<div align="center">RESEARCH ARTICLE</div>

A PERSONALITY-BASED AGGREGATION TECHNIQUE FOR GROUP RECOMMENDATION

Emre YALCIN [1, 2, *] , Alper BILGE [3]

[1] Department of Computer Engineering, Faculty of Engineering, Eskisehir Technical University, Eskisehir, Turkey
[2] Department of Computer Engineering, Faculty of Engineering, Sivas Cumhuriyet University, Sivas, Turkey
[3] Department of Computer Engineering, Faculty of Engineering, Akdeniz University, Antalya, Turkey

ABSTRACT

The main goal of a group recommender system is to provide appropriate referrals to a group of users sharing common interests rather than individuals. Such group referrals are commonly produced by utilizing aggregation techniques that analyze the propensities of the whole group by combining the preferences of the users in the group. Although there exist various aggregation techniques in the literature, they usually rely on the assumption that each member of the group has equal importance on the final decision of the group. However, the decision-making process of a group is a complicated process that is strongly correlated with not only group members' experience about the domain of interest but also their behavioral aspects; therefore, the influence of the individuals might be dependent on user personalities. In this study, we propose a personality-aware aggregation technique termed as the *Personality weighted Average (PwAvg)*, which determines the influence degree of each member in the group using five fundamental personality traits, *openness*, *agreeableness*, *emotional stability*, *conscientiousness*, and *extraversion*; and then utilizes them to weight the preferences during the aggregation process. Experiments performed on two real-world benchmark datasets demonstrate that the PwAvg technique significantly outperforms three baseline aggregation techniques, especially for large user groups. Empirical outcomes also show that utilizing the PwAvg with emotional stability trait achieves more qualified group recommendations compared to others.

Keywords: Group recommender system, Aggregation technique, Personality traits

1. INTRODUCTION

With the prevalent usage of the Internet, a plethora of online services realizing people to perform their daily activities (e.g., booking a hotel, listening to music, watching movies) have been developed. Nevertheless, the rapid expansion of such systems causes difficulty for individuals to find useful/relevant information within a vast amount of data sources, which is referred to as the *information overload* problem [1]. Recommender systems are automated tools that overcome this problem by locating engaging content while filtering out irrelevant information [2].

Recommender systems traditionally aim to suggest appropriate products/services to individual users by considering their interests and preferences [2]. Here, the main goal is to recommend items fulfilling individuals to the maximum extent. However, people show a tendency to perform certain activities together with a group of users rather than doing alone e.g., watching a movie with family members [3] and going to a restaurant for lunch with colleagues [4]. Additionally, people have to act together with a community in some circumstances, such as taking a trip with a tour [5-7], working out in a fitness center [8], and using public transportation. Such scenarios require more complicated recommendation mechanisms since now the target audience is not a single person, but a group of people getting together for the same reasons.

Group recommender systems (GRSs) are introduced to handle the issue of recommending appropriate content appealing to a group of users sharing common interests instead of individuals [9]. Recently, various GRSs in several different domains (e.g., music [8, 10-12], touristic attractions [5-7], TV programs [3], movies [13-17], and restaurants [4]) have been proposed. The main goal of these systems is to produce useful information for group members by examining the characteristics and tendencies of a group. In doing so, such systems combine the preferences of group members by utilizing methods termed as aggregation techniques to construct a group profile that represents the preferences of the group [18, 19]. These techniques are crucial components for determining appropriate group preferences, hence directly influence the quality of provided group recommendations.

There exist several aggregation techniques for providing group recommendations [18]. The vast majority of these techniques generate group profiles by considering the different aspects of the ratings provided by group members. In other words, group profiles are generally determined based on high-average [5, 7, 16, 17], frequency of rating counts [18, 19], rankings [20], or highest/lowest ratings [13, 21]. Although such techniques can provide reasonable recommendation results, they usually disregard the psychological aspects of members in the group during the aggregation process. However, it is a known phenomenon that the decision-making is related to the emotions and personalities of individuals, as users with similar personal characteristics are more likely to have similar preferences/interests [22, 23].

Also, the existing aggregation techniques usually rely on the assumption that each member of a group has similar impacts on the final decision of the group. However, intuitively, individuals might have different influences in decision making in a community, and the degree of influence of a member in the group might be dependent on the personality traits [24]. For example, there can be particular members in the group who are persistent in their thoughts as they do not want to lose their utility, or as they believe that their own choices are ideal decisions for every member of the group. On the other hand, other types of people can be worried about the satisfaction of all the other members, at the cost of the personal one; thus, they are willing to reduce some utility to reach a suitable agreement for the entire group. In order to consider such different behavioral features of the individuals while producing group recommendations, it is required to study the personality traits of individuals via models in the area of the human sciences. One of the most used models in this sense is the big five-factor [25], which describes human personality through the following five fundamental traits: openness, agreeableness, emotional stability, conscientiousness, and extraversion.

