Supporting Group Discussions with Recommendation Techniques

Thuy-Ngoc Nguyen

Supervisor: Prof. Francesco Ricci

Faculty of Computer Science
Free University of Bozen-Bolzano

This dissertation is submitted for the degree of
PhD in Computer Science

February 2019
This thesis is dedicated to my beloved family, especially to my Mom & Dad for their undying love and support.

Nguyễn Thúy Ngọc
This thesis marks the end of my PhD, and thinking back on the journey, I am deeply grateful that I have been fortunate to meet, work with and learn from so many people.

I would like to, first and foremost, express my sincere gratitude to my supervisor Prof. Francesco Ricci for a tremendous amount of support, patience, and guidance throughout my entire PhD study. This thesis, undoubtedly, would not have been possible without his inspiration and assistance. I really very much appreciate all the time and effort he has put into providing invaluable and instant feedback on drafts of my research papers. It has always been an honor and intellectually rewarding experience for me to know and work with him.

I would like to extend my appreciation to Prof. Judith Masthoff and Dr. Silvia Rossi, who gave me constructive and helpful comments that greatly help in improving the thesis.

My most sincere thanks go to Dr. Derek Bridge who enabled me to undertake a great month of research at the Insight Centre for Data Analytics – University College Cork. I would also like to express my heartfelt thanks to Dr. Marko Tkalcic for always being willing to support, give academic advice, and encourage me to take up new challenges.

I truly appreciate that my journey has been filled with all of nice and brilliant colleagues and friends, who have brought a lot of joy to my life over the past three years. To all my amazing friends in Bolzano, some left and some new arrived, I owe a big thank you to all of you for your support, encouragement, laughter, and fond memories of coffee breaks, conferences, and hiking trips. To my dear colleagues in TUWien, many thanks for the enjoyable collaborations and the beautiful friendship that we have made.

Finally, I would like to express my eternal gratitude to my parents, my sister Thuy-Quynh, and my husband Duy-Nhat who, always, share with me many highs and lows during the period of study, and live through my ups and downs. Words cannot describe how grateful I am for everything you have done for me. I would not have made my journey this far and become who I am today without your sacrifices, unconditional love and support.
Abstract

Group recommender systems (GRSs) have been developed to support group decision making processes with recommendations expected to satisfy a group of people, not just a single person. Most of the previous research has assumed that users’ preferences are stable and independent, and hence systems solely based on individual preferences, in one shot, can predict a group choice. Unfortunately, in practice, what users choose in a group does not fully match their personal taste as the iterative interaction between group members drives them to adapt their behavior and choices to the group setting. Thus, designing effective GRSs necessitates the dynamic support for the group discussion and the decision making process.

In this thesis, we aim at leveraging interactive recommendation approaches to facilitate group members in discussing and settling on a final group choice. We particularly focus on: (i) the analysis of how people make decisions in groups, in order to build more useful GRSs; (ii) the development of a GRS that employs and integrates group recommendations into the discussion and decision making stage; (iii) an interactive group model that considers preference knowledge elicited before and during group interaction (i.e., respectively the long-term and session-based preferences); (iv) a generic simulation procedure that simulates social impacts on users’ behavior in order to explore the appropriate combination of long-term and session-based preferences in the alternative group scenarios; and finally (v) a novel group discussion simulation process that models agents’ conflict resolution styles in order to investigate their effect on the outcome of the decision making process supported by the GRS.

We have shown in an exploratory user study that the usability score of our system is greater than a standard benchmark and the proposed group model is able to enhance the perceived recommendation quality. The results of various simulation experiments further prove the efficacy of the proposed model in capturing the changes of the users’ preferences, measured in terms of system ranking performance, and indicate that the optimal combination of long-term and the session-based preferences depends on the specific group scenario. Finally, the evidence of our conflict simulation analysis has shed light on how conflict resolution styles are interlinked with other group factors and how they can influence the group recommendation performance, measured by observing the average individual’s loss in utility for selecting the collective choice rather than the personal one, and the difference in utility obtained by the users with the highest and lowest utility in the group.
# Contents

| List of figures | xi |
| List of tables | xiii |

1 **Introduction**

   1.1 Research Motivation ........................................ 1

   1.2 Research Questions and Hypotheses ............................ 2

   1.3 Contributions .................................................. 5

   1.4 List of Publications ........................................... 6

   1.5 Thesis Outline .................................................. 8

2 **State of the Art**

   2.1 Group Recommendation Generation ............................... 9

      2.1.1 Individual Recommendations ................................. 9

      2.1.2 Preference Aggregation Approaches ........................ 13

      2.1.3 Incorporating Social Psychological Concepts .............. 17

   2.2 Group Recommender Systems ...................................... 20

      2.2.1 Non-interactive Systems .................................... 20

      2.2.2 Interactive Systems ......................................... 20

   2.3 Group Decision Making and Observational Studies ............ 24

   2.4 Evaluation of Group Recommender Systems ...................... 26

      2.4.1 User Studies ............................................... 26

      2.4.2 Offline Evaluations ........................................ 26

3 **Observing Group Decision Making Processes** .................... 29

   3.1 Research Motivation and Hypothesis ............................ 29

   3.2 Study Procedure ............................................... 30

   3.3 Results ......................................................... 33

   3.4 Discussion and Conclusions ..................................... 36
4 Building a Chat-based Group Recommender System 41
  4.1 Application Scenario ................................................. 41
  4.2 Recommendation Logic .............................................. 46
  4.3 Evaluation Procedure ............................................... 50
  4.4 Results ................................................................. 53
  4.5 Data Observations and Analysis ................................. 56
  4.6 Discussion and Conclusions .................................. 59

5 Evaluating an Interactive Group Recommendation Model 61
  5.1 Research Motivation ............................................... 61
  5.2 Group Discussion Simulation Model ............................ 62
  5.3 Experiment Setup .................................................... 66
  5.4 Results ................................................................. 68
  5.5 Discussion and Conclusions .................................. 71

6 Simulating Group Discussions in Conflict Situations 73
  6.1 Research Motivation ............................................... 73
  6.2 Group Discussion Simulation Model ............................ 75
    6.2.1 Individual’s Conflict Resolution Style ...................... 75
    6.2.2 Group Discussion Procedure ............................... 78
  6.3 Experiment Setup .................................................... 83
    6.3.1 Independent Variables ........................................ 83
    6.3.2 Dependent Variables .......................................... 84
  6.4 Results ................................................................. 85
    6.4.1 Experiments with homogeneous groups ....................... 86
    6.4.2 Experiments with heterogeneous groups .................... 93
  6.5 Analysis of Real Groups ......................................... 99
    6.5.1 Methodology ..................................................... 99
    6.5.2 Results .......................................................... 102
  6.6 Discussion and Conclusions .................................. 103

7 Conclusions and Future Work .................................. 107

Bibliography ................................................................. 111
List of figures

3.1 The distribution of users’ ratings given to the ten cities ................................ 34
4.1 Group connection ................................................................. 43
4.2 Group discussion ................................................................. 44
4.3 Group decision making support ............................................... 45
4.4 Screenshots from the on-line mobile emulators that were used to collect data 53
4.5 System Usability Scale (SUS) .................................................... 54
4.6 The distribution of individuals’ ratings collected before the group discussions 56
4.7 The distribution of 33rd and 66th percentile user’s utility values ............... 58
5.1 The computation sequence in the simulated group discussion ................ 65
5.2 MRR of the group recommendations ........................................ 68
6.1 An illustration of how the probabilities that a user will give positive and negative evaluations can be transformed with respect to the level of cooperativeness ................................................................. 77
6.2 An illustration of the simulated group discussion process ..................... 78
6.3 MIL of groups with similar interests and uniform conflict resolution styles . 87
6.4 MMD of groups with similar interests and uniform conflict resolution styles 88
6.5 Performance of groups with similar interests and uniform conflict resolution styles, stopping the discussion at the first group choice .................... 89
6.6 MIL and MMD of groups with similar interests and mixed conflict resolution styles ................................................................. 90
6.7 Performance of groups with similar interests and mixed conflict resolution styles, stopping the discussion at the first group choice .................... 91
6.8 MIL and MMD of groups with diverse interests but similar conflict resolution styles ................................................................. 94
6.9 Performance of groups with diverse interests but similar conflict resolution styles, stopping the discussion at the first group choice .................... 96
xii  |  List of figures

6.10 Performance of groups with diverse interests and mixed conflict resolution styles ......................................................... 98
6.11 Performance of groups with diverse interests and mixed conflict resolution styles, stopping the discussion at the first group choice .................................................. 100
6.12 MIL of real groups across uniform conflict resolution styles ...................................................................................... 103
6.13 Summary of the results observed through this simulation study ................................................................................... 104
List of tables

2.1 Overview of the reviewed GRSs ................................................. 23
3.1 Number of participants and groups at each university ...................... 31
3.2 Contingency table: preferences match and excitement ....................... 34
3.3 Performance of aggregation strategies ........................................ 35
4.1 Group recommendation quality .................................................. 55
4.2 Group choice satisfaction and group identification ............................ 57
4.3 Probability of each type of feedback in low, medium, and high utility ranges 59
5.1 Table of ANOVA2 analysis for groups of size 5 ............................... 69
6.1 Summary description of the adopted design for simulating the TKI conflict resolution styles .................................................... 76
6.2 Overview of the employed independent variables .............................. 84
6.3 Pairwise comparisons in terms of MIL for groups of size 2 whose members have similar interests and uniform conflict resolution styles, stopping at the first group choice .................................................. 92
6.4 Pairwise comparisons in terms of MIL for groups of size 2 whose members have similar interests and mixed conflict resolution styles, stopping at the first group choice .................................................. 93
6.5 Pairwise comparisons in terms of MIL for groups of size 2 whose members have diverse interests but similar conflict resolution styles ............... 95
6.6 Pairwise comparisons in terms of MMD for groups of size 2 whose members have diverse interests but similar conflict resolution styles ............ 95
6.7 Pairwise comparisons in terms of MIL for groups of size 4 whose members have diverse interests but similar conflict resolution styles, stopping at the first group choice .................................................. 97
6.8 Group discussion behaviour ....................................................... 102
CHAPTER 1

Introduction

In this chapter, we introduce the motivation and objectives of this thesis, and then we list the research questions and hypotheses that guide our study. Finally, we summarize our contributions and outline the thesis structure.

1.1 Research Motivation

These days with the development of the Internet, a single e-commerce website can offer users up to millions of items in different categories. As a consequence, users may encounter difficulties in filtering out such an overwhelming information content to make decisions and choose suitable items. Recommender Systems (RSs), therefore, have been developed to mitigate such an information overload problem by providing the users with suggestions for items that are most likely to match their needs and interests [78].

