths that are based on CF (users' similarity); and finally, content metapaths (CP) of the format ⟨U-P-P⟩ [see Fig. 6(d)] that represents metapaths that are based on the content filtering (items similarity). Similarly, by exploring longer metapath, we get more hybrid filtering paths (based on both CF and CBF, in addition to interest mining and item mapping); for example, metapaths of length $P_l = 3$ could be based on CBF (i.e., ⟨U-P-P-P⟩), CF (i.e., ⟨U-U-U-P⟩), hybrid filtering (i.e., ⟨U-U-P-P⟩), or a combination of filtering and interest mining (i.e., ⟨U-T-T-P⟩, ⟨U-T-P-P⟩, and ⟨U-U-T-P⟩).

The importance of each metapath is characterized by its weight w_p. The path weight is computed by the sum of its edges' weight over its length P_l. Formally, let P^n : $\{v_1, v_1, \ldots, v_n\}$ be a metapath with a length of $P_l = n$, and the path weight of P is denoted as w_p, which is the sum of all the links' weights within P, as shown in (3), where $w_{v_i, v_{i+1}}$ represents the weight of link that connects the nodes

Algorithm 3 DiscoverMetaPaths

```
Input u_s, l_max, ε
Output FNL
1: VIST ← Ø
2: P ← Ø
3: FNL ← Ø
4: for i = 1 to l_max do
5:   if (i = 1) then
6:     VIST ← VIST ∪ {u_s}
7:     for NGB ∈ Γu_s do
8:       P ← P ∪ {u_s → NGB}
9:       VIST ← VIST ∪ {NGB}
10:    end for
11:  else
12:    TEMP ← Ø
13:    for CURN ∈ P do
14:      NODE ← p_c[i]
15:      if (NODE = item) and (w_{p_c} > ε) then
16:        FNL ← FNL ∪ {p_c}
17:      end if
18:      if (ΓNODE−VIST ≠ Ø) then
19:        for NGB ∈ ΓNODE−VIST do
20:          TEMP ← TEMP ∪ {CURN → NGB}
21:          VIST ← VIST ∪ {NGB}
22:        end for
23:      end if
24:      P ← P − CURN
25:    end for
26:    P ← TEMP
27:  end if
28: end for
```

v_i and v_{i+1}

$$w_p = \frac{\sum_{i=1}^{n} w_{v_i, v_{i+1}}}{P_l}. \quad (3)$$

In order to predict a possible recommendation for a given user node, we explore all the instances of metapath with a maximum path length $l_{\max} = 3$. Because short metapaths are more semantically significant compared with longer metapaths. Therefore, we prioritize shorter metapath by considering that the contribution of a path weight to the overall link prediction score is inversely proportional to the metapath length P_l. The link prediction score between user u_i and item p_j with the metapath maximum link constrain as $l_{\max} = l$ is computed using (4). To predict the N-most recommended items for a given user, we extract all metapaths by exploring the interest graph with a fixed length and link prediction score constraints

$$\delta_{i,j}^{l} = \sum_{k=2}^{l} \frac{\sum_{r \in P_{i,j}(k)} w_r}{k-1}. \quad (4)$$

To compute the recommended items for a given user, we extract all metapaths instances between the user and potential recommended items by exploring the user–interest–item graph with a fixed length and link prediction score constraints. The pseudocode shown in Algorithm 3 presents the steps of metapath discovery. The algorithm takes as input the user source node u_s, the maximum metapath length to explore $l_{\max}$, and the link prediction score threshold ε. We denote P as the set of the temporarily explored metapaths, P is updated by adding new explored paths or removing dead paths (paths that have no neighbors or paths that do not end with an item node), and FNL is the set of the final metapaths. The set of visited nodes is denoted as VIST, and Γv denotes the set of neighbors of node v. NODE and CURN are temporary variables used to denote the current node and current path respectively in each iteration. In Lines 5–11, a path from the source node u_s to every neighbor node is created and inserted into the set of metapaths P, and node u_s and its neighbors are marked as visited nodes VIST. In lines 13—25, for each path CURN from P, the last node of these paths is visited and added to the final metapaths if it is an item node, and recursively, all the nodes that have not been visited before are added as a potential metapaths. Algorithm 4 shows the pseudocode of recommendation process. Initially, if the user is still in the cold-start phase (lines 2–7), the recommended items are to be filtered based on the topical interests that were extracted from the user's social media data and by associating these topics with the related items according to their OPD categories. Otherwise, the metapaths starting from the source user u_s are discovered and grouped according to the metapath types (interest metapath, friendship metapath, and content metapath), and the items that are in the intersection of these metapaths sets are given propriety in the recommended items' set.