In this study, we propose a novel aggregation technique termed as Personality weighted Average (*PwAvg*) that takes into account all personality traits stated above to provide high-quality group recommendations. More specifically, the PwAvg measures the influence of members in a group by using their personalities and employ them to weight their preferences in the phase of aggregating.

The rest of the study is organized, as follows: The following section explains the prominent aggregation techniques in the literature. Section 3 gives a brief literature summary of well-known GRSs. Section 4 presents the proposed group recommendation framework, including a novel aggregation mechanism, and the following section demonstrates experimental work and empirical outcomes. Finally, Section 6 concludes the study and explains future research directions.

2. AGGREGATION TECHNIQUES

As emphasized in the Introduction, group recommendations are produced based on group preferences estimated by utilizing an appropriate aggregation technique. In the literature, there exist various aggregation techniques with different characteristics to construct group profiles to be used for providing referrals to groups [18, 19]. However, we consider the most prominent ones, including Average [5, 7,

16, 17]**,** Average without Misery [10], and Least Misery [13, 16, 17]; to investigate the performance of the technique proposed in this study. In the following, we attempt to describe how these techniques aggregate the ratings provided by group members for calculating a group rating.

- **Average (*Avg*):** This technique provides a consensus among group members by considering the ratings of all users of the group. It estimates a group rating for an item by merely calculating averages of group members' preferences for the corresponding item. Also, this technique relies on the assumption that the influence degree of each member of the group is equal.
- **Average without Misery (*AwM*):** This technique is similar to the Avg method; however, the AwM ignores items having any rating below the predefined threshold. In other words, it only calculates an average for items whose all provided ratings are above the threshold.
- **Least Misery (*LM*):** This technique considers only the minimum rating among the ratings provided by group members in calculating group ratings. In other words, the LM technique sets the lowest rating in a group as the group preference.

To show how the Avg, AwM, and LM techniques work, we also provide a user-item matrix in Table 1, which presents recommendations for five items to a group with four members on a 10-star scale. Here, the unrated items are denoted with (-), and the threshold for the AwM technique is selected as 6.

Table 1. Group ratings calculated by the Avg, AwM, and LM

		i_1	i_2	i_3	i_4	i_5
	u_1	7	10	-	3	8
	u_2	8	-	10	-	6
	u_3	-	7	8	4	7
	u_4	8	9	5	1	3
Group rating	*Avg*	7.66	8.66	7.66	2.66	6
	AwM	7.66	8.66	-	-	-
	LM	7	7	5	1	3

3. RELATED WORK

Over the past two decades, various GRSs have been introduced for different scenarios in several domains, such as music [8, 10-12], movies [13-17], restaurants [4], tourism [5-7], and so on. However, the utilized methods to combine individual preferences in these systems differ, as there is no single aggregation technique that can achieve high-performance in all applications [18].

MusicFX [8] is an intelligent system that selects background music for a group of individuals working out at a fitness center by considering the probabilities of favorite music genres among users. As the aggregation technique, MusicFX employs the AwM and considers only genres where all preferences of users are above a predefined threshold. Adaptive Radio [10] is a server that selects songs to be played for a group of individuals by considering the information about what type of music they do not desire to hear. In doing so, it combines such negative preferences with a without misery aspect. The FlyTrap [11] and Adaptive In-Vehicle Multimedia System **[12]** are other salient examples of GRSs in the music domain. More specifically, the former produces a virtual DJ that makes a music playlist for users in a room by considering both the musical preferences of the users and background knowledge of music genres, songs, and artists. As the aggregation technique, FlyTrap employs a variant of the simple computation method, counting the number of times that the users in the group prefer a song. The latter, on the other hand, suggests a series of tracks for a group of people sharing a vehicle for traveling by constructing a group profile based on the average of passengers' preferences about songs.

PolyLens [13] is proposed as an extension of famous MovieLens [14] and provides movie recommendations for groups of users instead of individuals. To produce group recommendations,

PolyLens employs group profiles that are constructed by the LM technique. Another example of GRSs in the movie domain is gRecs [15], which first identifies groups of users with similar interests and then recommends a movie list according to the group ratings estimated with the LM. HappyMovie [16] is developed as a Facebook application providing movie recommendations to groups based on the tastes of the users and the trust among members in the group. FlexiFeed [17] is an approach that constructs feedback vectors to represent user interests better and utilize them to maximize the satisfaction of individuals from produced group recommendations in the domain of movie. Also, HappyMovie and FlexiFeed employ both Avg and LM methods to provide group recommendations.

