Social Utilities and Personality Traits for Group Recommendation: A Pilot User Study

Silvia Rossi and Francesco Cervone
Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione, Universita' degli Studi di Napoli "Federico II", Napoli, Italy

Keywords: Social Choice, Group Recommendation, Big-Five, Group Decision Making.

Abstract: Recommendations to a group of users can be provided by the aggregation of individual users' recommendations using social choice functions. Standard aggregation techniques do not consider the possibility of evaluating social interactions, roles, and influences among group's members, as well as their personalities, which are, indeed, crucial factors in the group's decision-making process. Instead of defining a specific social choice function to take into account such features, the proposed solution relies on the definition of a utility function, for each agent, that takes into account other group members' preferences. Such function models the level of a user's altruistic behavior starting from his/her agreeableness personality trait. Once such utility values are evaluated, the goal is to recommend items that maximize the social welfare. Performance is evaluated with a pilot user study and compared with respect to Least Misery. Results showed that while for small groups LM performs slightly better, in the other cases the two methods are comparable.

1 INTRODUCTION

Recommender Systems are powerful tools that provide suggestions about items for users. The list of recommended items is the result of a decision process that aims to understand, for example, which product a specific user wants to buy, which music to listen, which movie to see or which news to read. In this context, the goal of Group Recommender Systems (GRSs), differently from Individual Recommender Systems (or simply RSs), is to recommend items to a whole group of users taking into account all the individual preferences. The study of GRSs is still an open research problem. Preferences and tastes of individuals are collected in the same way of RSs, but the real problem falls into the correct aggregation of these preferences in order to recommend items that maximize global satisfaction, or at least, minimize dissatisfaction of all the members.

The problem of aggregating individual preferences has been widely studied in Mathematics, Economics and Multi-agent systems (MAS), with the definition of Social Choice functions. In literature, several solutions have been proposed (Masthoff, 2011). The most common technique is the Average Satisfaction (AS): it treats all group members as being equal by averaging preferences (or recommendations) of all the member in order to produce a final list for the entire group. Least Misery (LM), instead, cares about the possible dissatisfaction of some members by choosing items that minimize it. Anyway, no one of the standard techniques considers that there are also other factors that can influence the group decision. Groups can be dynamic, as well as members' behaviors depending on situations. Real *group decision making* is a complex mechanism that involves relationships among the members, users' personality, as well as their experience about the domain of interest.

In this work, we study the influence of individual users personality in the group decision-making process. In a realistic scenario, indeed, the personalities of the group members can have an impact on group decisions. For example, there can be people that rarely change their minds because they believe that their own decision is the best for everyone, or simply because they do not want to reduce their utility in favor of others. Other types of people instead, can be worried about the satisfaction of all the other members, at the cost of the personal one. Thus, the latter ones are willing to lose some utility in order to reach a valid and suitable agreement for the entire group. In order to consider these elements, it is necessary to study users' personalities through some models proposed in human sciences area. One of the most common is certainly the Five-Factor Model (FFM).

Rossi, S. and Cervone, F.
Social Utilities and Personality Traits for Group Recommendation: A Pilot User Study.
DOI: 10.5220/0005709600380046
In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART 2016) - Volume 1*, pages 38-46
ISBN: 978-989-758-172-4

According to this, the behavioral features of a person can be summarized and described through five factors, called also Big-Five. They are: openness, conscientiousness, extraversion, neuroticism and agreeableness (Costa and MacCrae, 1992). In particular, in this work, we focus on the role of the *agreeableness* factor in the definition of a utility function that models altruistic behavior.

In literature, some approaches are starting to model GRSs that weights users preferences in different ways according to a specific user-related parameters. For example, in Gartrell et al. (2010) users preferences are weighted according to their expertise, while in Rossi et al. (2015) users' dominance and influence are taken into consideration. In the proposed approach, we do not consider weights in the aggregation process, but the utility function used to evaluate users rating on individual items takes into account the whole group preferences and the user agreeableness trait. In this sense, our work is related to the approach of Salehi-Abari and Boutilier (2014), where individual empathetic utilities are defined taking into account local relationships with neighborhoods in a social network. However, in Salehi-Abari and Boutilier (2014) the Authors do not specify how to evaluate such numerical relationships, while they focus on computational aspects of scaling up with large networks of friends. While the role of personality has been addressed before, in literature, to improve the performance of RSs (Nunes and Hu, 2012; Hu and Pu, 2011), up to our knowledge, this is the first attempt to introduce personality factors in group decision making through the use of personality-based utility functions. The only relevant approach is the one of Quijano-Sanchez et al. (2013), where the personality of every individual in the group is evaluated in terms of conflicting resolution styles. The user's ranking of an item is modified by considering all the couples of users and their mutual influence.