Although a lot of remarkable progress has been made to provide high quality recommendations for individual users, in many realistic situations, the recommended items are supposed to be consumed collectively, not individually. For example, a group of friends or a whole family may be looking for a restaurant or a destination to experience together, hence the recommendations are expected to satisfy all group members. These scenarios have led to the research on Group Recommender Systems (GRSs) [47, 60]. Not surprisingly, recommending to groups has been highlighted as a real challenge as the recommendation that suits one user may actually be unsuited for another, considering that people could have different expectations and preferences. As a result, conflict is accepted as an unavoidable and ubiquitous aspect of group life. In fact, there is no clear path to identify optimal or even satisfactory suggestions for a group.

Notwithstanding the ubiquity of group decision making (e.g. in travel and tourism), the subject of GRSs is, undoubtedly, in its infancy, considering the lack of industrial/commercial
systems genuinely dedicated to support the process of making group decisions. That fact is understandable, given that by far most of the research on GRSs is focusing on algorithmic aspects, which often boils down to predicting group choices or ranking group recommendations, and much less work has been devoted to understanding the interplay between decision making behavior of users and recommendation techniques. In particular, numerous research studies are based on the simplifying assumption that users’ preferences are stable and independent (i.e., they are influenced neither by the group decision making process nor the interactions between group members), and hence the systems, with a single short, can predict a group choice by aggregating users’ individual preferences in a fair and acceptable way, such as by suggesting items that average the satisfaction of all group members or minimize the misery of the least satisfied user [59, 62, 69]. This assumption, however, overlooks the dynamics of people’s behavior in a group decision making process wherein what users choose does not completely match their individual interests as they tend to be affected by opinions and reactions of the other group members [61]. As a result, another thread of research on GRSs has tried to model the users’ behavior in group interactions by adopting more interactive recommendation techniques [42, 46, 63]. Nevertheless, in practice, how to materialize the principle of using the interactive recommendations to effectively support the group decision making process that is built up over a series of group interactions, still calls for a foundational study and exploration [25, 78].

In the light of this analysis, the main objectives of this thesis are threefold. First, we aim at facilitating group decision making processes by designing a useful user-system interaction that integrates the recommendation process into the group discussion and decision making stage. In order to accomplish this goal, we model the dynamics of users’ preferences as they evolve in the course of a group discussion by leveraging interactive recommendation techniques. Ultimately, we focus on exploring the impact of different group situations and compositions on the performance of group recommendations.

1.2 Research Questions and Hypotheses

The stated goals lead us to the following research questions and hypotheses:

**How to build a useful GRS supporting group discussion and the decision making process?** Social psychology studies on group dynamics have emphasized the importance of the full decision making process adopted by a group, in determining the quality of the decision outcome, especially the discussion phase in which the information is shared among group members is highlighted as a key element to understand and predict the collective
choice [39, 89]. In the context of GRs, however, it is unknown what kind of support from the systems the users need when facing the task of making a collective choice. We argue that an observational analysis of how people make decisions in groups can pave the way for designing more effective systems. Additionally, group decisions are typically made in several stages, we therefore hypothesize that a GRS must support the iterative process of discussing and reaching a consensus rather than simply offering a list of group recommendations. The key issue thereby is to figure out how the group recommendations can be employed and integrated into the group discussion and decision making process. This question is addressed in Chapter 3 and 4.

How to model users’ preferences in group settings? Empirical studies in GRSs have shown that there is no clear and unique best preference aggregation strategy [60], which implies that a group choice is not dictated entirely by the knowledge of individual preferences prior to the group discussion, and that the group choice is possibly influenced by other factors. Among various possible factors associated with predicting the group choice, we focus on the “process” factor because user preferences often evolve and change over a series of group interactions, and the process of sharing information and opinions within groups is a major determinant of the group decision outcomes [39, 89]. Thus, we conjecture that group members’ preference models should exploit not only the individual long-term preferences, i.e., the ones acquired independently of the group context, but also the group-induced preferences, i.e., those that can be inferred from observing the interactions of the members with a GRS in the course of group discussion. Importantly, we stick to the working hypothesis that the individual preferences of the users must be continuously updated and revised according to their revealed feedback by only observing the items that they like or criticize during the discussion, without explicitly asking their feature-level preferences. This problem is tackled in Section 4.2 of Chapter 4.

How to evaluate the effectiveness of an interactive recommendation model in various group situations? Given an interactive group recommendation model that concerns both individual long-term and session-based preferences, the question that we attempt to unravel is how to properly exploit these two types of preferences. Our hypothesis is that the relative importance of the long-term and session-based preferences varies depending on specific group settings. To validate this hypothesis, we propose evaluating the performance of the preference model with alternative ways of combining those preferences in different group situations. This task, in the context of group recommendations, is complex and multifaceted due to two reasons. First, it is difficult or even impossible for user studies to extensively
test the efficacy of the model in various group situations, especially when it requires repeated interaction settings with a high degree of interactivity. As in traditional RSs, off-line experiments can be used to evaluate the effectiveness of GRSs. These approaches, however, are somewhat hindered by the absence of public data sets that capture the preferences and choices of users in real group contexts. In fact, when it comes to the evaluation of interactive recommendation models, the research literature on RSs has also applied simulation procedures to emulate the behavior of the users while interacting with a conversational RS [55, 93]. For example, in order to assess an interactive RS, user-system sessions wherein a user incrementally modifies a query to finally select or add a product was simulated [55]. We therefore argue that a simulation-based analysis can also be effectively utilized to address the problem of evaluating GRSs, which is investigated in Chapter 5.

How to assess the effect of users’ conflict resolution styles on the outcome of group decision making process? Connecting with the previous research question, this one revisits the purpose of evaluating the performance of an interactive GRS, yet, in a setting where group members who interact with the system have diverse conflict resolution styles. Specifically, according to the Thomas-Kilmann Conflict Mode Instrument (TKI), each conflict resolution style is characterized by two fundamental dimensions, namely assertiveness and cooperativeness, that are, in essence, the extent to which the individual attempts to satisfy their own and other person’s concerns, respectively [49, 88]. So far, the impact of the individual’s conflict resolution style on the outcome of the decision making process supported by a GRS has been unexplored. In fact, existing research on group recommendations has mostly used TKI to weigh the influence of group members, e.g., a person who is more assertive is assumed to have greater influence and hence given greater weight [72, 75]. On top of that, measuring such impact goes beyond the capability of a user study as the effect of the group members’ conflict resolution style can probably be interconnected with other group factors. For instance, it is unclear, what the quality of the outcome will be if group members tend to compete against each other even when they share similar preferences, or conversely, if they have diverse tastes but they treat others’ preferences with the same weight. To this end, we focus on how to improve the proposed simulation model, so that it can help us deepen our understanding of the interplay of individual conflict resolution styles and the group decision making process. The answer to this question is provided in Chapter 6.
1.3 Contributions

Driven by the research questions and hypotheses formulated in the previous section, we hereby summarize the main contributions of the thesis as follows:

**Observing group decision making processes.** Regarding the first research question, we have proposed and analyzed the results of an observational study on how people make decisions in the specific scenario of tourism group decision making. The study design considered: (i) the individual preferences for a small set of travel destinations (before group interactions); (ii) the interactions between group members during the decision making process; and (iii) the individual evaluations of the task and the choices of the group. The experimental results have confirmed that group preferences are constructed during the decision making process, and the analysis of the data acquired by observing users’ interactions in the group decision making task gave important indications for designing more effective GRSs.

**Building a chat-based group recommender system.** With the lessons learned from the observational analysis, we have developed an interactive mobile GRS that is equipped with a chat-based interface and various decision support functions; from recommending new alternatives for a group discussion to suggesting a final group choice among options that are discussed by the group. In this manner, we contribute to understanding how to integrate recommendation technologies into the discussion and decision making stage.

Our next contribution is related to the second research question. Specifically, we have proposed an interactive group recommendation model that harnesses both the long-term and session-based preferences, which is implemented in a GRS to provide an interactive recommendation functionality. The model considers the following stages: before any group discussion unfolds, it captures the individual user’s preferences expressed in the form of users’ ratings given to items and builds a utility function for each user. Then it iteratively updates the inferred utility functions by observing the evaluations revealed during the group discussion. The proposed model also overcomes the difficulties posed by the explicit acquisition of feature-level feedback by eliciting users’ preferences at the item level. Hence, it enables a user, in the context of a group discussion, to express only evaluations (like/dislike) for the proposed items. Based on these evaluations, it automatically infers the importance weight of each single feature for the user. We carried out an exploratory user study to investigate the system usability and the acceptance of the obtained group recommendations. The results of the study are encouraging as the proposed system attains a very good usability score, and it can also increase the user-perceived recommendation quality.
Evaluating an interactive group recommendation model. Another contribution of this thesis is the development of a group discussion simulation approach that is generic and reusable. We used it to test the impact of alternative combinations of long-term and session-based preferences on the performance of the proposed recommendation model in different group scenarios. In particular, we simulate users’ actions in three situations: (i) when the group has no effect on the user preferences; (ii) when the group setting nudges group members to better align their preferences; and (iii) when the group setting causes group members to differentiate even more their preferences. The empirical results prove the goodness of the proposed model in correctly capturing the changes of the user’s needs, and bolster our claim that the appropriate use of long-term and the session-based preferences depends on the specific group scenario.

Simulating group discussions in conflict situations. Finally, we extended the previously mentioned group discussion simulation procedure, so that it can be used to experimentally manipulate different group compositions and analyze the responses of group members as a function of their preferences and their conflict resolution styles. To retain the essential characteristics of a realistic system-mediated group discussion as closely as possible, the simulation model is informed and fine tuned to match the observations that we have previously made on how users interact with a GRS. The analysis of a range of simulated group discussions has contributed to clarify how the outcome of the group decision making process supported by the GRS is influenced by the conflict resolution styles together with various group factors namely, inner-group similarity, interaction length and group size. More concretely, the simulation analysis demonstrates that when the groups are composed of users with competing conflict resolution style then they receive the largest average utility loss for choosing a collective choice instead of the individual ones, but irrespective of the adopted conflict resolution styles, there is no distinct difference in their utility for the group choice.

1.4 List of Publications

We hereby give the list of publications grouped by their type and ordered by the date.

• journal articles:


• conference papers:


• workshop, poster and doctoral consortium papers:


1.5 Thesis Outline

The remaining chapters are structured with respect to the contributions of the thesis.

- Chapter 2 surveys related research including: a) the prevalent approaches to generate individual and group recommendations; b) state-of-the-art GRSs; c) the background of social psychological concepts used in our study; and d) the different methods for evaluating GRSs.