Pocket Restaurant Finder [4] is an environment recommending a series of restaurants for a group of users by utilizing a variant of the Avg technique to aggregate their interests in location and food (e.g., taste, cuisine type, restaurant amenities, and price category). Travel Decision Forum (TDF) [5], CATS [6], and Intrigue [7] are some other popular applications in the context of tourism, and each of them utilizes the Avg technique to provide group referrals. More specifically, TDF assists in the decision-making of a group of users planning to travel together by providing group recommendations based on their preferences. CATS, on the other hand, produces a list of recommended ski-packages for groups of users with similar interests. Intrigue is a system that helps the organization of guided tours by suggesting touristic attractions according to the properties of the tour participants.

In literature, there exist only a few studies that take into account the user personality for providing appropriate group recommendations. For example, Recio-Garcia et al. [26] introduce an approach providing personality-aware group recommendations by investigating how conflicts of group members influence the recommendation process. For measuring the behaviors of individuals in conflict situations, they build a profile for each member in the group using a measure named *Conflict Mode Weight*. They integrate such profiles into the recommendation process to weight the preferences of users. Also, Rossi et al. [24] propose a new group recommendation approach that recommends suitable items for user groups by utilizing a personality-based utility function, which models the level of a user's altruistic behavior according to their agreeableness personality trait. Finally, Sanchez et al. [27] present a novel method that considers both the group personality composition and the social relationships among users to improve the quality of the produced group recommendations.

The behavioral characteristics of group members play a vital role in analyzing the propensities of the group appropriately. Although a few recent studies address this issue [24, 26, 27], there is still a need to develop novel aggregation mechanisms that comprehensively investigate the roles of personality traits during the aggregation process and utilize them to provide group recommendations.

4. A GROUP RECOMMENDATION FRAMEWORK UTILIZING A PERSONALITY-BASED AGGREGATION TECHNIQUE

In this section, we introduce our group recommendation framework that consists of two fundamental steps, as depicted in Fig. 1: *(i)* identifying groups of users with similar interests by *k*-means algorithm, *(ii)* providing top-*N* group recommendations based on the group ratings estimated through a novel aggregation technique, namely, *Personality weighted Average*, which utilizes the personal traits of group members during the aggregation process.
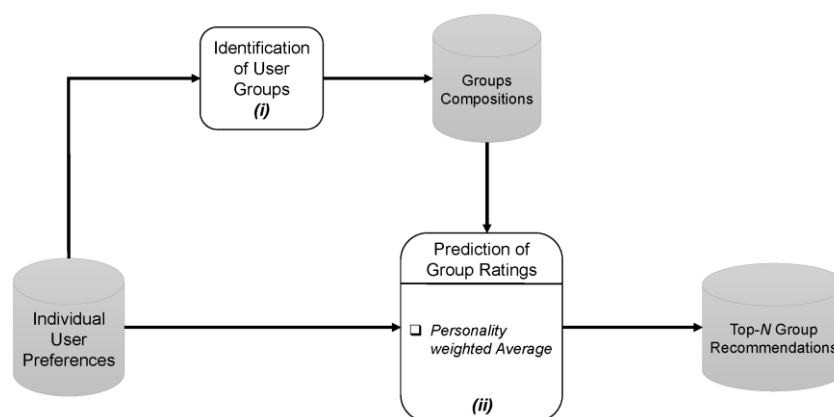
Figure 1. The proposed group recommendation framework based on novel personality-based aggregation techniques.

4.1. Identification of Groups

In the literature, although there exist many studies utilizing established groups [6, 7, 13], groups of users with similar interests in a community are commonly identified automatically based on the preferences of individuals [18, 19, 28]. Detecting harmonious user groups is crucial for the success of the utilized GRSs, as it is intuitively easier to gratify like-minded individuals rather than a randomly ensembled mass. Such automated identification of the user groups is also beneficial because *(i)* the procedure of partitioning users into groups is a continuous process requiring regular updates because of the changes in the interests of individuals in time and *(ii)* manual identification of groups becomes challenging as the number of users proliferates.

To identify groups of individuals having similar preferences in a community, there exist at least two common strategies as follows: *(i)* calculating correlations between all individuals and utilize them to form groups of users [29], and *(ii)* detecting user groups by applying a clustering algorithm such as *k*-means [18, 19, 30], *k*-medoids [31], or hierarchical clustering methods [32]. Although the former is an efficient strategy to detect groups of similar users, the computation time required to compute correlations among all users drastically increases as the number of items/users proliferates, which leads to a problem of time-complexity. The latter, on the other hand, is a practical strategy in identifying groups of like-minded individuals and more efficient in terms of time complexity when compared to the former strategy. Therefore, to identify groups of similar users, we follow the latter approach by applying the *k*-means clustering algorithm on the original user-item rating matrix due to its simplicity and efficiency [18, 19]. Note that, we employ the adjusted cosine metric while performing the *k*-means clustering algorithm since it is one of the most successful metrics in computing similarities between two users [33]. After the identification of user groups, we estimate the group ratings by the personality-based aggregation technique explained in detail in the following sections.