This paper is organized as follows. In Section 2, we introduce the Five-Factor Model used to evaluate users' agreeableness personality trait values, while, in Section 3, we introduce the utility function used to model altruistic behaviors and the evaluation, by personality tests, of parameters that characterize such function. In Section 4, we present the developed application and the metrics used to evaluate with real users the proposed utility function with respect to Least Misery (LM). Finally, results are discussed in Section 5.

2 BIG-FIVES AND PERSONALITY TRAITS

Research has shown that personality is a primary factor which influences human behaviors. The **Five-Factor Model** (FFM) describes human personality using five factors, also known as **Big-Five** (McCrae and Costa, 1987). *Neuroticism* represents an emotional instability characterized by negative emotions like fear, anger, sadness and low self-esteem (McCrae and Costa, 1989). High neuroticism level people rarely control their impulses and cope the stress (McCrae and Costa, 1985). *Extraversion* is an indicator of assertiveness and trust. Extravert people easily create interpersonal relationships and love working and being together with others (Costa and McCrae, 1995). *Agreeableness* describes the level of sympathy, availability and cooperativeness. People with low level of this factor are competitive, skeptics and antagonistic. It measures how much a person is nice and altruist (Costa and McCrae, 1995). *Openness* represents the inclination to openness to new experiences, having an active imagination and a preference about the will to find new ideas (Costa and McCrae, 1995). Closed people are less flexible and rarely understand others' point of views. Finally, *conscientiousness* describes how much an individual is responsible, disciplined and dutiful (McCrae and Costa, 1989).

Several studies also demonstrated that personality traits have an impact in group decision making. Anderson et al. (2001) tried to correlate Big-Five with the *status*[1] of college students. A student with a high level of status implies a considerable level of attention by others towards him. Furthermore, status relates to respect and influence in the social context. Authors measured students' status through an assessment made by the student themselves and according to the positions that they play in the college organization. Results showed that the extraversion has a positive correlation with the status; conversely, neuroticism has a negative correlation, but only for men. The latter result can be explained because Brody (2000) said that sadness, depression, fear, shame and embarrassment (neuroticism features) are viewed as "unmanly". Thus, men who show these emotions are evaluated more negatively with respect to women. For the latter, these emotions are considered ordinary.

Sager and Gastil (2006) studied the impact of Big-Five in producing a *supportive communication*. The latter is opposed to *defensive communication* in which group members see others as a threat and try to stay one step ahead them. In *supportive communication*

[1] Status describes the role that a person has in a social group.

instead, all the members allow others to express their own opinions and consider other choices. Extraversion, agreeableness and openness has a positive correlation in this context.

Ma (2005) studied the influence of personality in conflict situations that happen during negotiations. He correlates *conflict resolution styles*, which represent different people reactions in case of conflict, with big five by using two factors: the *assertiveness*, which describes how much an individual try to satisfy his needs, and *cooperativeness*, which suggests the level of collaboration and the intention of maximizing others' utilities. Results showed that: neuroticism is negatively correlated to *compromising*; extraversion is positively correlated to *competing* and *collaborating*; agreeableness is negatively correlated to *competing* and positively to *compromising*; finally, conscientiousness and openness have no significant correlations.

Starting from the previous considerations, in this work, we decided to move a first step by considering only a single personality trait at the time starting from the agreeableness trait, while leaving the combination with the others as a future work. We decided not to rely on the evaluation of neuroticism, which has negative correlations with status (a neurotic person could be excluded during a decision process). Conversely, extraverts could be viewed as leaders, but at the same time they are more competitive than collaborative: all of those are qualities that are suitable to our case (the choice of a movie), but already considered in related works. A conscientious person could have a positive impact when deciding important decision, but we do not believe it is significant for our specific goal. This is the same for open individuals which could agree to see a movie not suitable with respect to their preferences, but whose trait could be more difficult to model with respect to the agreeableness factor.

We believe that in choosing a movie in a group of close friends, agreeableness (that is related to altruistic behavior (Costa and McCrae, 1995)) plays an important role. It is obvious that, unless all the components have the same movie tastes, someone will have to give up their desires to see a movie which others do not like. Therefore, agreeable people will make compromises and, for this reason, we decided to analyze the impact of this factor in a GRS. People with high a value of agreeableness, just because of its descriptive characteristic, are more altruistic: basically, it means that this type of people care about the satisfaction of the entire group, or are more willing to compromise in order to obtain a solution that works for the whole group.