Lines 7–10 enumerate all the neighbors of the source node and lines 13–25 (and eventually lines 19–22) are the primary computational blocks in Algorithm 3. If we study the worst case graph structure, which is a complete graph (fully connected graph), where every user is similar to all other users and interested in all topics and also connected to all the available products (even though, in this case, we do not have to run Algorithm 3, there is no unknown link that we need to predict). Algorithm 3 still runs in linear time complexity. Let G be a complete graph (fully connected) with n nodes and $n = x + y + z$ (x: user nodes; y: topic nodes; and z: product node). The run time of the block (lines 7–10) is $O(x+y+z-1)$ to add all the graph nodes to the visited nodes group VIST and their generated paths to P. The block (12–25) also runs in linear time of $O(x + y + z - 1)$ as well; Even, it includes a nested loop (lines 19–22) that could result a quadratic time, lines 19–22 will never be reached due to the if-condition block in lines 18–23 (as the studied graph is a complete graph, and VIST will contain all the graph nodes [added in block (7–10)]; therefore, $\Gamma\mathrm{NODE} - \mathrm{VIST} = \emptyset$). Hence, the overall time complexity of Algorithm 3 is $O(n)$.

IV. System Evaluation

In this section, we present the details of the collected data set, evaluation metrics and baselines, and the analysis of the obtained results.

A. Baselines

To evaluate the performance of the proposed product recommendation system, we have compared it with different

Algorithm 4 RecommendProducts

Input u_s, l_s
Output R
$R \leftarrow \emptyset$
if $(CS(u_s))$ **then**
  for $t \in I_s$ **do**
    $PR \leftarrow Product\_interest(t)$
    $R \leftarrow R \cup PR$
  end for
else
  $P = DiscoverMetaPath(u_s)$
  $IP = InterestPaths(P)$
  $FP = FriendPaths(P)$
  $CP = ContentPaths(P)$
  $RecPaths = TopNPaths(IP \cap FP \cap CP, FP \cap CP, CP \cap IP)$
  for $Path \in RecPaths$ **do**
    $PR \leftarrow Path[last_node]$
    $R \leftarrow R \cup PR$
  end for
end if

baselines that use various recommendation techniques, such as deep learning, metapath analysis, network embedding, and session-based. The proposed system is compared with the following baselines:

1) GNN-SEAL (Graph Neural Networks) [22]: GNN-SEAL is a link prediction framework that formulates link prediction problem as a subgraph classification problem. For every predicted link (user–item link in our case), GNN-SEAL determines its h-hop enclosing subgraph A and computes its node information matrix X (which contains structural labels, latent embeddings, and the explicit attributes of nodes). After that, the framework feeds (A, X) into a graph neural network (GNN) to classify the link existence so that it can learn from both graph structure features (from A) and latent/explicit features (from X) simultaneously for link prediction. The framework is open source, and the code is available on GitHub.[1]

2) metapath2vec (Metapath and Network Embedding) [20]: metapath2vec formalizes metapath-based random walks to build the heterogeneous neighborhood of a node and then uses the heterogeneous skip-gram model to perform node embeddings and, subsequently, user–item link prediction. metapath2vec is open source, and its implementation code is available on Github.[2]

3) DGRec (Session-Based) [23]: DGRec is a session-based recommendations' framework that employs dynamic-graph-attention neural network to model the context-dependent social influence and recurrent neural network to model dynamic user interest. Finally, DGRec gives the recommendation by merging the user's interests and preferences and his/her social influence. DGRec is open source, and its implementation code is available on Github.[3]

4) LightFM (Cold Start) [34]: LightFM is a cold-start alleviation framework that uses a hybrid matrix factorization model to represent items (products in our case) and users as linear combinations of their content features' latent factors. LightFM is parameterized in terms of d-dimensional user and item feature embeddings e U f and e I f for each feature f. Every feature is also modeled by a scalar bias term (b U f for user and b I f for item features). The model's prediction for user u and item i is then given by the dot product of user and item representations, adjusted by the user and item feature biases. LightFM is open source, and the implementation code is available on GitHub.[4]

5) CF-CBF: This is the hybrid filtering system that combines the users' viewing similarity and product similarity to determine the neighborhood set and recommends new items.

B. Evaluation Metrics

Any product recommendation system is evaluated by measuring the accuracy and coverage of its recommended items. To test the efficiency of Meta-Interest and compare it to the afore-mentioned baselines, we determine the recommended items by each baseline, display it in the user's feed along with other irrelevant items, and measure the accuracy rate of the relevant items. Formally, let $F = R \cup I$ be the set that represents all items in user u's feeds, where $R = \{p_1, p_2, \ldots, p_r\}$ is the set relevant items, and $I = \{p_1, p_2, \ldots, p_i\}$ is the set of irrelevant items. After showing F in user u's feeds, we denote $V = \{p_1, p_2, \ldots, p_v\}$ as the set of viewed items. In this context, we are interested in the following values: 1) true positives: the group of relevant items that have been viewed by the user TP $= \{x \ / \ x \in R \bigcap V\}$; 2) false positives: the group of irrelevant items that have been viewed by the user FP $= \{x \ / \ x \in I \bigcap V\}$; and 3) false negatives: the group of relevant items that have not been viewed by the user FN $= \{x \ / \ x \in R, \ x \notin V\}$. We have used the following metrics.