4.2. Prediction of Group Ratings

Group recommendations are usually generated based on the group profiles that embody the interests of a group of people as a whole. The aggregation techniques are useful tools to construct such group profiles by combining the preferences of users in the group. Although there exist various aggregation techniques that produce qualified group recommendations by analyzing the properties of provided ratings, they usually do not aware of the personal characteristics of members in the group while constructing group profiles. However, the psychological aspects of individuals, such as emotions and personality, play a vital role in decision-making since people having similar personal characteristics are more likely to have similar interests [24]. Also, many recent studies in the field of recommender systems have utilized personality traits to develop user-oriented recommendation approaches [26, 27, 34, 35].

In psychology, personality is defined as the patterns of motivation, behavior, feeling, preference, and thought, which are expressed in different circumstances [24]. In the context of recommender systems, it can be defined as a user profile, which is independent of both context and domain, i.e., it does not change with location, time, or some other context/domain [27]. Such user profiles are usually generated through the big five-factor model of personality [25], which identifies the personality based on the following five fundamental traits: *Open to experience* (i.e., curious, imaginative, insightful); *Agreeable* (i.e., forgiving, generous, trusting, sympathetic); *Emotionally stable* (i.e., self-confident, stable, balanced); *Conscientious* (i.e., reliable, organized, efficient, responsible); and *Extraverted* (i.e., energetic, talkative, assertive, active).

To achieve well-constructed group profiles that reflect the characteristics of the members and consequently improve recommendation quality, we propose to utilize such personality information of group members in the phase of aggregating. In the following, we introduce a novel aggregation strategy based on the personality traits of group members.

4.2.1. The *Personality weighted Average* technique

The existing aggregation techniques combine individual preferences by commonly assuming that each user in a group has a similar influence on the overall group taste. However, different members might have different impacts on the decision-making process of the group, as mentioned in the Introduction. To address this issue, we propose the *Personality weighted Average* (*PwAvg*) technique that incorporates the personality traits of members in the group along with the Avg technique. For each member of the group, the PwAvg initially calculates an influence score that measures the relative importance of the corresponding user on the group in terms of a particular personality trait, and then utilize it to weight the preference of the member while performing aggregating.

In this study, we use the Personality dataset [36], where the personality traits of a person, i.e., openness (*o*), agreeableness (*a*), emotional stability (*es*), conscientiousness (*c*), and extraversion (*e*), are assessed with the scores based on a 7-star rating scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Suppose that u is a member of a group G and provide a rating for item i, and $s_{u,t}$ is the score of u for trait t, where $t \in \{o, a, es, c, e\}$. The PwAvg_t that is the personality weighted average technique relying on t calculates a group rating $R_{G,i}$ for i using the formula given in Eq. (1).

$$R_{G,i} = \frac{\sum_{u \in G} r_{u,i} \times s_{u,t}}{|G|} \tag{1}$$

Note that the range of the trait scores of users might differ for each group, which requires a normalization process to determine the relative importance of users in a group more accurately. For this purpose, we transform $s_{u,t}$ scores into [0, 1] scale by min-max normalization before calculating $R_{G,i}$.

5. EXPERIMENTAL STUDIES

In this section, we evaluate the performance of the proposed aggregation technique in terms of accuracy with several experiments on real-world datasets.

5.1. Datasets and Evaluation Metric

In the experiments, we use the Personality dataset [36] collected by the GroupLens research team at the University of Minnesota. In this dataset, user preferences on movies are presented with discrete ratings on a ten-star rating scale ranging from 1 to 5. The Personality dataset also contains assessment scores on a seven-star rating scale about five personality traits for each user (i.e., openness, agreeableness, emotional stability, conscientiousness, and extraversion), as stated in Section 4.2.1. Since the original collection is large and sparse (about 98.4%), we employ two subsets of the collection in which each user

and item has at least 5 and 10 ratings, referred to as Per5 and Per10, respectively. Table 2 presents detailed information about Per5 and Per10.