3 AN UTILITY FUNCTION FOR AGREEABLENESS

Generally speaking, the aim of a Recommendation System (RS) is to predict the relevance and the importance of items (for example movies, restaurants and so on) that the user never evaluated. More formally, given a set of n users ($U = \{1,\dots,n\}$) and a set of m items ($M = \{1,\dots,m\}$), in our domain application it is a set of movie, an individual recommendation system, for each user i, aims at building a *Preference Profile* of the user i over the complete set M, starting form some initial ratings (typically a value from 1 to 5) each user provides on some elements of M. Once each user $i \in U$ has a preference profile $\succ_i$ over M ($\succ_i = \{x_{i,1},\dots,x_{i,m}\}$) with $x_{i,j} \in \mathcal{R}$, which represents the user i rank (as directly expressed by the user i or evaluated by the system) for the j movie, the goal of a GRS is to obtain $\succ_U = \{x_{U,1},\dots,x_{U,m}\}$, where $x_{U,j}$ is the correspondent ranking for the j movie, as evaluated for the group. Typically, this is obtained by implementing a social choice function $SC : \succ^n \rightarrow \succ_U$, that aggregates all the preferences profiles in $\succ_U = \{x_{U,1},\dots,x_{U,m}\}$. Note that our goal is not to guess the exact value of $x_{U,j}$ the whole group would assign to the movie j, but to properly select the k-best movie in the group preference profile (the ones with the highest rating) and suggest them to the group.

In this work, we propose to consider the agreeableness factor in the process of building recommendations for groups. Instead of defining a specific social choice function that considers such factor, the proposed solution relies on the definition of an individual *utility* function (to build up the preference profile) to evaluate the rating of items for each user, that takes into account the whole group. Such utility function could be interpreted as "the user satisfaction if the recommender system chooses that item for the group". For the utility function, we used a model developed by Charness and Rabin (2002).

Let us consider a group of n users, which have a rating for each element j of the domain (e.g., an expected rating provided by an individual RS x_{ij}). The utility function is composed by two terms; the first concerning a "disinterested" social-welfare criterion, defined as follows:

$$W(x_{1,j}, x_{2,j}, \dots, x_{n,j}) = \delta \cdot \min(x_{1,j}, x_{2,j}, \dots, x_{n,j}) + (1-\delta) \cdot (x_{1,j} + x_{2,j} + \dots + x_{n,j}) \tag{1}$$

where $\delta \in [0,1]$, $x_{1,j}, x_{2,j}, \dots, x_{n,j}$ are called *payoffs* and they are the individual rating of each of the n users of the j-th item. The aim of the first addend is reducing inequity (by helping the worst-off person);

indeed, this value increases proportionally to the minimum payoff. This factor is, basically, a generalization of LM technique just because utility is given by the minimum satisfaction among the users. The second addend is responsible for the maximization of the *social welfare* and increases proportionally to the sum of the individual payoffs. Setting $\delta = 1$, indeed, the function cares only about inequity, just like LM. Setting $\delta = 0$ instead, the function focuses on global satisfaction.

The utility of user i about the item j is a weighted sum of the disinterested social-welfare W and its own payoff defined as follows:

$$U_{i,j}(x_{1,j}, x_{2,j}, ..., x_{n,j}) = (1 - \lambda_i) \cdot x_{i,j} + \lambda_i \cdot W(x_{1,j}, x_{2,j}, ..., x_{n,j}) \quad (2)$$

where $\lambda_i \in [0,1]$ means how much a person pursues its own interest or the social welfare. Setting $\lambda_i = 0$ user takes care only of his personal interest. $\lambda_i = 1$ represents the classical disinterested behavior, of him/her, who does not take part in the group decision, moving the weight of the function on the entire group satisfaction (including its own). Hence, according to Charness and Rabin (2002), the considered function evaluates both the personal and the group satisfaction, depending on an altruistic factor λ_i. Once the new items utilities are evaluated, the goal is to recommend movies that maximize the *social welfare*.

Since the agreeableness factor is positively correlated with an altruistic behavior, here, our goal is to calibrate the λ_i parameter with respect to the individual agreeableness levels, so that the most altruistic have a high value and viceversa.

3.1 Personality Test

Personality can be acquired in both explicit and implicit ways (Dunn et al., 2009). The former measures a user's personality by asking the user to answer a list of designed personality questions. These personality evaluation questionnaires have been well established in the psychology field (Gosling et al., 2003). The implicit approaches acquire user information by observing users' behavioral patterns. Typically, explicit personality acquisition interface are preferred (Dunn et al., 2009) by the user. However, implicit methods require less effort from users. In our study, we adopted the explicit way to measure users' personality, because the evaluation of a single personality trait would require using only a small set of questions, and so a minimum required effort from the user.