Precision: The portion of relevant viewed items in the total viewed items, and it is computed as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{5}$$

Recall: The portion of relevant viewed items in the total relevant items, and it is computed as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{6}$$

F-Measure: It is also called the balanced F-Score; it is the harmonic average of the precision and recall; and it is computed as follows:

$$F = \frac{2 \ P \ R}{P + R}. \tag{7}$$

[1] github.com/muhanzhang/SEAL
[2] ericdongyx.github.io/metapath2vec/m2v.html
[3] github.com/DeepGraphLearning/RecommenderSystems/tree/master/socialRec
[4] github.com/lyst/lightfm

TABLE IV
DATA SET STATISTICS

Parameter	Value
Number of users	2228
Number of articles	25873
Number of items	6230
Cold start users	575
Cold start items	1520

C. Data Set Description

We have integrated the Meta-Interest product recommendation system with a social network platform called Newsfulness[5] that we have implemented earlier for automatic personality recognition projects. Newsfulness enables the user to view and shares news articles from various news publishers. During registration, the users go through the TIPI Big-Five personality questionnaire [35] to capture their personality traits. Newsfullness collects published articles from different English-speaking news websites, and the collected articles are from the following outlets (BBC, CNN, Aljazeera, France24, Russia-Today, Reuters, The Guardian, and The New York Times). The gathered articles are from all the news classes (politics, business, sports, health, travel, education, entertainment, art, science, and technology) from different geographic regions categories. The products' recommendation system was implemented by fetching products from different online stores (mainly JD, Banggood, and Amazon). The statistical details of the used data set are presented in Table IV.

D. Result Discussion

To tune the optimal value of the users' similarity parameter α and products' similarity parameter β, we observe the optimal value of α and β that maximize the F-Measure of the proposed system. Figs. 7 and 8 show the optimal value of α and β in different topics of interest count and viewed items count, respectively. As we can observe from Fig. 7, during the cold-start phase with no topic of interest at all, $\alpha = 1$, and at this point, the users' similarity is based only on personality similarity measurement. With the increase in previously detected topics of interest, the value of α gradually decreases and finally stabilizes with $\alpha = 0.2$ when the user passes the cold-start phase and had enough topical interest and previously viewed items. Similarly, the optimal value of β during the cold-start phase for the new item with no views is $\beta = 1$, and with the increase in the number of views, the value of β decreases to finally stabilize with $\beta = 0.5$, as shown in Fig. 8. For the size of Top-n recommended products, in our experiment, we set $N = 20$, as choosing a larger value will lead to uncertainty of whether the users did not view the products' feed because they are not interested in them, or they did not view them because there are too many items in the products' feed. If we ignore this uncertainty and just consider that the user did not view the product out of his disinterest, this will lead to an increase in false positives and false negatives as well, hence the decrease in the overall system performance. As we can observe from Fig. 9, the

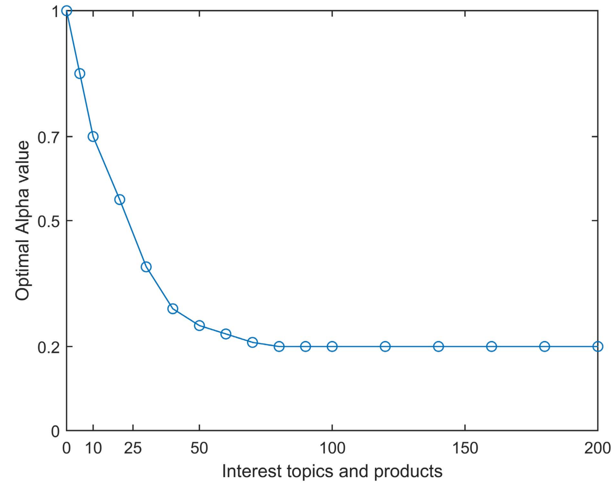


Fig. 7. Users' similarity parameter tuning.