- Chapter 3 gives a summary of the observational study procedure and its key findings. It further discusses the lessons learned and the implications for designing GRSs.

- Chapter 4 presents the interaction design and functionality of the interactive mobile GRS, called South Tyrol Suggests for Groups (STSGroup), that we have developed as a prototype system for testing our research hypotheses. Following that is a detailed description of the proposed group recommendation model. The chapter also provides the analysis of users’ interactions with the system and the experimental results obtained from the exploratory user study.

- Chapter 5 describes the simulation procedure that we have designed to extensively assess the performance of our group recommendation model in the simulated group scenarios. It further illustrates and explains the results of the simulation experiment.

- Chapter 6 introduces a group discussion simulation process wherein we model the dynamics of users’ behavior in alternative group compositions, characterized by the conflict resolution styles, inner-group similarity, interaction length and group size. The chapter highlights new insights into the effect of the different group compositions on the outcome of the group decision making process supported by interactive GRSs.

- Chapter 7 summarizes our findings and ends with an outlook for future research.
STATE OF THE ART

In this chapter, we first provide an overview of the most recent approaches to generate group recommendations. Next, we shift the focus onto fully functioning GRSs that support both the preference elicitation and recommendation process. We then provide the necessary background on social psychology and observational studies on the group decision making process. Finally, we discuss different methods for evaluating the effectiveness of GRSs.

2.1 Group Recommendation Generation

This section covers widely used techniques for: (i) generating individual recommendations; (ii) aggregating individual preferences or recommendations in order to generate collective recommendations; and (iii) incorporating social psychological concepts to improve group recommendation quality.

2.1.1 Individual Recommendations

Many recommendation techniques have been proposed for supporting individual decision makers; they can be broadly classified into four categories: content-based filtering, collaborative filtering, knowledge-based and hybrid systems [1]. With a different angle of view, one can differentiate RSs by the extent to which users engage in interactions with the systems, resulting in single-shot and conversational recommendations [74]. We also discuss the difference between the task of context-aware and group recommendations.

Content-based Filtering

Content-based Filtering (CB) recommends items that are similar to those that a given user liked in the past [54, 71]. The core idea behind CB is to match the features (i.e., attributes)
of the target user profile, the representation of the user’s preferences and interests, with the features of the items. Typically, systems implementing a CB technique represent items as a set of features that are extracted from items’ meta-data or descriptions, and then they learn the profile of the target user by analyzing the features of the items that he or she rated previously. Based on the resulting profile, various machine learning techniques can be applied to predict which items would interest the user.

The main advantage of CB techniques is that it does not suffer from the new item problem, i.e., the difficulty of accurately recommending items that are not yet rated by any user. Hence, CB techniques are often preferred in domains where it is necessary to recommend brand-new items or where data sparsity is high, but having said that, they still encounter the new user problem, i.e., not being able to generate reliable recommendations for new users who have not yet provided enough information about their tastes. The real drawbacks of CB techniques, however, are the necessity of domain knowledge and the lack of serendipity, i.e., users are often recommended with unsurprising items having similar content to those already rated [54].

**Collaborative Filtering**

Collaborative Filtering (CF) techniques base their recommendations on the “wisdom of the crowd” since they rely on the opinions of like-minded people who share their tastes with a target user, referred to as user-based CF algorithms [77]. A variant of the CF technique is item-based CF, which generates recommendations based on the similarity between items rather than between users [36]. The underlying idea of how user-based and item-based CF algorithms work is similar, so in the next lines we will focus on user-based CF. More in general, CF can be divided into memory-based and model-based methods [1]. The former, at request time, makes predictions based on all the ratings given by the user’s nearest neighbors which are often identified with similarity metrics such as Pearson correlation coefficient or cosine distance. The latter, by contrast, uses the historical information about the users’ ratings as a training data set to build a predictive model beforehand. The learned model is later used to generate predictions of a user’s preference for an item, and it is retrained periodically. So far, the most prevalent model-based technique is latent factor model such as Matrix Factorization in which users and items are characterized by vectors of latent features inferred from historical user-item preference patterns [52].

CF is probably the most widely adopted technique due to the fact that it requires no prior knowledge of the application domain, and it often outperforms other techniques when sufficient ratings are provided. Nevertheless, the cold start problem, i.e., the poor recom-
2.1 Group Recommendation Generation

Recommendation quality that is observed when there is a lack of historical information about users’ ratings, remains the main issue of CF techniques.

**Knowledge-based Systems**

The knowledge-based (KB) approach attempts to make a matching between collected user requirements and the available options. If no solution is found then it automatically repairs inconsistent requirements based on extracted domain knowledge and usually provides reasons for the generated recommendations. Two important classes of KB approaches are case-based [21, 86] and constraint-based methods [38]. Basically, case-based and constraint-based approaches differ in the way they generate recommendations. Case-based methods store and retrieve previous recommendation sessions, so-called cases, and then reuse the similar observed cases to recommend appropriate items. On the other hand, constraint-based approaches mainly exploit predefined rules for relating user requirements to item features (e.g., filter rules or incompatibility rules).

The major strength of KB approaches is their flexibility to incorporate various types of rules, e.g., decision rules or business rules, into the recommendation process. Additionally, they do not suffer from the cold start problem as the users’ needs are elicited within the recommendation process. However, KB methods require a great effort to build in terms of knowledge extraction, representation and system design for a specific domain.

**Hybrid Systems**

Hybrid RSs combine multiple knowledge sources coming from different RS techniques. They are designed to gain better recommendation performance by overcoming the drawbacks of a single recommendation technique in isolation [22]. For example, a hybrid RS might exploit both intelligence of the crowd and item features by combining CF and CB approaches. There are several possible methods for hybridizing various sources of recommendation components, such as applying weighted linear combination to combine different recommendation components, or presenting different recommendation components side-by-side in a combined list [22]. In that sense, the primary issue of hybrid RSs is to identify the most appropriate knowledge sources and recommendation techniques for a given task and how they can be most effectively combined.

**Single-shot versus Conversational Recommendations**

Recommendation techniques can be distinguished by the extent to which users engage in the recommendation process. In a *single-shot* fashion, users are presented with a single
set of recommended items based on the historical information about their preferences and after that, the recommendation process usually ends, which means the next user request is treated independently of the previous ones. Conversely, conversational RSs are interactive, so that users can better elaborate their requirements in the course of the interaction with the systems by giving feedback on the recommended items, which is considered in the next set of recommendations [74]. The benefit of interactive (i.e., conversational) approaches lies in the fact that they help users to navigate the item space as efficiently as possible, i.e., to find the suitable items with the fewest interaction cycles. Moreover, such interactive systems assist users in clarifying their own preferences over a series of recommendation cycles, which is especially useful for those who are unable to specify their preferences from the outset.

Critiquing is a form of feedback mechanism that has been widely adopted in numerous interactive systems since it supports naturalistic negotiations wherein the systems alternatively recommend items and elicit user feedback in terms of critiques on specific features of the recommended items, so that the recommendations are revised based on the recent users’ preferences for specific features [27, 66]. For example, “I would like a cheaper restaurant” is a critique pointing out an unsatisfied preference for the “price” feature and asking for a new suggestion with a lower price. Or, “I would like a hotel with a swimming pool” is a critique confirming the importance of feature “swimming pool”. Simply put, with critiquing, the “recommend - review - revise” process is repeated until the desired item is found, hence the system can better capture the user’s needs and preferences to improve the recommendation quality. In the critiquing literature, there are two common feedback mechanisms: unit and compound critiques [76]. The standard form is unit critiquing that enables users to critique a single feature at a time, whereas the generation of compound critiquing allows users to provide feedback on multiple features with a single critique.

Context-aware versus Group Recommendations

Conventional RSs solely base the recommendations on the knowledge of user-item interactions like users’ ratings or clicks and ignore the fact that users interact with the systems in a particular context such as in a particular location or time of the day. However, contextual information has proven to be valuable for improving the effectiveness of the systems. For example, a user may watch action movies when being alone and watch comedy movies when being with his or her family. Arguably, the presence of the family or the social context, in that example, should play a crucial role in determining the movie recommendation. Therefore, Context-Aware Recommender Systems (CARSs) have been developed to generate more
relevant recommendations by adapting recommendations to the specific contextual situation of the user [2].

Unlike individual RSs that try to generate personal recommendations for a single person, Group Recommender Systems (GRSs) shift the focus onto providing collective recommendations, i.e., recommending items to groups of users whose preferences can possibly be different from each other. The recommended items, in this case, are expected to satisfy the preferences of all group members as much as possible. Group decision making problems are ubiquitous in real-life situations, so the applications of group recommendations can be found in a wide range of domains, from TV programs [97], movies [69], music [62], restaurants [42], tourism [5, 63] to social network websites [83].

Although CARSs can regard the information about which group is going to consume the recommended item as a kind of companion context [19], the general techniques applied for CARSs are unable to directly solve the group recommendation problem because of two general reasons. First, the variety of all groups can produce a context variable that contains a large number of possible contextual values, since the context involves not only individual-level factors, such as personal preferences or characteristics, but group-level factors like group compositions or interpersonal relationship as well [46, 32]. Therefore, to apply general CARS techniques one is expected to enter into the structure of this variable and acquire users’ ratings for each possible context, which is nearly impossible. Moreover, what makes the group recommendation task unique and challenging is that the satisfaction of an individual is also depending on that of the other people in the group [61]. The members, for instance, can probably adjust their preferences to accommodate those of other members. Consequently, most of the research on group recommendations so far has been focused on very simplified and specialized cases.

2.1.2 Preference Aggregation Approaches

To generate group recommendations, the traditional research activities in GRSs have mostly revolved around the problem of identifying the aggregated recommendations that can satisfy, as much as possible, individual preferences of group members, i.e., typically in the form of item ratings given by the users. Particularly, with respect to the moment of aggregating individual data, preference aggregation techniques can be categorized as recommendation (i.e., prediction) aggregation and profile aggregation [60]. The former generates individual recommendation lists for each member and then combines those lists to form a single set of recommendations for the group. Similarly, it can also compute individual rating predictions for each user and aggregate these predicted ratings to identify the top scoring items for the group [47]. The latter, in the first place, aggregates the profiles of all group members to
build a joint profile and then creates recommendations based on that joint profile. Besides, Berkovsky and Freyne [14] proposed also the combination of the prediction and profile aggregation, called hybrid switching. It is yet unclear which of the techniques should be preferred as previous studies have compared these techniques in the food and movie domains and it was observed that the winning approach varies depending on the considered domain [14, 17, 29]. For example, in the food recommendation domain it was shown that profile aggregation outperforms prediction aggregation [14], but in the movie recommendation scenario, the experimental results showed the opposite [17].