There are several questionnaires to predict a person's Big-Five factors. Some of these consist of a lot of questions, in certain cases some hundreds. Too long questionnaires can reduce the people's attention level. A lot of research groups tried to cut the number of questions preserving the accuracy. The most famous small questionnaires are NEO-FFI (NEO Five Factor Inventory) composed by 60 questions (Costa and MacCrae, 1992), the 50-item IPIP-FFM (International Personality Item Pool - Five Factor Model) (Goldberg, 1992), the 44-item BFI (Big Five Inventory), the TIPI (Ten-Item Personality Inventory) (Gosling et al., 2003) and, finally, the **Mini-IPIP** (Mini International Personality Item Pool) (Donnellan et al., 2006). We chose the last because it is very short, but at the same time effective. It consists of 20 questions, 4 for each personality factor. Since our focus only on the evaluation of the user agreeableness, we extracted the 4 related questions (see Table 1; questions should be read in first person). The answer of each question can be a number from 1 to 5, where 1 means "very inaccurate" and 5 "very accurate".

Table 1: Mini IPIP questionnaire for agreeableness evaluation.

#	Text
1	Sympathize with others' feelings.
2	Am not interested in other people's problems. (R)
3	Feel others' emotions.
4	Am not really interested in others. (R)

(R) = Reverse Scored Item.

3.2 Agreeableness and Group Recommendation

Let us assume that each user completed the *agreeableness* evaluation test. Generic user i has a personal value $\alpha_i \in [1,5]$, obtained by averaging individual answer values. Since the λ_i parameter of Equation 2 should belong to the range $[0,1]$ and should depend directly from α_i, we defined it as follows:

$$\lambda_i = \frac{\alpha_i - 1}{4}. \quad (3)$$

This way, when $\alpha_i = 5$ (high level of *agreeableness*), $\lambda_i = 1$, so the user cares only about the group satisfaction. When $\alpha_i = 1$, instead, $\lambda_i = 0$, consequently, user's utility depends only on his personal satisfaction. As we already explained, in the Equation 1 there is the δ parameter that deals with weighting inequity aversion against global satisfaction. We set it equal to 0.5 for giving the same importance to both goals. Utility function 2 refers to the utility of a particular item for a specific user. Thus, preferences $x_{1,j}, x_{2,j}, ..., x_{n,j}$ stand for the rating predictions of a generic item for the n members of the group.

To build recommendations for the groups, we chose the merging recommendations technique. Firstly, the system creates a list of 10 items L_i for each user i evaluated by an RS. Later, it merges all the lists in a single one which we call L:

$$L = \bigcup_{i \in U} L_i \tag{4}$$

where, G is the set of members of the group. Our GRS computes for each user the rating predictions of all the items in L. The latter are exactly the $x_{1,j}, x_{2,j}, \ldots, x_{n,j}$ parameters of Equation 2 (where $n = |U|$ and j is the movie). $U_{i,j}$ for each user i and for each item j is the computed. $x_{U,j}$ denotes the utility of the group if the GRS chooses item j defined as follows:

$$x_{U,j} = \sum_{i \in U} U_{i,j}. \tag{5}$$

Our goal is to maximize the *social welfare*, indeed, system will recommend the 10 items with the highest $x_{U,j}$ value.

4 EXPERIMENTAL STUDY

To conduct our experiments, we developed a client-server application. Client was an Android[2] app and the server was developed in Java[3] using Spring framework[4] hosted by Tomcat Servlet Engine[5].

4.1 Movie Recommendation Server

We built a RESTful web service JSON-based in order to communicate with the Android app. We adopted the Apache Mahout library[6] to predict the user ratings, and chosen the MovieTweetings (Dooms et al., 2013) dataset to train the system and to populate the *Ratings Repository*. MovieTweetings consists of movie ratings contained in well-structured tweets on the *Twitter.com* social network. This information is contained in three files: *users.dat*, *ratings.dat* and *movies.dat*, which provide, respectively, the user identification number, his/her associated ratings and a list of movies. The dataset is updated every day, therefore, its size is constantly changing. At the last access, it contained about 35000 users, 360000 ratings and 20000 movies.