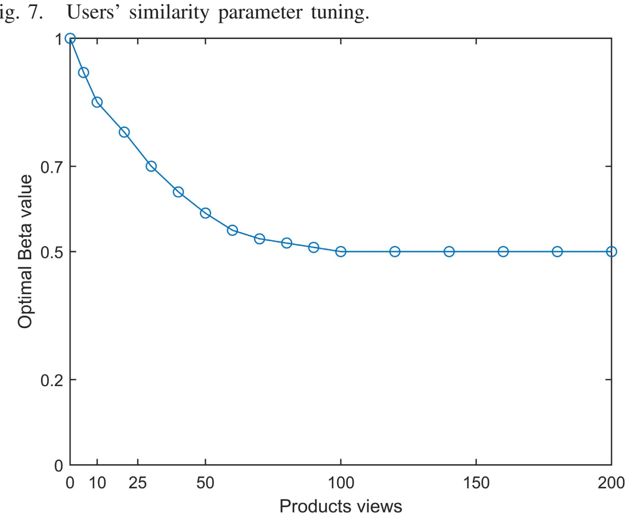


Fig. 8. Products' similarity parameter tuning.

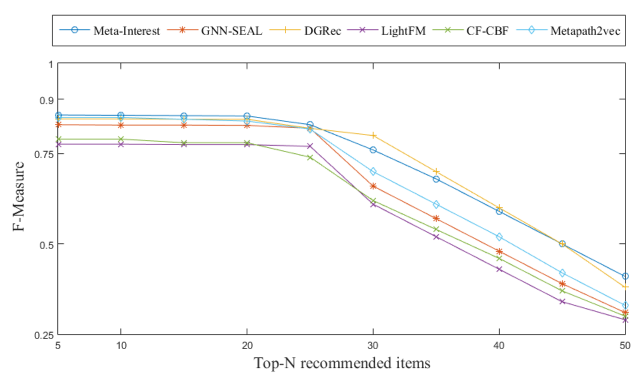


Fig. 9. Top-n recommendation parameter tuning.

F-Measure of the proposed system and all the studied baselines have decreased dramatically when the value of N is over 22.

The precision, recall, and F-measure of Meta-Interest compared with the baseline schemes are shown in Fig. 10. As we can observe, the proposed system, Meta-Interest, and the session-based system, DGRec, clearly have the highest precision (0.854 and 0.845) and recall (0.868 and 0.855), respectively. The superiority of the proposed system is because of the personality biased approach that filters the relevant

[5]www.newsfullness.live/data set

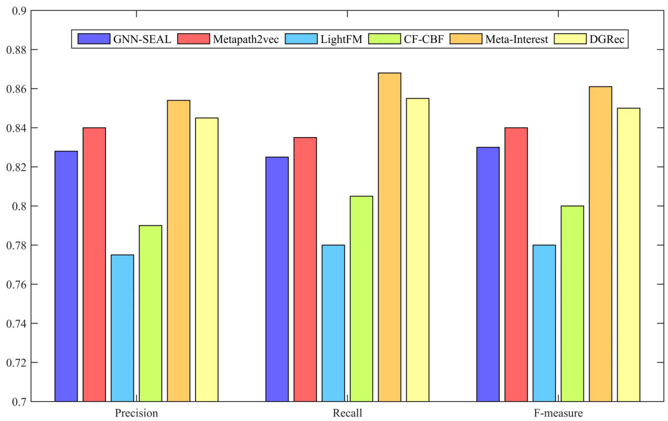


Fig. 10. Overall system evaluation.

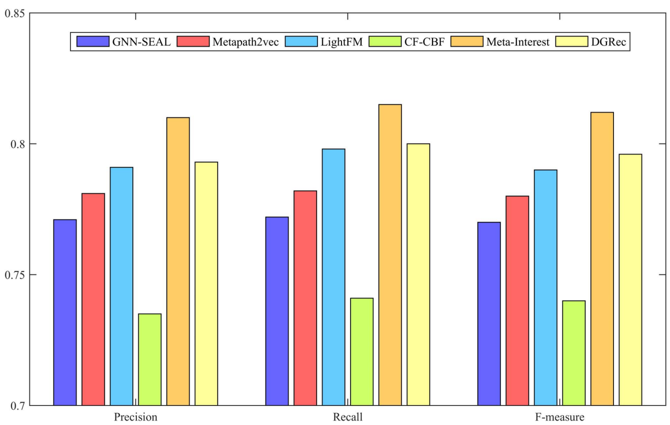


Fig. 11. System evaluation under cold start (new users).

items that are related to the personality facets of the user, while other approaches view the user's personality traits just as additional information that helps find the similarity and construct the network embeddings or features. The second reason for the superiority of Meta-Interest (and also DGRec compared with other baselines) is the ability of Meta-Interest and DGRec to alleviate the cold-start effects, and hence maintain stable precision, and recall values all over the phases. Unlike the network representation method, metapath2vec, and the deep-learning method, GNN-SEAL, that come third and fourth with 0.84 and 0.828 of precision value and 0.835 and 0.825 of recall value, respectively, LightFM performs quite well in the cold-start phase (as we will see later in other figures); however, it fails to cope with a large amount of diverse data in later stages, which leads to a drop in its precision and recall values.