Whichever technique is applied, i.e., either recommendation, prediction or profile aggregation, how individual inputs can be aggregated to reach a collective consensus is a fundamental topic, leading to various preference aggregation strategies proposed in the literature of GRSs so far.

Social Choice Theory

Overall, most of the available preference aggregation strategies are derived by Social Choice Theory that studies how a final aggregated ranking list can be achieved, given diverse individual ranking preferences on the same set of choices [6], e.g., how the ranked list of candidates in elections, provided by each voter can be combined to arrive at a single ordering. According to the Arrow’s impossibility theorem [6], there does not exist an optimal voting method that is fully fair, i.e., the one that can satisfy a few fundamental requirements altogether. As a result, it requires a variety of aggregation strategies to be in place.

In the context of group recommendations, Masthoff provided a comprehensive overview of various aggregation strategies that have been derived from Social Choice Theory [59, 60]. In the following we give a brief description of several commonly used strategies, in which each option is assigned a joint score and the option with the largest score is considered the collective choice.

- **Plurality Voting.** Each user votes for his or her most preferred option, and the one with the majority of votes wins.

- **Utilitarian Strategy.** This can be done in multiple ways:
  - **Additive:** ratings of group members for each option are added, and the one with the largest sum wins.
  - **Multiplicative:** ratings of group members for each option are multiplied, and the one with the largest product wins.
• **Borda Count.** Points are awarded to each option according to its position in the individual’s preference list: the last option gets no points, the second last gets one point, and so on. To obtain the group preference ordering, the individually awarded points are added up, and the option with the largest number of points wins.

• **Approval Voting.** Counts the individuals with ratings for an option above an approval threshold and the option with the largest number of approvals is selected (e.g. 5).

• **Average.** The average of the individual ratings is computed and the one with the highest score wins.
  
  – **Average without Misery.** The average of the individual ratings is computed, but without items scoring below a certain threshold.
  
  – **Weighted Average.** The average of the individual preferences is computed, but with the weights reflecting the importance of the individuals.

• **Least Misery.** The minimum of the individual ratings is a group score as it maximizes the utility of the least happy member.

• **Most Pleasure.** The maximum of the individual ratings is a group score as it maximizes the utility of the most happy member.

• **Fairness.** Individuals take turns to receive their preferred options.

Each aggregation strategy has its own strength and weaknesses as it is aimed at optimizing different criteria. The pure **Average** method, for instance, pursues fairness by equally considering all the group members’ interests. **Least Misery** strategy, semantically, is useful for cases where some members have extreme preferences that can act as a veto (e.g., a vegetarian cannot eat meat). Likewise, we have **Plurality Voting with Veto** that allows a member to prevent an item from being selected, even when a majority of group members have accepted it [57]. **Borda count** focuses on the difference in a ranking list of alternatives given by each group member, rather than the scores themselves while **Multiplicative** strategy tries to amplify the distinction between the scores by multiplying them. The detailed discussion about the use of alternative strategies can be found in [59]. Overall, **Average** and **Least Misery** have been widely adopted in group recommendation studies [4, 46], but the results obtained from empirical experiments showed that there is no superior strategy among the proposed aggregation strategies, and the choice of which strategy to use depends on characteristics of specific domains, tasks and group scenarios [60].

Considering the efficacy of the predicted recommendation ranking rather than the accuracy of the predicted ratings, Baltrunas et al. investigated the performance of different rank
aggregation strategies (e.g., Spearman footrule rank and Borda count) [8]. Specifically, by using simulated data of user groups, the authors compared the goodness of individual and group recommendations in terms of Normalized Discounted Cumulative Gain (nDCG) to understand when the group recommendations can be more effective than the individual ones. The empirical results showed that this happens when the individual recommendations are not particularly good, i.e., when the RSs is not able to make good personalized recommendations. It is worth noting that the experimental evaluation was conducted under the assumption that the preferences of users are stable when being in a group. Without this assumption, there is no stable ground truth to directly make the comparisons between the effectiveness of individual and group recommendations.

**Alternative Aggregation Strategies**

Motivated by the ideas of a statistical dispersion and auction negotiation, Salamó et al. [81] introduced two alternative aggregation strategies, called “purity” and “completeness” to reach a consensus on the group recommendations. The purity of an item measures how many positive preferences are given to that item, considering the preferences of all group members, and then the deviation from the mean of the members’ preferences is used to measure the dispersion. On the other hand, the goal of measuring the completeness of an item is to favor high group members’ individual satisfaction score according to their preferences while penalizing big score differences between the members.

Some group recommendation studies have addressed the preference aggregation problem by leveraging Multi-agent Systems (MAS) and their negotiation mechanism. In the work of Carvalho and Macedo [23], a group decision process, users and items are treated as a non-cooperative game, game players and game actions, respectively. Then the payoff function is the predicted satisfaction of each member from the items chosen by other members in the decision process. The feasible recommendation is viewed as the Nash Equilibrium of the game. This approach is similar to the classical methods in GRSs, considering that users’ preferences are assumed to be always stable, i.e., they do not change when users are in different groups. To deal with more sophisticated user behaviors in groups, a line of studies derived from MAS models group members as user agents who can adopt different behaviors during the negotiation process. For instance, Rossi et al. [80] modeled users’ actions associated with their conflict resolution styles obtained through the Thomas-Kilmann Conflict Mode Instrument (TKI), a model that has been formulated in the attempt to rationalize how a person behaves in conflict situations [49]. Or, in the work of Villavicencio et al. [94], the user agents carry out a cooperative negotiation process based on Monotonic Concession
Protocol that defines the rules of the agreement criterion, which agent makes the concession in case of no agreement reached, and how much that agent should concede.

Furthermore, the issue of aggregating individuals’ preferences can be tackled with Multi-Attribute Utility Theory (MAUT) that is used to aggregate different criteria into a function needed to be maximized [35]. Concretely, Stettinger et al. [87] employed Group-based MAUT that represents each member’s MAUT value as the weighted average of the individual ratings given to the features of a specific option, and finally calculate the sum of all these individual MAUT values to rank that option.

2.1.3 Incorporating Social Psychological Concepts

When it comes to making decisions in groups, there are numerous potential factors that might govern people’s satisfaction as well as their choice [39, 60]. Therefore, social psychological concepts of users’ behavior in groups are typically used as a source to determine what factors influence the decision outcomes. An important line of studies on GRSs has been dedicated to identifying these influential factors and integrating them into the systems.

Personality

Personality refers to an individual’s distinct characteristics in terms of emotional, interpersonal, experiential, attitudinal, and motivational patterns [64]. It is known, according to research studies in social psychology, to be strongly correlated with the behavior of the user in a group decision making process and its outcome [39].

In the scope of this work, we restrict ourselves to the TKI model of personality as it focuses on the interaction between group members in a conflict situation, rather than on the characteristics of individual users themselves [90]. According to the TKI model, the behavior of an individual, in a conflict situation, can be described by two dimensions: assertiveness and cooperativeness that are respectively, the extent to which the individual attempts to satisfy the interests of themselves and other people. These basic dimensions of behavior are used to define five different conflict resolution styles that people can follow: competing (assertive and uncooperative), accommodating (unassertive and cooperative), avoiding (unassertive and uncooperative), collaborating (assertive and cooperative), and compromising (moderately assertive and cooperative). The conflict resolution style of a user is typically assessed by using questionnaires [49] or derived from the user’s Big Five personality traits [96], composed of five factors, i.e., Openness to new experiences, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.
The role of conflict resolution styles, in the context of group recommendations, has been investigated only to some extent. Recio-Garcia et al. [75] took into account the group members’ personality, modeled according to TKI, to weigh the influence of their preferences during the recommendation process; the stronger influence was given to the more assertive members. The authors tested their approach in the movie recommendation domain by comparing different group recommendation algorithms with and without considering the conflict mode weight. The results showed that the accuracy of the group recommendations was improved when using the conflict personality factor. This approach was extended by Quijano-Sanchez et al. [72], where the authors considered both personality strength and social trust between group members to formulate an influence-based rating prediction. The predicted preferences of individuals in a group context, hence, depend not only on their personality strength but also on their trust to the other group members. Unlike the previous studies that treat the personality traits as weights in the aggregation process, Rossi et al. [79] defined an individual utility function that takes into account the extent to which someone is altruistic, which derives from their level of Agreeableness. A pilot user study conducted on movie recommendations showed that in case of large group sizes, the group recommendations generated by the proposed utility function are better than those produced by the Least Misery strategy in terms of prediction accuracy, but this does not hold for groups of size two. In another work, Rossi et al. [80] proposed a negotiation mechanism for group recommendations, where the behavior of agents representing group members, is determined by the conflict resolution styles in the TKI model. Concerning opinion shifting, Barile et al. [10] with a user study, indicated that there is a direct correlation between the presence of conflicts and opinion shifting; i.e., in the presence of peaceful setting, the initial opinion of an individual shifts towards the opinion of another person. On the contrary, the presence of conflict leads to the opposite effect: individual’s opinions drift further apart.

User Roles and Social Relationships

In the work of Ardissono et al. [5], intra-group roles such as children and the elderly were considered, and based on these characteristics, group members were divided into homogeneous subgroups. Recommendations were built by using the weighted average of the importance given by each subgroup to items. Berkovsky and Freyne [14] also exploited the role of member within a family as weights in the preference aggregation process. The influence of users, according to Quintarelli et al. [73], was defined by comparing an individual choice of a user with previous choices of the groups in which the user participated. It is based on the assumption that the more often the individual preference of a group member is selected as the collective choice, the higher the influence of that member. For example, if the user was
part of four different groups and his or her selection was chosen as the final group choice in three out of four cases, then the influence of the user is the ratio of three to four.

Regarding social relationships, Gartrell et al. [40] utilized three group characteristics: relationship strength, expertise level, and dissimilarity among group members’ preferences to propose a heuristic group consensus function, in which the relationship strength is used to determine the aggregation strategy, the expertise level is for weighing group members, and the dissimilarity is used to modulate the consensus score by either penalizing or rewarding the preference differences. Barile et al. [10, 11] also investigated the impact of the social relationship, which was measured by the frequency of users’ interactions (e.g., weak, intermediate and strong), on opinion shifting. Their experiments showed that a general positive shifting occurs when the social tie gets stronger. To address the consensus problem in social networks, Salehi-Abari and Boutilier [82] introduced a social choice mechanism wherein the preferences of an individual depends on their own intrinsic utility and their empathetic utility deriving from the satisfaction of their neighbors in the network. The recommendation task is then translated into a weighted form of traditional preference aggregation taking into account local relationships with the neighbors. In the tourism domain, Christensen et al. [28] built a group profile by analyzing three sources of information: user demographics, users’ preferences on item features, and social relationships between users. The social relationships, acquired by explicitly asking group members, are grouped into four categories: close relationship, hierarchical relationship, acquaintance and unknown.