Table 2. Detailed information about Per5 and Per10

Dataset	#Users	#Items	#Ratings	Sparsity (%)
Per5	1,819	14,868	984,499	96.4
Per10	1,819	10,375	954,985	94.9

To examine the performance of the proposed aggregation techniques, we utilize the normalized Discounted Cumulative Gain (*n*DCG) metric, which is commonly used in group recommendation studies [18, 29]. The *n*DCG measures the quality degree of the ranked recommended list by considering both the position of the items in the recommended list and the actual ratings of the items. Suppose that u is a user in a group G, and $r_{u,i}$ is the actual rating of u for item i. If $\{i_1, i_2, ..., i_N\}$ is the ranked items generated as a recommendation list for G, then the Discounted Cumulative Gain (DCG) and *n*DCG for each user in the G are computed using the formula given in Eqs. (2) and (3), respectively.

$$DCG_N^u = r_{u,i_1} + \sum_{n=2}^{N} \frac{r_{u,i_n}}{\log_2(n)} \tag{2}$$

$$nDCG_N^u = \frac{DCG_N^u}{IDCG_N^u} \tag{3}$$

where $IDCG_N^u$ indicates the maximum gain for u, which is obtained with the ideal re-ordering N items. In calculating *n*DCG scores for each user, it is required to know the actual ratings of users for the items in the recommendation list. However, the actual ratings of users are usually missing because of the sparsity issues, which is one of the most challenges in recommender systems. To cope with this problem, we estimate the actual ratings of users using a user-based collaborative filtering algorithm to have ground truth for each user, as in [37, 38].

5.2. Experimentation Methodology

In the study, we follow a five-fold cross-validation experimentation methodology to evaluate the proposed aggregation technique. We randomly divide the set of items into five subsets to perform the cross-validation, i.e., each subset contains 20% of the items. At each iteration, we utilize one of the subsets as the test set and the remaining subsets as the training set. The training set is used to determine user groups, while the test set is utilized to investigate the performance of both the proposed and baseline aggregation techniques. Note that overall *n*DCG scores are calculated by taking the average of five-fold experimental accuracy results.

To construct user groups, we employ the *k*-means clustering algorithm, as stated in Section 4.1. For investigating the effect of group size on the aggregation techniques, we perform several experiments with groups of different numbers *k* varying from 2 to 64. After identifying groups, for each number of groups, we estimate group ratings using the PwAvg and baseline aggregation techniques (i.e., Avg, AwM, and LM). Also, we employ five different variants of the PwAvg technique according to the utilized personality trait. After predicting a group rating for each item, we produce group recommendations by selecting top-*N* items, where *N* is set as 3, 5, 10, and 20. Finally, we calculate *n*DCG scores of the individuals based on the recommended top-*N* lists and then average them.

5.3. Experimental Results and Discussion

To evaluate the accuracy performance of the proposed *Personality weighted Average (PwAvg)* technique on predicting group ratings, we performed various trials with different parameters, including the utilized personality trait, the size of recommendation list (*N*), and the number of groups (*k*). Also, we compared empirical outcomes against the baseline techniques, Avg, AwM, and LM, for both Per5 and Per10 datasets, as presented in Tables 3 and 4, respectively. We also performed statistical significance *t*-tests to compare obtained *n*DCG results for the baselines and PwAvg variants, as presented in the footnote of tables.

Table 3. *n*DCG results for Per5 dataset

| | | *Number of Groups (k)* | | | | | |
| | | *Large Groups* | | *Medium Groups* | | *Small Groups* | |
top-*N*	Aggregation Technique	2	4	8	16	32	64
3	Avg	0.647	0.671	0.708	0.732	0.743	0.750
	AwM	0.635	0.692	0.718	**0.739**	**0.745**	0.749
	LM	0.652	0.671	0.710	0.736	0.739	**0.751**
	PwAvg *openness*	0.629	0.699	0.722	**0.730**	0.702	0.674
	agreeablenesss	0.683†	0.708†	0.704	0.723	**0.704**	0.672
	emotional_stability	**0.694**†	**0.720**†	**0.722**	0.726	0.699	**0.683**
	conscientiousness	0.681†	0.699	0.692	0.703	0.696	0.677
	extraversion	0.668	0.669	0.716	0.721	0.683	0.678
5	Avg	0.666	0.664	0.689	0.724	0.736	**0.751**
	AwM	0.653	0.676	0.689	**0.728**	**0.740**	0.751
	LM	0.650	0.689	0.707	0.724	0.746	0.742
	PwAvg *openness*	0.664	0.671	0.691	0.707	0.699	0.646
	agreeablenesss	0.705†	0.692	0.692	0.722	0.697	0.639
	emotional_stability	**0.711**†	**0.715**†	**0.705**†	0.725	0.715	0.644
	conscientiousness	0.689†	0.704†	0.700†	0.715	0.710	0.633
	extraversion	0.657	0.694	0.679	0.718	0.706	0.638
10	Avg	0.640	0.688	0.721	**0.721**	**0.731**	**0.743**
	AwM	0.659	0.690	0.727	0.720	0.730	0.739
	LM	0.642	0.689	0.730	0.727	0.731	0.739
	PwAvg *openness*	0.641	0.693	0.720	0.708	0.716	0.676
	agreeablenesss	0.659	0.697	0.738	0.716	0.711	0.674
	emotional_stability	**0.677**†	**0.723**†	**0.741**†	0.718	0.717	0.682
	conscientiousness	0.660	0.721†	0.733	0.713	0.713	0.679
	extraversion	0.677†	0.707†	0.738	0.705	0.708	0.681
20	Avg	0.660	0.688	0.719	0.731	**0.738**	**0.735**
	AwM	0.659	0.688	0.716	0.734	0.733	0.724
	LM	0.660	0.681	0.714	0.732	0.732	0.723
	PwAvg *openness*	0.655	0.693	0.717	0.734	0.683	0.653
	agreeablenesss	0.687†	0.707†	0.715	0.730	0.683	0.644
	emotional_stability	**0.705**†	**0.719**†	0.718	**0.740**	0.680	0.654
	conscientiousness	0.693†	0.718†	0.727	0.737	0.692	0.653
	extraversion	0.704†	0.702	**0.729**	0.735	0.669	0.658