One of the main reasons for incorporating the user's personality in the product recommendation systems and interest mining schemes is to alleviate the effects of the cold-start problem [6]. In this regard, we have tested the performance of Meta-Interest and the studied baselines under the cold-start settings. The cold-start settings include two tests.

1) Cold-start users' test, in which only the new users are considered in the precision and recall measurements. In our experiment, a user is considered in the cold-start phase if the number of viewed articles and items is less than 20.

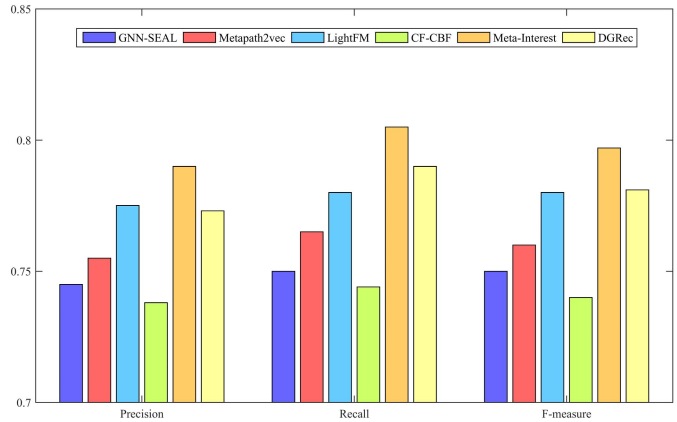


Fig. 12. System evaluation under cold start (new items).

2) Cold-start items' test, in which we consider only the new items that have not been viewed or rated by any user. Figs. 11 and 12 show the results of the cold-start users' test and cold-start items' test, respectively.

As we can observe, the proposed system and the session-based system, DGRec, still have the upper hand even in the cold-start phase as both systems are robust in cold-start settings, as explained early. However, we can notice that the LightFM is ranked third and obviously outperforms metapath2vec and GNN-SEAL because LightFM was originally designed to mitigate the cold-start effects, and as mentioned early, LightFM has a poor performance when the amount of the data increases. To further study the relationship between the amount of available data and the performance of the proposed system compared with the baselines, we measure the performance of Meta-Interest and the other baseline systems while changing the percentage of training set size from 10% to 100%. Fig. 13 shows the precision, recall, and F-measure values of the studied systems with different training set sizes. We can clearly observe that Meta-Interest outperforms the other baselines with only a small training set size, with only 10% of the training set and Meta-Interest scores 0.768 and 0.765 in precision and recall, respectively. With the increase in training set size, Meta-Interest steadily improves to reach 0.854 and 0.868 in precision and recall using 100% of the training set (around 10.07% improvement compared with 10% training). LightFM ranks second in terms of precision and recall with 10% of the training set; however, it ends in the fifth place (better only than the conventional CF-CBF) with full training set 100%, whereas deep-learning-based schemes (GNN-SEAL) and network embeddings approach (metapath2vec) have a low performance with small training data. When trained with 10% of the data set, GNN-SEAL scores 0.63 and 0.64 in precision and recall, respectively. However, GNN-SEAL and metapath2vec performance increase dramatically with the increase in the training data size. For instance, when trained with the full training data, metapath2vec scores 0.84 and 0.835 (around 23% improvement compared with 10% training). That is because metapath2vec uses network embedding, which requires the presence of dense node links to capture the network structure.