The recommendation engine provides rating predictions when the recommendation API is invoked. To achieve this goal, we used *item-based City Block distance*, also known as *Manhattan distance*. In Mahout implementation, the generic movie j is represented by a boolean vector:

$$j = [x_{1,j}, x_{2,j}, \ldots, x_{k,j}], \tag{6}$$

where k is the number of users in the dataset and $x_{i,j} = 1$ if user i rated the movie j. The distance between two movies rated by user i is the sum of the absolute value of the differences of the two associated vector components. More formally, the distance between items j and h is:

$$d(j,h) = \sum_{i=1}^{k} |x_{i,j} - x_{i,h}|. \tag{7}$$

4.2 Android Application

An Android application was developed in order to gather the information needed by the server to provide recommendations to single users. In order to simplify the operations, the experiment consists in some sequential steps so that each phase unlocks the next one.

The first duty for the user, when he/she accesses the application, is to sign up/sign into the system. The user signs up to the system by entering username, password, gender, age, and education level. When the interaction starts, users have first to provide a certain number of movie ratings (at least for 20 movies), with a value in the range $[1,10]$ (see Figure 1-left) in order to define their profile. The user is provided with an interface to get movie lists and to store movies ratings. If a user is in the training stage, he/she can browse movies by ordering them by most rated or best rated, or searching for a specific movie (filtering by genre or title).

Once the user rated twenty movies, the app automatically shows the personality questionnaire (see Figure 1-right). It consists of four questions, as previously explained.

After this first stage, a user can get movie recommendations from the server. When the server gets the recommendation request, once calculated the best movies for the user, it retrieves additional details about the film, like, for example, the director, writers, actors and genres using OMDb[7] web service. Fortunately, MovieTweetings data set stores, for each movie, its IMDb id, which can be used to address the OMDb service. The Android application shows on the screen the recommendations for the user through textual and graphical descriptions.

[2]https://www.android.com
[3]https://www.java.com/
[4]https://spring.io
[5]http://tomcat.apache.org
[6]http://mahout.apache.org/
[7]http://www.omdbapi.com - The Open Movie Database is a free web service to obtain movie information.

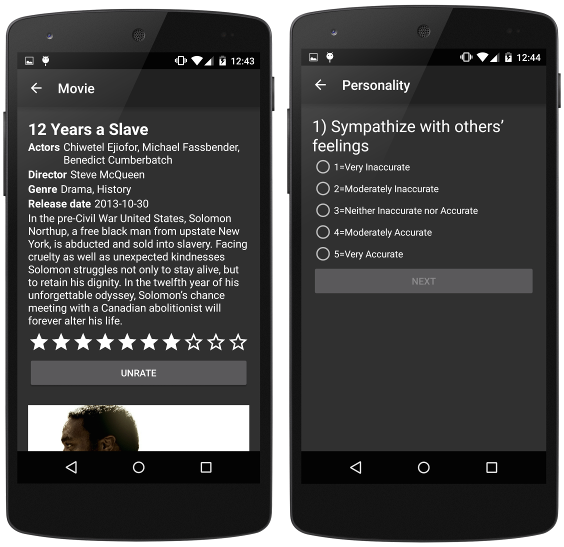

Figure 1: Movie rating interface (left) and first question of personality test (right).

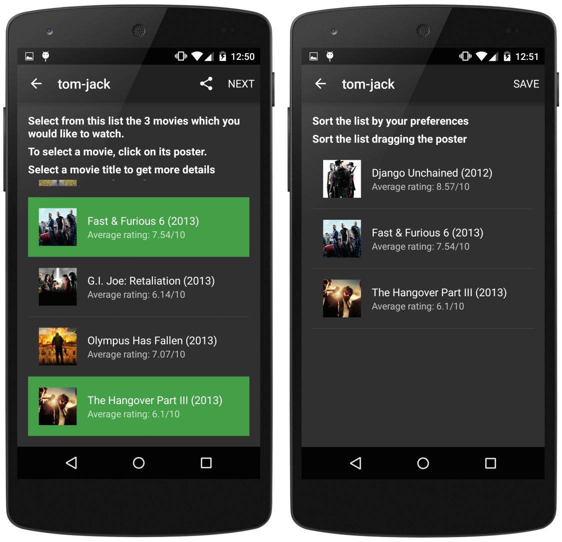

Figure 2: Recommended movies (left) and sorting page (right).

4.3 Methodology

The design of this study is a within-subjects, counterbalanced, repeated measures experiment. The goal of our study is to compare the proposed technique, described in the previous paragraphs, with respect to the Least Misery. We selected the latter because it achieves good performance especially for small groups (O'connor et al., 2001).