$\dagger$ For significance at 95%; w.r.t. the best performing baseline technique

Table 4. *n*DCG results for Per10 dataset

top-*N*	Aggregation Technique		Large Groups 2	Large Groups 4	Medium Groups 8	Medium Groups 16	Small Groups 32	Small Groups 64
			Number of Groups (k)					
			Large Groups		*Medium Groups*		*Small Groups*	
			2	**4**	**8**	**16**	**32**	**64**
3	Avg		0.698	0.736	0.725	0.738	0.742	0.758
	AwM		0.680	0.716	0.722	0.731	**0.744**	**0.760**
	LM		0.651	0.708	0.714	0.733	0.739	0.751
	PwAvg	*openness*	0.688	0.718	0.710	0.735	0.689	0.677
		agreeablenesss	0.707	0.729	0.715	0.726	0.681	0.669
		emotional_stability	**0.713**†	**0.745**†	**0.752**†	**0.739**	0.693	0.682
		conscientiousness	0.697	0.707	0.722	0.727	0.682	0.671
		extraversion	0.686	0.718	0.741	0.737	0.694	0.678
5	Avg		0.710	0.678	**0.728**	**0.734**	0.744	**0.757**
	AwM		0.659	0.673	0.725	0.733	**0.746**	0.754
	LM		0.680	0.686	0.705	0.726	0.738	0.746
	PwAvg	*openness*	0.677	0.657	0.722	0.713	0.715	0.682
		agreeablenesss	0.684	0.689	0.715	0.715	0.722	0.677
		emotional_stability	**0.727**†	**0.707**†	0.723	0.714	0.727	0.679
		conscientiousness	0.702	0.693	0.723	0.711	0.732	0.668
		extraversion	0.701	0.706†	0.721	0.708	0.722	0.676
10	Avg		0.714	0.710	0.710	0.733	**0.741**	**0.755**
	AwM		0.662	0.683	0.710	**0.735**	0.741	0.753
	LM		0.711	0.702	0.722	0.727	0.732	0.747
	PwAvg	*openness*	0.698	0.716	0.716	0.723	0.687	0.666
		agreeablenesss	0.711	0.716	0.720	0.719	0.681	0.657
		emotional_stability	**0.726**†	**0.725**†	**0.738**†	0.720	0.692	0.665
		conscientiousness	0.714	0.719	0.732†	0.729	0.686	0.658
		extraversion	0.696	0.724†	0.723	0.708	0.678	0.660
20	Avg		0.694	0.702	0.738	**0.737**	**0.738**	0.746
	AwM		0.610	0.689	0.719	0.730	0.727	0.731
	LM		0.674	0.700	0.724	0.729	0.730	0.737
	PwAvg	*openness*	0.674	0.700	0.736	0.717	0.698	0.675
		agreeablenesss	0.715†	0.709	0.737	0.719	0.685	0.667
		emotional_stability	**0.716**†	**0.716**†	**0.738**	0.735	0.693	0.680
		conscientiousness	0.710†	0.709	0.731	0.719	0.701	0.671
		extraversion	0.687	0.714	0.735	0.728	0.697	0.668

† For significance at 95%; w.r.t. the best performing baseline technique

As can be seen in Tables 3 and 4, the *n*DCG results of the experiments conducted for both datasets demonstrate that the accuracy performance of all baseline techniques improves as the number of groups increases, i.e., the size of the groups diminishes. On the other hand, the PwAvg variant techniques achieve the highest nDCG scores, especially when large groups are of interest. The reason for this consequence is that increasing group size leads to having groups consisting of individuals with diverse characteristics; thus, considering not only preferences but also personalities of group members while performing aggregation allows producing high-quality group recommendations.