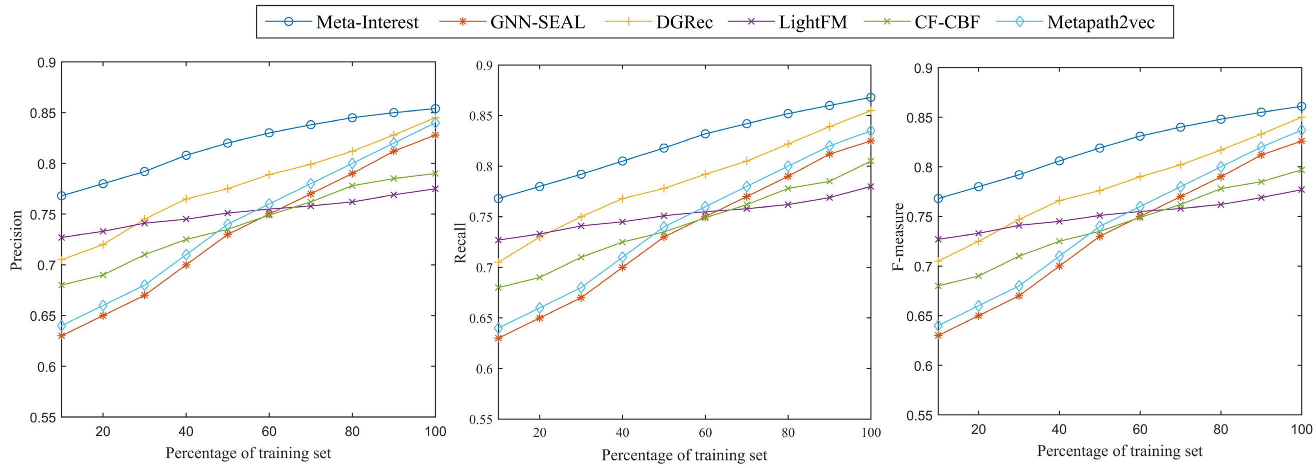


Fig. 13. System evaluation with different sizes of the training set.

TABLE V
SPEED COMPARISON (s)

System	Total computational time	Update computational time
Meta-Interest	16.05	3.56
metapath2vec	14.6	14.6
GNN-SEAL	15.5	15.5
DGRec	16.2	4.66
LightFM	19.04	5.92
CF-CBF	16.15	5.41

In a practical situation with a large graph of millions of nodes and links that require intensive computational power, the speed of the recommendation system is crucial to keep a reasonable response time. Therefore, it is important to analyze the speed and time complexity of the proposed system compared with the compared baselines. Table V shows the time complexity of the proposed system compared with the studied baselines. The shown values in Table V are the average of 100 times testing. The time complexity of Meta-Interest and all the baselines were tested on Dell Inspiron 173000 Laptop, with tenth-generation Intel Core i7-1065G7 Processor (8-MB Cache, up to 3.9 GHz) and 16-GB RAM (2 × 8 GB, DDR4, 2666 MHz), running Ubuntu 19.04 operating system. As we can observe from Table V, when it comes to the total computational time required for the system to compute the recommendation for all users, Meta-Interest is not the fastest system, metapath2vec has the lowest computational time of 14.6 s, and GNN-SEAL ranks second with 15.5. However, for the update operation where we add a new block of users and items and compute the time required for the system to compute the recommendation of these new users, metapath2vec needs to recalculate the network embeddings in lower dimensional space all over again, which costs as high as the initial time required to compute all the recommendations, which is not the case with Meta-Interest, as we just need to recalculate the weights of the newly added metapaths.

V. CONCLUSION

In this article, we have proposed a personality-aware product recommendation system based on interest mining and metapath discovery, and the system predicts the user's needs and the associated items. Products' recommendation is computed by analyzing the user's topical interest and, eventually, recommending the items associated with those interests. The proposed system is personality-aware from two aspects: first, because it incorporates the user's personality traits to predict his topics of interest; second, it matches the user's personality facets with the associated items. Experimental results show that the proposed system outperforms the state-of-art schemes in terms of precision and recall especially in the cold-start phase for new items and users.

However, Meta-Interest could be improved in different aspects.

1) In this work, the users' personality traits' measurement was conducted through questionnaires. Integrating an automatic personality recognition system, which can detect the users' personality traits based on their shared data, into Meta-Interest is one of our future directions.
2) The proposed system uses big-five to model the user's personality. Extending Meta-Interest to include other personality traits models, such as the Myers–Briggs type indicator, is a future direction.
3) The proposed system could be further improved by integrating a knowledge graph and infer topic–item association using semantic reasoning.

ACKNOWLEDGMENT

The authors would like to thank all the active users of Newsfullness that agreed to be a part of the Meta-Interest experiment.

REFERENCES

[1] G. Piao and J. G. Breslin, "Inferring user interests in microblogging social networks: A survey," *User Model. User-Adapted Interact.*, vol. 28, no. 3, pp. 277–329, Aug. 2018. [Online]. Available: http://link.springer.com/10.1007/s11257-018-9207-8

[2] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 273–291, Jul. 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6834774/

[3] V. Martínez, F. Berzal, and J.-C. Cubero, "A survey of link prediction in complex networks," *ACM Comput. Surv.*, vol. 49, no. 4, pp. 1–33, Feb. 2017.

[4] H.-C. Yang and Z.-R. Huang, "Mining personality traits from social messages for game recommender systems," *Knowl.-Based Syst.*, vol. 165, pp. 157–168, Feb. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S095070511830577X

[5] W. Wu, L. Chen, and Y. Zhao, "Personalizing recommendation diversity based on user personality," *User Model. User-Adapted Interact.*, vol. 28, no. 3, pp. 237–276, Aug. 2018.

[6] H. Ning, S. Dhelim, and N. Aung, "PersoNet: Friend recommendation system based on big-five personality traits and hybrid filtering," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 3, pp. 394–402, Jun. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8675299/

[7] B. Ferwerda, M. Tkalcic, and M. Schedl, "Personality traits and music genres: What do people prefer to listen to?" in *Proc. 25th Conf. User Model., Adaptation Personalization*, Jul. 2017, pp. 285–288.