Once the questionnaire is completed, a user can start the test by completing the following steps. Firstly, a user creates a group giving it a name, and adds in it one or more members using their usernames. The system, then, recommends a list of 10 movies (see Figure 2-left). In order to generate the list of 10 items for the group, for each technique (the used utility function and LM) the GRS recommends 10 movies ordered in terms of their ratings. To merge them in one list of ten items, we developed an iterative algorithm that, at each step, adds an item from each list (starting from the items with the highest rating) in the output set. If the current item is already in the set, the algorithm skips to the next one, and so on. From this list, the group has to collectively choose three movies that they would like to see together (see Figure 2-left). Finally, the group has to sort the three selected movies (see Figure 2-right) in their joint preference order.

4.4 Evaluation Metrics

Since groups select the best 3 from a 10 movies set, in order to evaluate and compare our method with LM, we considered the following metrics.

precision@3: is the ratio between the number of movies guessed by the GRS (using a specific method) and the sum of the latter and the remaining movies (only considering the first three movies). If G is the set of the groups that participated to the experiment, I_g and P_g represents, respectively, the 3 movies selected by group g and the 3 movies with the highest prediction, then:

$$precision@3 = \frac{1}{|G|}\sum_{g\in G}\frac{|I_g \cap P_g|}{|I_g \cap P_g| + (3 - |I_g \cap P_g|)} \tag{8}$$

nDCG@3: evaluates the ranking of predicted movies with respect to the real ranking j chosen by groups.

$$nDCG@3 = \frac{1}{|G|}\sum_{g\in G}\sum_{j=1}^{3}\frac{rel_{gi_j}}{\max(1, log_2 j)} \tag{9}$$

where,

$$rel_{gi_j} = \begin{cases} 1 \text{ if } i_j \in I_g \\ 0 \text{ otherwise} \end{cases} \tag{10}$$

x_success@3 , for each group, is 1 if the algorithm guessed at least x movies in the 3 selected by the group. With $1 \le x \le 3$:

$$x\_success@3 = \frac{1}{|G|}\sum_{g\in G} x\_success@3_g \tag{11}$$

where,

$$x\_success@3_g = \begin{cases} 1 \text{ if } |I_g \cap P_g| \geq x \\ 0 \text{ otherewise} \end{cases} \quad (12)$$

5 RESULT ANALYSIS

Experiments lasted about two weeks and, as summarized in Table 2, we recruited 68 users (48 groups) with an average age of 27 years, the most of whom were students.

Table 2: User stats.

Number of users	68
Average age	26.9
Minimum age	14
Maximum age	55
Males	45
Females	23
Middle school	1
High school	9
Bachelor students - undergraduate	26
Bachelor students - graduate	11
Master students - undergraduate	11
Master students - graduates	10

Group members were directly selected by one of the users, as described above, and their intersection is not necessarily empty (e.g., some users joined more than one group). A single group was allowed to join the test only for one time. The number of considered group was 48 with an average number of members equals to 2.6. In Table 3, we reported the number of groups considered in the experiments for each group dimension.

Table 3: Group stats.

# mebmers	Amount
2	22
3	23
4	2
5	1
Total	48

When comparing the two techniques, we analyzed their results separately for the groups of dimension two and for the groups of more than two members. The main reason of this choice is that about the half of the groups were composed by two members (see Table 3) and it is known that LM shows the best performances in this case.

precision@3. Results of precision@3 are summarized in Figure 3. From the charts, we can see the better performance of LM on two members groups. It is not a surprise, because LM excels in cases like this. ANOVA test confirms that difference between the two techniques is significant ($F = 3.076$, p-value$= 0.09$).

Regarding groups with more than two members, once again LM is better than our technique, but difference, in this case, is too small to be significant for ANOVA test ($F = 0.11$, p-value$= 0.74$).

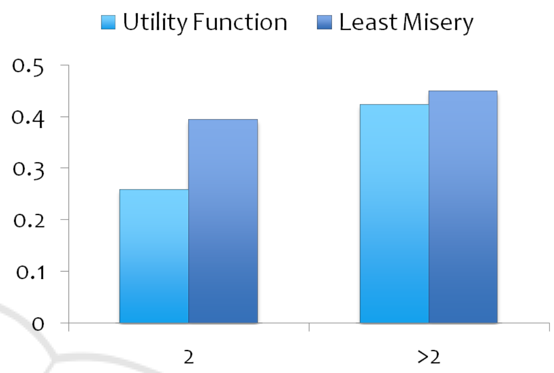

Figure 3: precision@3.

nDCG@3. As for the previous metrics, LM overcomes the proposed technique (see Figure 4), but not enough according ANOVA ($F = 1.547$, p-value$=$ 0.22 for the two members groups and $F = 0.056$, p-value $= 0.81$ for the others).