With the concept of social influence, Masthoff and Gatt [61] introduced several satisfaction functions derived from analyzing emotional contagion, the process of being affected by other members’ emotion, and conformity, the act of adjusting one’s judgment towards the majority. In their model, the satisfaction of one group member depends not only on the extent to which the user’s personality is susceptible to emotional contagion but also on the user’s relationship with each other member. In particular, four basic types of social relationships were studied, i.e., communal sharing (user’s satisfaction is positively affected by the satisfactions of the other members), authority ranking (user’s satisfaction is influenced by the satisfactions of the members who have higher priority), equality matching (users’ satisfactions are independent), and market pricing (user’s satisfaction is negatively affected by the other members’ satisfactions). Recently, Delic et al. [31] have delved into the effect of social relationship and social centrality as the respective characteristics of a group and group members on the different outcomes of group decision making process, e.g., group choice and group satisfaction.
2.2 Group Recommender Systems

Similarly to individual RSs, GRSs can also be classified as non-interactive and interactive systems, referred to as passive and active user-system interaction [60].

2.2.1 Non-interactive Systems

A GRS is considered non-interactive, if it adopts a single-shot recommendation strategy, returning a single set of items to the group of users. Simply put, the system bases its recommendations entirely on information about what users have liked or disliked in the past and do not elicit any additional users’ information in the group context. This type of systems is somewhat appropriate for low-risk item domains where the members are less likely to have strong constraints and making a wrong choice can only cause a small loss, e.g., it is more popular in domains of music or books.

In this direction, we can cite MusicFX [62], a GRS that determines music suited to a group of people working out together in a fitness center; Let’s Browse [53] that provides web page recommendations for group members; PolyLens [69] which recommends movies to small groups of users with similar interests; and Intrigue [5], a GRS suggesting sightseeing destinations for heterogeneous groups of tourists.

2.2.2 Interactive Systems

In interactive GRSs, users are actively engaged in a group recommendation process and their incremental feedback elicited from group interactions is used to revise the recommendations. In the travel and tourism domain, interactive GRSs have great potential because of two reasons. First, most of the time users travel in groups, of various size and nature, such as a group of friends, so they hardly know exactly the preferences of each other at the outset. Moreover, when it comes to tourism products, a choice is typically composed of several more elementary ones (e.g., transportation, activities), and it can also be evaluated along several different dimensions, aka criteria. Hence it is evidently difficult for users to articulate all their preferences at once and to make a decision in a one-shot fashion. In these cases, interactive systems are perfectly suited to help group members construct and adjust their preferences over a series of recommendation sessions. This section, thereby, is devoted to interactive GRSs in the tourism and travel domain.

Travel Decision Forum [46] tries to recapture the flavor of a face-to-face interaction by using animated characters representing group members. The system helps group members in reaching an acceptable organization of a vacation and supports their asynchronous com-
munication. The group members are also allowed to interact with a character representing a mediator who directs the interaction between the group members.

Also exploiting agents acting on behalf of group members, *Trip@dvice* [12] adopted co-operative negotiation methodologies to support the group decision making process. Particularly, individual recommendations are first generated and considered as proposals coming from the group members. Then, among these proposals, the negotiation mediator applies one of the aggregation strategies, such as *Average* or *Least Misery*, to offer an option that is regarded as an agreement.

Exemplifying the critiquing technique, *Collaborative Advisory Travel System* (CATS) [63], is a GRS that helps a group in planning a skiing vacation. The system enables each group member to share and express their personal preferences as critiques for specific item features such as price, car parking, or ski room. With a critique, users can reject a feature value (in case of nominal features), or increase/decrease a feature value (in case of ordinal features).

*Where2eat* [42] is a mobile application that offers restaurant recommendations for groups of users. Specifically, it implements interactive multi-party critiquing, an extension of the basic critiquing technique to a computer-mediated conversation between two users. The novel aspect of this work is to provide real-time negotiation support to groups of two members, which enables them to make proposals and counter-proposals generated by critiquing.

Considering that users can discuss not only recommendations but also their features, *Hootle+* [58] is a system that assists a group of users in selecting hotels. It is designed to support a discussion and negotiation on features of a desired hotel. The group members can define and propose features that the item collectively selected should possess. They can also accept or reject the discussed features as well as individually assign an importance weight to each of them. The system then provides and refines group recommendations based on the outcome of the self-managed negotiation process.

With a wider scope and functionality, *Choicla* [87] is a group decision support environment allowing users to flexibly configure the process of a decision task in a domain-independent setting. For example, a group can decide to use *Plurality Voting* for selecting which movie to watch and *Least Misery* for choosing a restaurant to dine out. The system makes use of different aggregation strategies as well as Group-based MAUT to rank the recommended items in the result sets.

As we noticed, most of the cited interactive GRSs elicit user preferences at the feature level. The feature level critiquing yet tends to put a cognitive burden on users who are expected to determine a specific value for an item feature, and this requires a great effort especially when the number of features is large [50, 67]. Moreover, so far the majority of the reviewed systems only consider the long-term (individual) preferences and disregards
the session-based (group-induced) preferences that evolve during group discussions. Table 2.1 gives an overall summary of the reviewed GRSs that are classified according to four dimensions: (i) group recommendation approach (profile or recommendation aggregation); (ii) whether the recommendation is interactive or not; (iii) type of user’s preferences used in each system (long-term or session); and (iv) adopted preference aggregation strategies.

In summary, the observed advantages and drawbacks of interactive GRSs motivate the development of a model that derives users’ preferences from their evaluations acquired at the item level, i.e., automatically inferring the importance of item features for each user based on what items they like and dislike. Besides, user and group profiles are expected to be pivotal elements not only in how they model the long-term individual preferences, but also in how they take into account the preferences that can be derived from observing a series of group interactions.
### Table 2.1 Overview of the reviewed GRSs

<table>
<thead>
<tr>
<th>System</th>
<th>Rec. approach</th>
<th>Interactive</th>
<th>Preferences</th>
<th>Aggregation strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MusicFX</strong></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>Average without Misery</td>
</tr>
<tr>
<td><strong>Let's Browse</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Average; Fairness</td>
</tr>
<tr>
<td><strong>Polylens</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Least Misery</td>
</tr>
<tr>
<td><strong>Intrigue</strong></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>Weighted Average</td>
</tr>
<tr>
<td><strong>Travel Decision Forum</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Average</td>
</tr>
<tr>
<td><strong>Trip@device</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Average; Least Misery; Most Pleasure</td>
</tr>
<tr>
<td><strong>CATS</strong></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>Average without Misery</td>
</tr>
<tr>
<td><strong>Where2eat</strong></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hootle+</strong></td>
<td>✓</td>
<td></td>
<td></td>
<td>Plurality Voting with Veto</td>
</tr>
<tr>
<td><strong>Choicla</strong></td>
<td>✓</td>
<td></td>
<td></td>
<td>Plurality Voting; Least Misery; Most Pleasure; Group-based MAUT</td>
</tr>
</tbody>
</table>
2.3 Group Decision Making and Observational Studies

While up to now, a few studies on GRSs have explored the issue of how group members make choices and how the process of making choices can be supported [25], a substantial amount of research on group decision making process has been conducted in the field of social psychology.

An extensive study on group dynamics including the influence of the different aspects, from group compositions to group decision process structure, on the group choices is comprehensively covered in [39]. Specifically, social scientists studying the functional theory of group decision making have pointed out that groups who adopt the full decision process consisting of the four phases of Orientation, Discussion, Decision and Implementation (ODDI model) are more likely to make better decisions [39]:

1. **Orientation**: a group defines important aspects and goals of the decision making process, such as the problem needed to be solved, strategies and procedures used to reach a decision.

2. **Discussion**: group members discuss possible alternatives, exchange collected information, and express opinions.

3. **Decision**: based on the previous phases, the group makes a choice using a decision scheme, e.g., voting or consensus reaching. If the decision cannot be reached the group can return to any of the previous phases.

4. **Implementation**: a decision is implemented and evaluated.

Among these phases in the ODDI model, the **Discussion** step is regarded as a crucial part of the process as it is the place where group members can exchange information, process it thoroughly, and rely on it to arrive at the final decision. In line with that finding, Tindale and Kemeda [89] emphasized and demonstrated that “social sharedness”, the extent to which preferences or information is exchanged and shared within groups, is an integral element to understand the group decision making process.

To perform an observational study and record interactions within small groups, there are many existing approaches. One of them is the Interaction Process Analysis (IPA) proposed by Bales [7, 39]. IPA is a coding method for observing group interactions and it is widely used as it increases the objectivity of observations. The approach requires an observer to identify a “unit” of interaction for each group member. According to a basic definition proposed by Bales, a “unit” can be a single simple sentence or its equivalent. In case of a complex sentence containing an independent clause and at least one dependent clause, or a
compound sentence joined by “and”, “but” and so on, it should be broken down into a single expression “unit”. In addition to speech, a “unit” of interaction comprises facial expressions, gestures, body attitudes, emotional signs, etc. Then, for each group member, the observer categorizes each “unit” of interaction into one among twelve behavior categories:

1. Show solidarity/“Friendly” (e.g., expressing gratitude or appreciation, apologizing, or smiling directly at another, etc.)
2. Show tension release (e.g., showing satisfaction, enjoyment, relish, pleasure, etc.)
3. Agree (e.g., agreement reflected through verbal or nonverbal expressions)
4. Give suggestion (e.g., mentioning a problem to be discussed)
5. Give opinion (e.g., stating judgment or inference)
6. Give information (e.g., reporting factual, verifiable observations or experiences)
7. Ask for suggestion (e.g., requesting guidance in problem-solving process)
8. Ask for opinion (e.g., questions seeking value judgment, beliefs or attitudes)
9. Ask for information (e.g., questions requesting a simple factual, descriptive or objective type of answer)
10. Disagree (e.g., rejecting another person’s statement)
11. Show tension (e.g., appearing startled, blushing, showing embarrassment, etc.)
12. Show antagonism/“Unfriendly” (e.g., attempting to override the other in conversation, interrupting others, making fun of others, criticizing, tricking, etc.)