[8] B. Ferwerda, E. Yang, M. Schedl, and M. Tkalcic, "Personality and taxonomy preferences, and the influence of category choice on the user experience for music streaming services," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 20157–20190, 2019.

[9] Z. Y. Hafshejani, M. Kaedi, and A. Fatemi, "Improving sparsity and new user problems in collaborative filtering by clustering the personality factors," *Electron. Commerce Res.*, vol. 18, no. 4, pp. 813–836, Dec. 2018. [Online]. Available: http://link.springer.com/10.1007/s10660-018-9287-x

[10] S. Dhelim, N. Huansheng, S. Cui, M. Jianhua, R. Huang, and K. I.-K. Wang, "Cyberentity and its consistency in the cyber-physical-social-thinking hyperspace," *Comput. Electr. Eng.*, vol. 81, Jan. 2020, Art. no. 106506. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0045790618334839

[11] A. Khelloufi *et al.*, "A social relationships based service recommendation system for SIoT devices," *IEEE Internet Things J.*, early access, Aug. 14, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9167284/, doi: 10.1109/JIOT.2020.3016659.

[12] F. Zarrinkalam, M. Kahani, and E. Bagheri, "Mining user interests over active topics on social networks," *Inf. Process. Manage.*, vol. 54, no. 2, pp. 339–357, Mar. 2018.

[13] A. K. Trikha, F. Zarrinkalam, and E. Bagheri, "Topic-association mining for user interest detection," in *Proc. Eur. Conf. Inf. Retr.* Basel, Switzerland: Springer, 2018, pp. 665–671.

[14] J. Wang, W. X. Zhao, Y. He, and X. Li, "Infer user interests via link structure regularization," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 2, p. 23, 2014.

[15] S. Dhelim, H. Ning, M. A. Bouras, and J. Ma, "Cyber-enabled human-centric smart home architecture," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct. 2018, pp. 1880–1886. [Online]. Available: https://ieeexplore.ieee.org/document/8560294/

[16] S. Faralli, G. Stilo, and P. Velardi, "Automatic acquisition of a taxonomy of microblogs users' interests," *J. Web Semantics*, vol. 45, pp. 23–40, Aug. 2017.

[17] S. Dhelim, N. Aung, and H. Ning, "Mining user interest based on personality-aware hybrid filtering in social networks," *Knowl.-Based Syst.*, vol. 206, Oct. 2020, Art. no. 106227. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0950705120304354

[18] J. Kang and H. Lee, "Modeling user interest in social media using news media and wikipedia," *Inf. Syst.*, vol. 65, pp. 52–64, Apr. 2017.

[19] Q. Liu, E. Chen, H. Xiong, C. H. Q. Ding, and J. Chen, "Enhancing collaborative filtering by user interest expansion via personalized ranking," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 42, no. 1, pp. 218–233, Feb. 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6006538/

[20] Y. Dong, N. V. Chawla, and A. Swami, "Metapath2vec: Scalable representation learning for heterogeneous networks," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: ACM, Aug. 2017, pp. 135–144. [Online]. Available: https://dl.acm.org/doi/10.1145/3097983.3098036

[21] C. Shi, B. Hu, W. X. Zhao, and P. S. Yu, "Heterogeneous information network embedding for recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 2, pp. 357–370, Feb. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8355676/

[22] M. Zhang and Y. Chen, "Link prediction based on graph neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 5165–5175.

[23] W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang, "Session-based social recommendation via dynamic graph attention networks," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*. New York, NY, USA: ACM, Jan. 2019, pp. 555–563. [Online]. Available: https://dl.acm.org/doi/10.1145/3289600.3290989

[24] P. I. Armstrong and S. F. Anthoney, "Personality facets and RIASEC interests: An integrated model," *J. Vocational Behav.*, vol. 75, no. 3, pp. 346–359, Dec. 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0001879109000657

[25] U. Wolfradt and J. E. Pretz, "Individual differences in creativity: Personality, story writing, and hobbies," *Eur. J. Personality*, vol. 15, no. 4, pp. 297–310, 2001.

[26] J.-H. Lee, J. Ha, J.-Y. Jung, and S. Lee, "Semantic contextual advertising based on the open directory project," *ACM Trans. Web*, vol. 7, no. 4, p. 24, 2013.

[27] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[28] J. Han and H. Lee, "Characterizing user interest using heterogeneous media," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 289–290.

[29] N. Naveed, T. Gottron, J. Kunegis, and A. C. Alhadi, "Searching microblogs: Coping with sparsity and document quality," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 183–188.

[30] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Soc. Netw.*, vol. 25, no. 3, pp. 211–230, 2003.