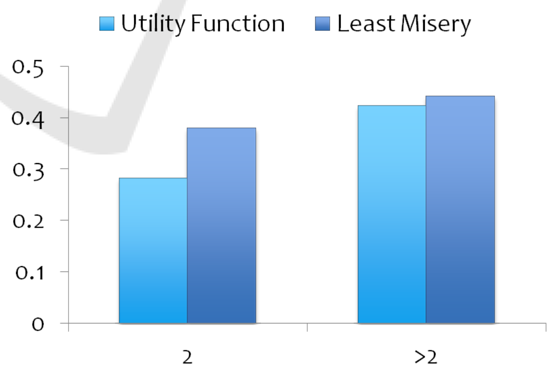

Figure 4: ndcg@3.

1_success@3. We recall that 1_success@3 counts the number of times that the algorithm guessed at least one movie among the three selected by the group. LM is always better than our technique (see Figure 5), but these results provide significant differences only when considering all the 48 groups together ($F = 3.847$, p-value $= 0.05$).

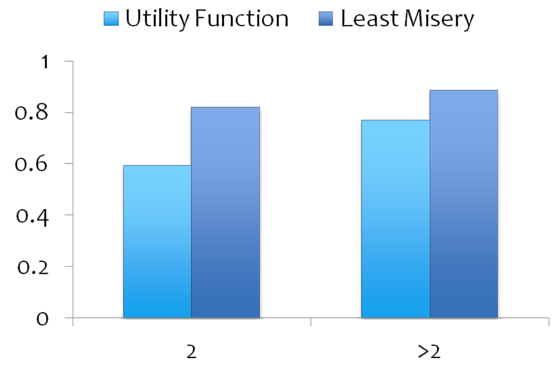

Figure 5: 1_success@3.

2_success@3. In this case, our method was more accurate in groups with more than two members (see Figure 6). Unfortunately ANOVA shows that these differences are due to chance ($F = 0.305$, p-value $= 0.58$).

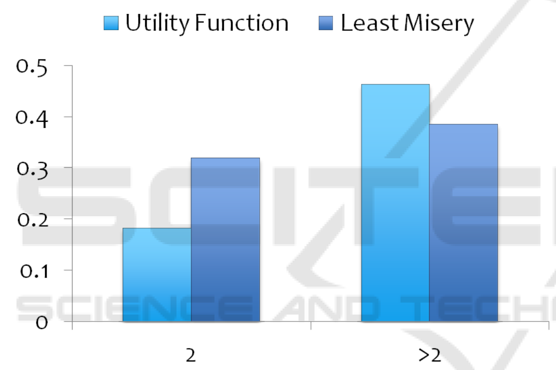

Figure 6: 2_success@3.

3_success@3. It is really rare that a technique is able to guess exactly the three movies chosen by the group. However, in some cases, it happened. Once a time LM wins but with no significant differences.

5.1 Discussion

Groups recommendation systems lack the appropriate dataset to be evaluated upon. On the contrary, user studies have the disadvantages of being expensive to conduct both in term of recruiting the proper users and engaging them, when they are volunteers. In this sense, founding the appropriate number of groups and with varying dimensions is challenging.

As expected, for two users groups LM is the best choice. In the other cases, we cannot say the same thing since the two methods are comparable and show a similar performance. Pearson correlation, evaluated on results distributions shows that the two techniques are not linearly dependent because, in most cases, its value is near zero. This result means that, in some circumstances, the proposed utility function shows a better performance than LM and in other cases the opposite occurs. Therefore, in certain groups, users do care to minimize the group unsatisfaction (i.e., the LM goal) regardless of the *agreeableness* value. Since we fixed the δ value to 0.5, with the aim to give the same weight to least misery and social welfare components of Equation 1, a better tuning of this parameter would results in choices that differentiate most from LM.

Moreover, we think these first results are effected by how we conducted the experiments. As already said, the 10 movies recommended to the groups are obtained by a sort of union of two lists independently generated by the two different techniques. Thus, in this set, there are movies chosen by both LM and the proposed utility function. Maybe, a between-subject experiment (by assigning the result of a single technique to each group) would lead us to different results. Moreover, the low number of experiments had a relevant impact in the significance analysis of the results. Finally, since most of the groups was composed by two or three members, we foresee that, with larger groups, our technique could have obtained better results.

6 CONCLUSIONS AND FUTURE WORK

In this work, we introduced a new method to predict which items are suitable for groups of users, taking into account users' personality. In particular, we evaluated the role of the *agreeableness* factor (e.g., one of the features of FFM), in order to weigh the importance of user's gain with respect to the global satisfaction.