The IPA system enables qualitative analysis of each group member’s behavior to be classified and quantified in a clear manner. For example, we can observe certain specific types of a user’s actions within a group and measure the number of occurrences of these actions, e.g., how many times the user expresses opinions or makes suggestions during the group discussion. A concrete application of the IPA system can be found in the observational and simulation studies presented in Chapter 3 and 6 of this thesis.
2.4 Evaluation of Group Recommender Systems

In order to evaluate the effectiveness of a group recommendation model, both user studies and offline experiments have been conducted [92]. Nevertheless, how to properly assess a GRS is still one of the major research topics [30, 60]. We discuss below two approaches.

2.4.1 User Studies

Basically, a user study is employed when the criteria used to measure the system performance are related to system usability and user experiences (e.g., perceived user’s satisfaction or recommendation quality) [14, 42, 58, 61, 87]. A user study can be carried out by directly interviewing participants about their experiences in their groups, and it is also possibly performed through crowd sourcing sites, such as, Amazon Mechanical Turk [4]. Nonetheless, measuring users’ satisfactions is difficult, considering that the satisfactions of users are interdependent with other factors such as users’ relationships or users’ emotions, so the measurement is needed to be examined under different group conditions. Thus, an indirect study has been employed to have more control over the user study design [10, 59, 61]. In this approach, real groups were not composed, hence the users were required to imagine that they were taking part in a fictional group instead. For example, to analyze the effect of social relationship between group members, Masthoff and Gatt [61] conducted an indirect experiment in which participants were asked to think of a member with whom they have a certain type of relationship, and then, they were requested to rate how their emotions would be affected by those of the imaginary member. The issue here is that the more fictional tasks are required, the less the study reflects the real world. Overall, employing user study cannot be the unique approach to evaluate the efficacy of a GRS since it is really problematic to extensively test the system performance across numerous different group compositions, especially when dealing with a high degree of interactivity.

2.4.2 Offline Evaluations

As in classical RSs, offline evaluations have also been considered solutions for GRSs. These approaches, yet are somewhat hindered by the absence of public data sets that capture the preferences of users in real group contexts. To this end, generating synthetic groups that are sampled from standard data sets (e.g., MovieLens or Netflix) is commonly applied [8, 29]. In particular, Baltrunas et al. reused the MovieLens data set to generate artificial groups with different sizes and different levels of inner-group preference similarity [8], i.e., groups with highly similar members are supposed to represent people with common tastes while
the low similarity groups represent members without any social relations, such as random people taking the same bus. So far, there are two options to evaluate the effectiveness of group recommendations:

- Comparing the generated group recommendations with the joint group evaluations for the recommended items, which are deemed to be the ground truth of group choices (i.e., the “true” group preferences). These “true” group preferences are supposed to derive from individual preferences, it is, therefore, required to decide an aggregation strategy integrating individual evaluations observed in the test set of each user into the group evaluations. This approach is yet limited since we have no knowledge about what the “true” aggregation strategy is applied by the group, and different strategies undoubtedly produce different results. An example of work following this direction can be found in [29, 41], in which proposed algorithms are compared to a standard set of CF algorithms (such as, popularity based, neighborhood and latent factor models) and the group evaluations are generated by various aggregation strategies, e.g., Least misery, Average, and so on.

- Comparing the generated recommendations for a group with the true individual preferences. This evaluation approach works on the assumption that an individual’s satisfaction only depends on the personal experiences, i.e., what users like or dislike as individuals. Particularly, Baltrunas et al. [8] compared the predicted group rankings with the rankings observed in the test set of each user in terms of nDCG. Likewise, in the work of Salamó et al. [81], users are asked to define their own “Perfect Product” that is supposed to contain all their ideal item’s features, then the system computes the average similarity between a single top recommendation for a group and the “Perfect Products” of each group member.

Both options, however, are based on the underlying assumption that the preferences of individuals are stable and independent of the group decision making process, which is not the case in most scenarios. In fact, the opinions or judgments of users in a group are likely to be influenced by the other group members, which has been observed and categorized as emotional contagion and conformity [61]. As a result, specific simulation approaches come into play, such as those previously employed to simulate the behaviour of users when they interact with a RS [15, 55, 93]. For example, user-system sessions wherein a user incrementally modifies a query to finally select or add a product, were simulated to perform evaluations [55].

In summary, from the literature review, we can conclude that the evolution of group members’ preferences during the decision making process, and their dependencies with
other group factors (e.g., the characteristics of user's behavior in conflict situations, the similarity among users' preferences, or the length of group interaction), have not been examined sufficiently so far, neither in user studies nor in offline experiments. We are convinced that it is essential to develop a simulated environment to test various properties of a group discussion before conducting empirical studies in real-world setting. Therefore, in this thesis we have designed a group discussion simulation procedure that empowers us to experimentally manipulate different group compositions and analyze the responses of group members as a function of their preferences and a range of group factors.
CHAPTER 3

Observing Group Decision Making Processes

In this chapter, we briefly describe an observational procedure, conducted in the context of GRSs, to study a group decision making process wherein group members face a travel decision task. We then summarize our key findings and discuss some important benefits that the study analysis can bring to the design of an effective GRS.

We note that this chapter is based on the collaboration work with Amra Delić, PhD student at Vienna University of Technology. The detailed statistical analyses, therefore, are not provided here as they will be part of the student’s PhD thesis and are published in [33].

3.1 Research Motivation and Hypothesis

The true ultimate goal of RSs is about supporting the decision making process, which becomes even more important in the context of GRSs since it can be used to assist group members in exchanging and processing information to make a well-informed decision [78]. To date, unfortunately, only a small fraction of studies has been devoted to designing full-fledged GRSs with decision and negotiation support functionality, such as Travel Decision Forum [46], Collaborative Advisory Travel System (CATS) [63], or Choicla [87]. In contrast, too much attention has been put on how to identify optimal recommendations, which is nearly impossible to define correctly in any group settings. In particular, as shown by the literature on GRSs, most of the existing group recommendation techniques have been strongly influenced by social choice theory that tries to solve the problem of aggregating users’ individual preferences in a fair and acceptable manner, whereas there are limited studies exploring aspects of group dynamics [60]. In fact, research studies coming from social science have emphasized the importance of the full decision making process adopted
by the group and indicated that the preference aggregation step is only one component of such a process [39].

Driven by these motivations, we argue that observing group decision making process in naturalistic settings can pave the way to understand how this task is executed along with how group members and group situation altogether affect the group decision. The data collected from the observation thereafter can be employed to develop GRSs that effectively support the process of making group decisions. Even though the observational analysis is a prevalent practice in social psychology disciplines [39, 89], it is still uncommon in the field of GRSs. To this end, we introduce the design of an observation procedure wherein groups of users were presented with a particular decision task, that is, selecting a travel destination to visit as a group, and then the groups were observed before, during and after the task. The observations specifically tap into personal characteristics, individuals’ preferences, decision outcomes, and users’ satisfaction with the final group choice.

More in general, the high-level research questions are: (i) whether the preference aggregation strategy applied by the GRS is the key component for predicting a group choice; and (ii) whether other facets of the decision making process influence the choice satisfaction of group members. In this work, these issues are examined in a concrete application domain of travel and tourism since tourism products are typically experienced in groups, and deciding where to travel is rather complex due to the requirement of considering various types of information (e.g., transportation, activities). In the end, the experimental results provide new insights into group decision making and group preference construction. More concretely, we demonstrate that mechanically aggregating users’ initial preferences is far from identifying the actual group choice, and individuals’ characteristics play an important role in grasping the users’ satisfaction with the collective choice. On the base of the findings, we discuss possible implications for building more effective GRSs that can facilitate the process of group discussion and decision making, rather than simply providing a list of group recommendations.

### 3.2 Study Procedure

The research study was held under the auspices of the International Federation for Information Technologies in Travel and Tourism (IFITT) and 11 universities worldwide. The first implementations of the study took place at the Delft University of Technology (TU Delft), the University of Klagenfurt (UNI Klagenfurt) and the University of Leiden (UNI Leiden) whereas an extended study was carried out at the Vienna University of Technology (TU Wien). The number of participants and groups at each university are detailed in Table 3.1.
Each study implementation comprises three phases: *before group discussion* (i.e., pre-questionnaire), *group discussion* and *after group discussion* (i.e., post-questionnaire phase). At each university, participants were arranged into groups of two, three or four members. In addition, at TU Wien each group selected two students, referred to as observers whose task was to observe and record their group activities while all the other group members, referred to as decision makers, took part in the decision making process. Without counting the observers, there were eight groups in total: two groups of two, one group of three and five groups of four decision makers. It is worth noting that the detailed recordings of the decision makers’ behavior were part of the study implementation at TU Wien only.

**Before group discussion.** In this phase, the decision makers filled out an online questionnaire that captured their personal characteristics and individual preferences for a small set of alternative destinations. Specifically, the pre-questionnaire comprises four sections:

1. Demographic data, e.g., age, gender, country of origin.

2. Tourist roles and Big Five factors:

   - Statements related to the 17 tourist roles [68], e.g., *Sun & Chill-out, Social & Sport, Knowledge & Travel, Action & Fun*.

   - Statements related to the Big Five factors [64], i.e., *Openness to new experiences, Conscientiousness, Extraversion, Agreeableness*, and *Neuroticism*.

3. Ratings or ranking of the ten predefined destinations, i.e., Amsterdam (for Austrian participants), Berlin, Copenhagen, Helsinki, Lisbon, London, Madrid, Paris, Rome, Stockholm and Vienna (for Dutch participants); and the experience related to those destinations, i.e., how many times participants have visited each destination.

4. Ranking of decision criteria for choosing a travel destination, e.g., budget, distance, sightseeing, social activities.

A five-point Likert scale was used for the questionnaire statements related to the 17 tourist roles and the Big Five factors.
**Group discussion.** The group meetings took place in a second phase. The task for decision makers was to jointly choose one of the ten previously rated destinations that they, as a group, would like to visit. Moreover, students were asked for their second choice in case the first choice would not be available. At TU Wien, each group had two observers who audio recorded and documented the group decision making process. The report template was constructed on the base of Bales’s IPA consisting of twelve categories of group members’ behavior as explained in Section 2.3. As previously mentioned, only at TU Wien the observers recorded behavior of the decision makers. Precisely, the report template covered the following information:

1. Decision making process planning: whether or not a specific plan for the group decision making process was used, and the duration of the decision process.
2. Group members’ roles, e.g., leader, follower, initiator, or information giver.
3. Group members’ behavior: twelve behavioral categories according to the IPA system.
4. Social decision scheme: when groups engage in a decision making task, usually they adopt a decision scheme to make a final choice, i.e., averaging - the group makes decisions by combining each individual’s preference using some type of computational procedure; voting - the group selects the destination favored by the majority of the members; reaching consensus - the decision is made when everyone agrees on a course of action and expresses satisfaction with the decision; observers could also provide a description of the decision scheme in their own words.
5. Strength of group members’ preferences: the observers rated users’ willingness to give up on their preferred options, on a scale from 1 - Very unwilling to 5 - Very willing.