[31] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[32] S. Dhelim, N. Aung, and H. Ning, "ComPath: User interest mining in heterogeneous signed social networks for Internet of people," *IEEE Internet Things J.*, early access, Nov. 10, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9253614, doi: 10.1109/JIOT.2020.3037109.

[33] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "PathSim: Meta path-based top-K similarity search in heterogeneous information networks," *Proc. VLDB Endowment*, vol. 4, no. 11, pp. 992–1003, Aug. 2011.

[34] M. Kula, "Metadata embeddings for user and item cold-start recommendations," in *Proc. 2nd Workshop New Trends Content-Based Recommender Syst. Co-Located 9th ACM Conf. Recommender Syst. (RecSys)*, Vienna, Austria, vol. 1448, 2015, pp. 14–21. [Online]. Available: http://ceur-ws.org/Vol-1448/paper4.pdf

[35] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," *J. Res. Pers.*, vol. 37, no. 6, pp. 504–528, 2003.

Sahraoui Dhelim (Member, IEEE) received the B.S. degree in computer science from the University of Djelfa, Djelfa, Algeria, in 2012, the master's degree in networking and distributed systems from the University of Laghouat, Laghouat, Algeria, in 2014, and the Ph.D. degree in computer science and technology from the University of Science and Technology Beijing, Beijing, China, in 2020.

His current research interests include social computing, personality computing, user modeling, interest mining, recommendation systems, and intelligent transportation systems.

Huansheng Ning (Senior Member, IEEE) received the B.S. degree from Anhui University, Hefei, China, in 1996, and the Ph.D. degree from Beihang University, Beijing, China, in 2001. He is currently a Professor and the Vice Dean of the School of Computer and Communication Engineering, University of Science and Technology, Beijing.

He is the Founder and the Chair of the Cyberspace and Cybermatics International Science and Technology Cooperation Base. He has presided many research projects, including the Natural Science Foundation of China and the National High Technology Research and Development Program of China (863 Project). He has authored or coauthored more than 100+ journal/conference articles and authored five books. His current research focuses on the Internet of Things and general cyberspace.

Dr. Ning serves as an Associate Editor for the IEEE Systems Journal (2013-currently), the IEEE Internet of Things Journal from 2014 to 2018, and as a Steering Committee Member of the IEEE Internet of Things Journal (2016-currently).

Nyothiri Aung received the M.E. degree in information technology from Mandalay Technological University, Mandalay, Myanmar, 2012, and the Ph.D. degree in computer science and technology from the University of Science and Technology Beijing, Beijing, China, in 2020.

She worked as tutors with the Department of Information Technology, Technological University of Meiktila, Meiktila, Myanmar, from 2008 to 2010 and a System Analyst of ACE Data System, Yangon, Myanmar, from 2012 to 2015. Her research interests include social computing, personality computing, and intelligent transportation systems.

Runhe Huang (Senior Member, IEEE) received the Ph.D. degree in computer science and mathematics from the University of the West of England, Bristol, U.K., in 1993.

She is currently a Full Professor with the Faculty of Computer and Information Sciences, Hosei University, Japan. She received a Sino-Britain Friendship Scholarship for her study in the U.K. Her research fields include cognitive computing, brain modeling, computational intelligence computing, big data, machine learning.

Jianhua Ma (Member, IEEE) received the B.S. and M.S. degrees from the National University of Defense Technology, Changsha, China, in 1982 and 1985, respectively, and the Ph.D. degree from Xidian University, Xi'an, China, in 1990.

He is currently a Full Professor with the Faculty of Computer and Information Sciences, Hosei University, Japan.

Dr. Ma is an Editor-in-Chief of *Journal of Ubiquitous Computing and Intelligence (JUCI)*, *Journal of Mobile Multimedia (JMM)*, *Journal of Autonomic and Trusted Computing (JoATC)*, and *International Journal of u- and e-Service, Science and Technology (IJUNESST)*, and an Assistant Editor-in-Chief of *International Journal of Pervasive Computing and Communications (JPCC)*. He is an Area Editor of the *International Journal of Computer Processing of Oriental Languages (JCPOL)*, an Associated Editor of the *International Journal of Distance Education Technologies (JDET)* and *Journal of Human-centric Computing (HCIS)* (Springer), on the Advisory Board of the *International Journal of Smart Home (IJSH)*, and on the Editorial Board of the *International Journal of Wireless and Mobile Computing (IJWMC)*, *International Journal of Security and Its Applications (IJSIA)*, *International Journal of Database Theory and Application (IJDTA)*, *International Journal of Computing & Information Technology (IJCIT)*, and *International Journal On Advances in Software*. He is the Founder Chair of the IEEE CIS Smart World Technical Committee and the IEEE SMC Cybermatics Technical Committee.