To evaluate this approach, we conducted a pilot user study on movie recommendations, where we compare the results of the proposed approach with respect to a Least Misery strategy (LM). Results showed that for small groups a LM performs slightly better. In particular, for two people's groups LM is the best choice; in the other cases the two methods are comparable and show a similar performance. We foresee, that our utility function will improve its effectiveness proportionally to the group size: the larger is the group, the greater will count altruism in the final decision. Hence, in future works, we should try to encourage users to create larger groups in order to better support our hypothesis.

Finally, we could study how to use other personality factors to build another utility function. We saw that *extraversion* is correlated to the leadership of a group, so in a decision process, it could be very critical. Furthermore, even *openness* could be decisive in such cases, because it can have the same weight of the *agreeableness* in our function. Open people are glad to try new experiences, so they could agree to view a movie for which the recommender system predicts a low value for them, but the opposite for other members.

REFERENCES

Anderson, C., John, O. P., Keltner, D., and Kring, A. M. (2001). Who attains social status? effects of personality and physical attractiveness in social groups. *Journal of personality and social psychology*, 81(1):116.

Brody, L. R. (2000). The socialization of gender differences in emotional expression: Display rules, infant temperament, and differentiation. *Gender and emotion: Social psychological perspectives*, pages 24–47.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly journal of Economics*, pages 817–869.

Costa, P. T. and MacCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual*. Psychological Assessment Resources.

Costa, P. T. and McCrae, R. R. (1995). Primary traits of eysenck's pen system: three-and five-factor solutions. *Journal of personality and social psychology*, 69(2):308.

Donnellan, M. B., Oswald, F. L., Baird, B. M., and Lucas, R. E. (2006). The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, 18(2):192.

Dooms, S., De Pessemier, T., and Martens, L. (2013). Movietweetings: a movie rating dataset collected from twitter. In *Workshop on Crowdsourcing and Human Computation for Recommender Systems, CrowdRec at RecSys 2013*.

Dunn, G., Wiersema, J., Ham, J., and Aroyo, L. (2009). Evaluating interface variants on personality acquisition for recommender systems. In Houben, G.-J., McCalla, G., Pianesi, F., and Zancanaro, M., editors, *User Modeling, Adaptation, and Personalization*, volume 5535 of *Lecture Notes in Computer Science*, pages 259–270. Springer Berlin Heidelberg.

Gartrell, M., Xing, X., Lv, Q., Beach, A., Han, R., Mishra, S., and Seada, K. (2010). Enhancing group recommendation by incorporating social relationship interactions. In *Proc. of the 16th ACM International Conference on Supporting Group Work*, pages 97–106. ACM.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26.

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.

Hu, R. and Pu, P. (2011). Enhancing collaborative filtering systems with personality information. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 197–204. ACM.

Ma, Z. (2005). Exploring the relationships between the big five personality factors, conflict styles, and bargaining behaviors. In *IACM 18th Annual Conference*.

Masthoff, J. (2011). Group recommender systems: Combining individual models. In *Recommender Systems Handbook*, pages 677–702.

McCrae, R. R. and Costa, P. T. (1985). Updating norman's" adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of personality and social psychology*, 49(3):710.

McCrae, R. R. and Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1):81.

McCrae, R. R. and Costa, P. T. (1989). The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of personality and social psychology*, 56(4):586.

Nunes, M. A. S. and Hu, R. (2012). Personality-based recommender systems: An overview. In *Proceedings of the Sixth ACM Conference on Recommender Systems*, RecSys '12, pages 5–6, New York, NY, USA. ACM.

O'connor, M., Cosley, D., Konstan, J. A., and Riedl, J. (2001). Polylens: a recommender system for groups of users. In *ECSCW 2001*, pages 199–218. Springer.

Quijano-Sanchez, L., Recio-Garcia, J. A., Diaz-Agudo, B., and Jimenez-Diaz, G. (2013). Social factors in group recommender systems. *ACM Trans. Intell. Syst. Technol.*, 4(1):1–30.

Rossi, S., Caso, A., and Barile, F. (2015). Combining users and items rankings for group decision support. In *Trends in Practical Applications of Agents, Multi-Agent Systems and Sustainability*, volume 372 of *Advances in Intelligent Systems and Computing*, pages 151–158. Springer International Publishing.

Sager, K. L. and Gastil, J. (2006). The origins and consequences of consensus decision making: A test of the social consensus model. *Southern Communication Journal*, 71(1):1–24.

Salehi-Abari, A. and Boutilier, C. (2014). Empathetic social choice on social networks. In *13th International Conference on Autonomous Agents and Multiagent Systems*, pages 693–700.