**After group discussion.** In a third phase, the participants filled out an online questionnaire about the group choices along with their overall experience with the group decision making process. Basically, the post-questionnaire consisted of the following sections:

1. First and second choice of the group.
2. Attractiveness of the ten predefined destinations, e.g., “Many destinations were appealing.”, “I did not like any of the destinations.”.
3. Satisfaction with the group choice, e.g., “I like the destination that we have chosen.”.
4. Difficulty of the decision making process, e.g., “Eventually I was in doubt between some destinations.”.
5. Perceived group identification of group members, e.g., “I identify with the other students in my group.”, and perceived preference similarity with the others, e.g., “I considered myself similar to the other members in my group in terms of our preferences.”.

6. Assessment of the task: participants were asked to select the statements to which they agree regarding the organization of the task, e.g., “The task was well described.”, “I did not understand what we should do.”, and their feedback, e.g., “The exercise was chaotic.”, “I learned something.”.

A five-point Likert scale was used to assess the attractiveness, satisfaction, difficulty and group identification.

It is noteworthy that the implementation of the study at each university followed the described structure, though slight variations existed in the second phase (e.g., the presence of observers). In total the collected data comprised 200 decision makers in 55 groups of two to five group members, plus 16 observers, two for each of the eight groups at TU Wien.

### 3.3 Results

The main objective of this study is to analyze the group decision making process and potential factors that, from the perspective of group dynamics, might influence individual satisfaction with the collective choice. To shed light on this, we examine individual (i.e., initial) preferences of group members, the actual group choices after the discussion, and the individual satisfaction with the final group outcome. In the next paragraphs, we will show that: (i) participants were excited about the group choice, even though their initial top choice was not selected by their group; ii) group choice is not just a plain aggregation of the individuals’ preferences, but that it is rather built up during the decision making process; and iii) users’ satisfaction is related to certain characteristics of individuals.

**Choice satisfaction and individual’s preference match.** In the first data analysis, we studied whether or not the decision makers were satisfied with the outcome of the group decision making process and tried to understand the impact of their initial preferences on that satisfaction.

As shown in Figure 3.1, the individual ratings were relatively high and the distribution was slightly right-skewed. This is rather similar to the rating distributions of other datasets such as MovieLens. The right-skewed distribution may be explained by the fairly high individuals’ satisfactions with the group choice. In particular, as expected, for participants
whose individual top choice matched the group selection (73 out of 200, 36.5%), the majority of them showed a high satisfaction with the destination chosen by the group (67 out of 73, 91.8%), i.e., they indicated that they were excited about the group choice. However, when the group choice was not in line with their preferred option (127 out of 200, 78.0%), most of them were still satisfied with the collective outcome (99 out of 127, 63.5%) (see Table 3.2). Apart from the reason that the decision makers perceived the ten offered destinations overall as very attractive, which was confirmed by the post-survey, a Chi-square test for the contingency in Table 3.2 shows that the two dimensions are, in fact, dependent $p = 0.01$. This means significantly more people are excited about a destination when it matches their individual preferences.

Table 3.2 Contingency table: preferences match and excitement

<table>
<thead>
<tr>
<th></th>
<th>Excited</th>
<th>Not excited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>67</td>
<td>6</td>
</tr>
<tr>
<td>No match</td>
<td>99</td>
<td>28</td>
</tr>
</tbody>
</table>

In general, the observation that participants were satisfied with the final group choice even though it did not match their initially preferred options signals that the group settings affect the individuals’ preferences to a certain extent. Consequently, it is important to figure out what additional factors actually determine this effect.

**Group choice and preference aggregation strategies.** In this second analysis, we explore whether or not common preference aggregation strategies used in GRSs are able to predict the outcome of the group decision making process.
To this end, we compared the actual group choices with recommendations that are generated by various aggregation strategies. Based on the group members’ individual ratings of the ten pre-defined destinations, which were acquired before group discussions, we computed the top two recommendations for each group of users according to the applied aggregation strategies. The ground truth was the actual group decisions on the top two destinations, which were collected after the group discussions. Ultimately, we calculated the precision of the predicted first and second group choice. The average precision computed on 55 groups is shown in Table 3.3.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Precision Top 1</th>
<th>Precision Top 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive</td>
<td>0.329</td>
<td>0.247</td>
</tr>
<tr>
<td>Multiplicative</td>
<td>0.333</td>
<td>0.257</td>
</tr>
<tr>
<td>Median</td>
<td>0.279</td>
<td>0.215</td>
</tr>
<tr>
<td>Least Misery</td>
<td>0.228</td>
<td>0.192</td>
</tr>
<tr>
<td>Most Pleasure</td>
<td>0.163</td>
<td>0.149</td>
</tr>
</tbody>
</table>

The multiplicative strategy, in general, outperformed the other methods, which is in line with previous results [59], but it was able to predict correctly only approximately one third of the group choices, and this is, in absolute terms, a weak performance. The analyzed performance of preference aggregation strategies confirms our assumption that they cannot fully predict the choices made by groups after a group discussion. Hence, an aggregation strategy applied in a mechanical way may not produce a recommendation that the group would choose. The analysis brings us to a conclusion that a useful GRS can be built without adhering to any of these strategies.

Overall, as it was shown, the initial users’ preferences and aggregation strategy are not good indicators of the actual group choice, which motivates us to investigate other factors that can play an important role in determining the outcome of a group decision making process. In particular, we formulate the hypothesis that the satisfaction of users does not only depend on the match between individual and group preferences, but also on a wide range of factors related to the characteristics of the individuals.

**Choice satisfaction and individuals’ characteristics.** In this third part, we study the relationship between the choice satisfaction and individuals’ characteristics. More specifically, we explore to which extent individual level factors, including personality traits, conflict resolution styles, travel behavioral patterns together with interaction behavior categories influ-
ence the group outcome measured by the choice satisfaction and the perceived difficulty of the decision making process. It is important to note that the analysis is entirely exploratory in nature as there were no hypotheses formulated before the statistics were performed.

Considering the Big Five factors of personality, we found out that choice satisfaction is significantly and positively correlated with Agreeableness and Conscientiousness, and negatively correlated with Neuroticism. The obtained correlations are in accordance with the personality theory saying that people with more agreeable and open personalities are easier to be satisfied, compared to those scoring high on Neuroticism.

Moreover, we noticed that the individual satisfaction with the group choice is also related to conflict resolution styles. Specifically, group members with collaborating or accommodating conflict resolution style were significantly more satisfied even when their initial preferences were not aligned with the group decision in the end. Decision makers with avoiding style, by contrast, were highly satisfied with the final group choice when it matched their own personal preferences, yet extremely dissatisfied in case of a mismatch.

Regarding behavioral categories recorded during the decision making process, it was found that choice satisfaction was significantly and negatively correlated with Give opinion and Ask for suggestion behavioral categories, and the perceived difficulty of the decision making task was significantly and positively correlated with Give opinion and Ask for opinion behavioral categories.

Aside from the individual’s satisfaction, we took the average choice satisfaction of the group members to study the satisfaction of a group as a whole. We then assigned the groups to two categories: groups with satisfaction higher than the average and those with satisfaction equal or below the average. The significant results of statistical tests indicated that more satisfied groups scored higher on the Social & Sport and lower on the Sun & Chill-out travel factor. With respect to personality trait, more satisfied groups also scored higher on the Openness to new experiences and lower on Neuroticism. Regarding the interaction behavior of users during the group decision process, we found that in low satisfied groups the observers recorded a significantly higher level of the Disagree behavioral category.

3.4 Discussion and Conclusions

As previously mentioned, the ultimate goal of the proposed observational study is to design more effective GRSs that are expected to help users to choose items that will make the group members more satisfied. The first conclusion that can be drawn from the study is that user’s preferences cannot be viewed as the only independent variables for the group decision making process as it is shown that people tend to accept the group choice even
when it does not fully match their initial preferences. The analysis, additionally, provides evidence that well-known aggregation mechanisms are unable to fully describe and predict the outcome of the group decision process, which may deviate from this normative rule. Finally, we demonstrate the correlation between the choice satisfaction of individuals and their personal characteristics. With these lessons learned from the observational analysis, we hereby discuss some important implications of the analysis of the data acquired by observing users’ interactions during the group decision making process.

**Group situation assessment.** The first implication of the study is that it is pivotal to estimate the “situation” in which each individual member is likely to behave in the group setting. For example, the sequence of the actions performed by the users during the group discussion, e.g., the frequency of the giving opinions or their reaction to the other members’ options, can be exploited to detect whether they will be competitive or collaborative in terms of conflict resolution styles [49, 88], without breaking the flow of interaction between users and distracting them from the main task, i.e., making group decisions. Undoubtedly, it is worth classifying individuals’ behavior into relevant conflict resolution styles as our exploratory analysis has illustrated the relationship between choice satisfaction and these characteristics. Besides, the literature has also showed that incorporating personality or conflict resolution styles as additional information can enhance the group recommendation accuracy [40, 75]. Overall, evaluating the group situation will pave the way for designing a truly adaptive and proactive GRS, which is explained in the next paragraphs.

As a first step, we have developed a mobile GRS that allows group members to be engaged in a group discussion where they can exchange messages together with proposing items that are thought to be suitable for their group and react to other group members’ proposals by giving feedback. The interactions between the members and the system in the course of the group discussion, are monitored and exploited to provide appropriate recommendations and choice suggestions. The details of the system, including application scenario and evaluations of recommendation performance will be described in Chapter 4.

**Adaptation of user’s preference model to the newly expressed preferences.** The next important usage of the observational data is to construct a dynamic recommendation model that integrates preference information derived from the discussion process (i.e., session-based preferences) into the baseline user preference model. It is clear from our study that the final group choice is not completely dictated by the initial long-term preferences of the participants. In the study, the participants were not requested to rate the destinations after their group discussions, which prevents us from telling how their preferences changed before and after the discussions. Arguably, they could like the choice selected by their group but not as much as another one.
users, i.e., the preferences expressed independently of any group influence. We therefore conjecture that the observed group interactions must be taken into account for better predicting which items may suit the group at the precise point in time when the group discussion takes place. Moreover, the findings of our study found correlation between the user’s activity in providing information or criticizing options and choice satisfaction. Hence this type of information must be employed to revise the initial user models learned by the system using the historical preferences of the users. For instance, if a content-based model was fitted to the known ratings of a user, then it can be updated by considering the items that the user liked or criticized during the group discussion. The proper balance between the two kinds of preferences must be identified based on the specific group situation.