In comparing results obtained from the Per5 and Per10 datasets, both PwAvg variants and all baseline techniques are more successful on the Per10 dataset for almost all schemes. The main reason for this finding is that the total number of given preferences for an item in the Per10 dataset outnumber the one in Per5, which makes the utilized aggregation method becomes more robust, as more preferences are involved in the aggregation procedure.

The empirical outcomes for both datasets also show that PwAvg variant techniques usually outperform the baseline techniques, especially when the groups are large. Among all personality traits, utilizing emotional stability is relatively more efficient for building the PwAvg technique when compared to other personality traits. Moreover, for both datasets, the improvements of the emotional stability based

PwAvG over all the baseline techniques appear to be statistically significant at 95% confidence level, especially when groups are large. Therefore, it can be concluded that recommending items highly-rated by emotionally stable group members when compared to items voted by unstable ones improve overall satisfaction.

6. CONCLUSIONS AND FUTURE WORK

Group recommender systems are useful tools that produce appropriate referrals to a group of people with common interests rather than individuals. In such systems, group recommendations are generally produced based on group ratings, which are determined by aggregating the preferences of members in the group. However, such an aggregation procedure is a complicated task as the groups usually consist of individuals with diverse personal characteristics, which requires to develop novel aggregation mechanisms.

In this study, we propose a personality-aware aggregation technique named *Personality weighted Average (PwAvg)* that estimates group ratings by considering not only the preferences of the group members but also the behavioral features of individuals. More specifically, the *PwAvg* measures the influence degree of members in the group through the following five fundamental personality traits: *openness*, *agreeableness*, *emotional stability*, *conscientiousness*, and *extraversion*; then, it employs them to weight their ratings during the recommendation process. The experiments performed on two benchmark datasets demonstrate that the PwAvg outperforms three baseline aggregation techniques, especially for large user groups. This result is achieved regardless of the size of the recommendation list, as confirmed by the statistical significance tests. Moreover, the empirical outcomes also suggest that building the PwAvg with *emotional stability* trait is more effective in terms of recommendation accuracy when compared to other ones. In conclusion, taking into account the different personal characteristics of group members while performing aggregating individual preferences contributes to producing high-quality group referrals as groups get crowded.

The obtained findings support that the personalities of group members are as significant as their experience about the domain of interest to provide appropriate group recommendations. Therefore, future research might include improving the proposed PwAvg method by considering more than one trait together in harmony. Also, the behavioral information of individuals can be utilized for building more suitable user groups.

REFERENCES

[1]   Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 2005; 17(6): 734-749.

[2]   Yalcin E, Bilge A. Binary multicriteria collaborative filtering. Turkish Journal of Electrical Engineering and Computer Sciences 2020; 28(6): 3419-3437.

[3]   Goren-Bar D, Glinansky O. FIT-recommending TV programs to family members. Computers & Graphics 2004; 28(2); 149-156.

[4]   McCarthy JF. Pocket restaurantfinder: A situated recommender system for groups. In: Workshop on Mobile Ad-Hoc Communication at the 2002 ACM Conference on Human Factors in Computer Systems, 2002.

[5]   Jameson A. More than the sum of its members: challenges for group recommender systems. In: Proceedings of the working conference on Advanced visual interfaces, 2004; 48-54.

[6]   McCarthy K, Salamó M, Coyle L, McGinty L, Smyth B, Nixon P. Cats: A synchronous approach to collaborative group recommendation. In: Florida Artificial Intelligence Research Society Conference (FLAIRS), 2006; 86-91.

[7]   Ardissono L, Goy A, Petrone G, Segnan M, Torasso P. Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. Applied artificial intelligence 2003;17(8-9): 687-714.

[8]   McCarthy JF, Anagnost TD. MusicFX: an arbiter of group preferences for computer supported collaborative workouts. In: Proceedings of the 1998 ACM conference on computer supported cooperative work, 1998; 363-372.

[9]   Jameson A. More than the sum of its members: challenges for group recommender systems. In: Proceedings of the working conference on Advanced visual interfaces, 2004; 48-54.

[10]   Chao DL, Balthrop J, Forrest S. Adaptive radio: achieving consensus using negative preferences. In:Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work, 2005; 120-123.

[11]   Crossen A, Budzik J, Hammond KJ. Flytrap: intelligent group music recommendation. In: Proceedings of the 7th international conference on Intelligent user interfaces, 2002; 184-185.

[12]   Zhiwen Y, Xingshe Z, Daqing Z. An adaptive in-vehicle multimedia recommender for group users. In: IEEE 61st Vehicular technology conference, 2005; 5: 2800-2804.

[13]   O'connor M, Cosley D, Konstan JA, Riedl J. PolyLens: a recommender system for groups of users. In:ECSCW, 2001; 199-218.

[14]   Harper FM, Konstan JA. The movielens datasets: History and context. Acm transactions on interactive intelligent systems 2015; 5(4): 1-19.

[15]   Ntoutsi I, Stefanidis K, Norvag K, Kriegel HP. gRecs: A group recommendation system based on user clustering. In: International Conference on Database Systems for Advanced Applications, 2012; 299-303.

[16]   Quijano-Sanchez L, Recio-Garcia JA, Diaz-Agudo B. Happymovie: A facebook application for recommending movies to groups. In: 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, 2011; 239-244.

[17]   Roy SB, Thirumuruganathan S, Amer-Yahia S, Das G, Yu C. Exploiting group recommendation functions for flexible preferences. In: IEEE 30th international conference on data engineering, 2014; 412-423.

[18]   Seo YD, Kim YG, Lee E, Seol KS, Baik DK. An enhanced aggregation method considering deviations for a group recommendation. Expert Systems with Applications 2018; 93: 299-312.

[19]   Boratto L, Carta S, Fenu G. Discovery and representation of the preferences of automatically detected groups: Exploiting the link between group modeling and clustering. Future Generation Computer Systems 2016;64: 165-174.

[20]   Masthoff J. Group recommender systems: aggregation, satisfaction and group attributes. In: recommender systems handbook, 2015; 743-776.

[21] Lieberman H, Van Dyke N, Vivacqua A. Let's browse: a collaborative browsing agent. Knowledge-Based Systems 1999; 12(8): 427-431.

[22] Nunes MASN. Recommender systems based on personality traits, PhD, Université Montpellier, France, 2008.

[23] Hu R, Pu P. A study on user perception of personality-based recommender systems. In: International conference on user modeling, adaptation, and personalization, 2010; 291-302.

[24] Rossi S, Cervone F. Social Utilities and Personality Traits for Group Recommendation: A Pilot User Study. In: ICAART (1), 2016; 38-46.

[25] Costa Jr, Paul T, McCrae RR. Four ways five factors are basic. Personality and individual differences 1992; 13(6): 653-665.

[26] Recio-Garcia JA, Jimenez-Diaz G, Sanchez-Ruiz AA, Diaz-Agudo B. Personality aware recommendations to groups. In: Proceedings of the third ACM conference on Recommender systems, 2009; 325-328.

[27] Quijano-Sanchez L, Recio-Garcia JA, Diaz-Agudo B. Personality and social trust in group recommendations. In: 22Nd IEEE international conference on tools with artificial intelligence, 2010; (2): 121-126.

[28] Boratto L, Carta S. State-of-the-art in group recommendation and new approaches for automatic identification of groups. In: Information retrieval and mining in distributed environments, 2010; 1-20.

[29] Baltrunas L, Makcinskas T, Ricci F. Group recommendations with rank aggregation and collaborative filtering. In: Proceedings of the fourth ACM conference on Recommender systems, 2010; 119-126.

[30] Boratto L, Carta S. Using Collaborative Filtering to Overcome the Curse of Dimensionality when Clustering Users in a Group Recommender System. In: ICEIS, 2014; 564-572.

[31] Khazaei E, Alimohammadi A. An automatic user grouping model for a group recommender system in location-based social networks. ISPRS International Journal of Geo-Information 2018; 7(2): 67.

[32] Cantador I, Castells P. Extracting multilayered Communities of Interest from semantic user profiles: Application to group modeling and hybrid recommendations. Computers in Human Behavior 2011; 27(4): 1321-1336.

[33] Schafer JB, Frankowski D, Herlocker J, Sen S. Collaborative filtering recommender systems. In: The adaptive web, 2007; 291-324.

[34] Tkalcic M, Kunaver M, Tasic J, Košir A. Personality based user similarity measure for a collaborative recommender system. In: Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges, 2009; 30-37.

[35] Nunes MAS, Hu R. Personality-based recommender systems: an overview. In: Proceedings of the sixth ACM conference on Recommender systems, 2012; 5-6.

[36] Nguyen TT, Harper FM, Terveen L, Konstan JA. User personality and user satisfaction with recommender systems. Information Systems Frontiers, 2018; 20(6): 1173-1189.

[37] Sacharidis D. Top-N group recommendations with fairness. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 2019; 1663-1670.

[38] Yalcin E, Ismailoglu F, Bilge A. An entropy empowered hybridized aggregation technique for group recommender systems. Expert Systems with Applications 2021; 166: 114111.