This conjecture is addressed and confirmed by a simulation experiment analyzing how long-term and session-based preferences can be appropriately combined in different group scenarios. Basically, it is observed that a combination strategy that weighs more the long-term preferences is fitted to the scenarios when the group setting has no impact on group members’ preferences, but when the group context pushes users to be either cooperative or uncooperative, users seem to benefit more from a system that takes into account the preferences observed from the group discussion, which reflect their newly emerging interests. The full description will be provided in Chapter 5.

**Adaptation of preference aggregation strategy to the estimated group situation.** After revising individual’s preference model, the subsequent goal is to identify which preference aggregation method could be applied. While we already mentioned that there is no single best aggregation strategy that fits all possible recommendation tasks and group settings, we argue that, given a family of candidate aggregation methods, the system can learn to predict the optimal ones for the specific situation of the group, based on the analysis of collected data. The most fundamental issue here, however, is to determine a criterion for evaluating preference aggregation strategies. For example, the system should accurately predict the aggregation strategy that the group in a certain specific setting would use, or encourage the group to deviate from it and try something else that would achieve a better choice in terms of different criteria such as fairness or diversity. Inevitably, the definition of the optimal criteria is multi-dimensional, so alternative solutions can be implemented and compared. Also, in order to understand which role the system should play, live user studies are indispensable.

**Adaptation of recommendation actions to the estimated group situation.** It is clear that an effective GRS should adapt its full recommendation strategy that might consist of different actions, instead of solely recommending items to the group. In that sense, the
system is expected to automatically adapt the presentation and explanation of the recommendations. For instance, based on the estimated group situation, the system can decide which item information is better to be presented to the members, what types of explanation should be provided to persuade the members to choose one of the recommendations, and when the recommendations are offered to the group. More in general, the system should take the role of a facilitator for the group decision making process rather than being a rigid mediator of users’ preferences. In fact, adaptive action selection was successfully introduced and employed in a conversational travel recommender system for individuals [56], and we are convinced that such an approach can bring even a greater benefit in a group decision support system.
In this chapter, we start with the presentation of a chat-based GRS that we developed as working prototype for putting in practice the lessons learned from our prior observational study. We then explain the interactive group recommendation model implemented in the system. Next, we describe an exploratory user study that was conducted to evaluate the system usability, perceived recommendation quality along with acceptance of choice suggestion. Finally, we analyze the users’ interaction data, which was collected in the group decision making process supported by the system.

4.1 Application Scenario

Our developed GRS is an Android-based application called South Tyrol Suggests for Groups (STSGroup1). It is an extension of STS [19], a context-aware point of interests (POI) recommender originally developed for individuals.

Motivation for the user-system interaction design. The system bases its interaction design on research studies of the functional theory of group decision making, which suggest that groups, who go through a four-stage process, Orientation, Discussion, Decision, and Implementation (ODDI model), are likely to make a better choice when facing decision tasks [39]. This finding is bolstered by our observational analysis of group decision making processes highlighting that user’s preferences are constructed during the course of making decisions, and hence it is essential to put emphasis on the decision making process itself instead of only on the preference aggregation step (see Chapter 3). Moreover, the motivation for designing user’s interaction that integrates information processing and communication

1Introduction video: https://www.youtube.com/watch?v=mBrEKhX0vP0
activities has been strengthened by a line of studies in the Computer Supported Cooperative Work field [37]. To test this hypothesis, the proposed GRS, thereby, is equipped with various decision support functions such as group discussion, group recommendations and choice suggestion, which are aimed at facilitating the various stages of the group decision making process.

Among the four stages, group discussion is regarded as the most crucial part of the decision making process, since its importance in understanding a group outcome, especially with respect to the amount of information shared among group members, has been proven [89]. When it comes to supporting real-time group conversations, chat applications such as Google Hangouts Chat\(^2\) and Slack\(^3\), are becoming increasingly popular in many organizations for coordinating teamwork activities [98]. Research studies have also demonstrated the benefits of chat in many different application aspects, such as fostering intimacy among friends and family [44], or enhancing collaboration in the workplace [43]. Arguably, designing the user-system interaction that effectively elicits user’s preferences and supports a natural flow of conversation between users is vital because, as pointed out by Human Computer Interaction research, it has a profound impact on the GRS effectiveness [25, 26].

Grounded on these findings, we decided to leverage a chat environment to expose the recommendation functionality to users within their conversation. Chats promote a back-and-forth discussion among group members. Moreover, the more time the members spend on the discussion, the more observations the system can make on their preferences. Our system, therefore, differs from many of these chat apps, in that it integrates group recommendations into the discussions. Simply put, it enables users to seamlessly add system-generated recommendations to their discussions while they are communicating.

The application domain of this work is recommending tourist attractions or points of interests (POIs) for a group of users. The system is dedicated to a travel and tourism recommendation scenario since most of the time users travel in groups, of various size and nature, such as, a family with kids, or a group of friends and colleagues. Moreover, users tend to contemplate various types of information (e.g., transportation, accommodations, activities) before arriving at a satisfactory option. Hence, the system is expected to provide a seamless recommendation process rather than a one-shot action (as is more popular in other domains like music or books).

**Interaction with the system.** Let us assume users are seeking a POI in South Tyrol (Italy) for their group to visit together. After signing up or logging in, they can specify their

\(^2\)https://gsuite.google.com/products/chat/
\(^3\)https://slack.com/
companions through appropriate system functions including: searching companions by user name (Figure 4.1a), sending connection requests to invite them to join a group discussion, or tagging companions (Figure 4.1b). Once a group of users is connected, the discussion is ready to start. It is noteworthy that the system is targeted at small group sizes, from two to five users, because when the size is small, group members are supposed to actively participate in a group discussion while with larger groups, some of them are likely to turn to be passive [20]. Additionally, users can always access functions that were previously available in STS. For instance, they can indicate important context variables, such as, their travel goal, or browse their personalized recommendations, which are computed by considering only their individual preferences (ratings for previously experienced POIs).

As soon as one group member sends a message to the others, a group discussion session is started (Figure 4.2a). The members can exchange messages in the chat component of the system, which is similar to a normal chat app. In addition, they can autonomously search in STSGroup for interesting items and propose them to their group discussion. All the proposed items are then displayed chronologically in the group discussion space, together with other messages related to the decision making problem. Next, the group members are
free to give feedback on the items proposed by other members in the form of: best choice (crown icon), like (thumbs-up), and dislike (thumbs-down) (Figure 4.2b). The proposed items can also be tagged by comments or emoji reactions.

We note that when a user proposes an item to the group discussion, the system automatically considers it to be the best choice of the user for the group unless (s)he explicitly gives a different evaluation. As part of the interaction, any member is also free to decide not to evaluate any of the discussed items. A summary comparison panel, which aggregates and compares the members’ dislike, like and best choice evaluations, is always shown on top of the screen in order to keep the group members aware of the others’ preferences. The panel is updated automatically when a new preference is expressed or when changes in the preferences are made by any group member. In general, the group members are able to perform as many interaction cycles as they desire, where a cycle is defined as a POI proposal made by one group member, followed by possible feedback given by the others.

During the course of the group discussion, the system can generate, upon the request of a user, novel group recommendations, in addition to POIs already proposed. This function offers the user a possibility to explore alternative options as new proposals (Figure 4.3a). The
recommended items for the group are augmented with explanations that provide a rationale for the system recommendations. The system, can, for example, refer to the item features that might draw the attention of the group members. By monitoring users’ evaluations given in a chat session, the system continuously updates all the group members’ preference models that are implemented with utility functions, and that are initialized before the discussion starts by considering the users’ group-independent ratings for POIs. The number of group members’ actions is also taken into account. Specifically, the more items a user rates, the higher the importance weight the user will have in the preference aggregation step of the recommendation computation. The details of the recommendation model that is used to suggest new candidate items from the entire item set are presented in the next section.

When facing difficulties in arriving at a final decision, any user can activate the choice suggestion function (Figure 4.3b). The system then ranks the discussed items to provide a choice suggestion. In principle, the items proposed in the group discussion can be ranked by the recommendation model, and yet, as in the primitive version, the system computes an accumulated score for each item, which is based on the item evaluations given by all the group members. Each item receives plus 2 and plus 1 for best choice and like feedback.
respectively, and minus 1 for a dislike evaluation. The ranking list and explanations are constructed with respect to this score. After several interaction cycles, wherein POIs are proposed, discussed and rated by individuals, the discussion might end up with a POI to actually visit or without one.

Interestingly, recently launched Google Maps’ group planning feature\(^4\) is also following a similar interaction design. Basically, that feature allows users to create a shortlist of restaurants or destinations and share it with friends they invite. Group members then will be able to see each destination's location on the map, and they can vote with a thumbs-up or thumbs-down, or add their own suggestions.

### 4.2 Recommendation Logic

Before any group discussion unfolds, users’ long-term preferences are acquired by our system in the form of item ratings. During a group discussion, it is then assumed that the group members may deviate somewhat from their previously observed preferences because of the influences of the other members and the group decision making process [61]. The system, therefore, exploits both sources of preferences to generate and continuously update the user’s preference model, which is represented with a utility function. Finally, the utility functions of the group members are aggregated to build a group preference model, i.e., group utility function. The recommended items are ranked according to this aggregated group model.

**Individual Long-term Preferences**

Each user’s preference model is represented by a utility function:

\[
U(u, i) = \sum_{j=1}^{N} w_{u_j} x_{ij}. \tag{4.1}
\]

The utility of an item gives a quantitative indication of the user’s preferences for the item, and we assume that a user prefers items with larger utility to those with smaller utility.

In the utility function (4.1), \(x_i = (x_{i1}, \ldots, x_{iN})\) is an \(N\)-dimensional Boolean feature vector that represents item \(i\), where \(N\) is the number of item features. If \(x_{ij} = 1\) then item \(i\) has the \(j\)-th feature, \(x_{ij} = 0\) otherwise. For instance, \(x_5 = (1, 0, 1, 0)\) means that item 5 has the first and the third features but not the second and the fourth ones. Long-term preferences of each user \(u\) are modeled by a vector of weights \(w_u\), also called a utility vector, representing the estimated importance that user \(u\) assigns to the \(N\) item features, where the sum of the weights

\(^4\)https://www.blog.google/products/maps/all-together-now-group-planning-google-maps/
is 1, i.e., $\sum_{j=1}^{N} w_{uj} = 1$ and $w_{uj} \geq 0$. Based on the user’s ratings acquired independently of any group setting, the user’s utility vector is built by applying a content-based approach:

$$w_{uj} = \frac{\sum_{i \in I_u} x_{ij} r_{ui}}{\left|\{i \in I_u : x_{ij} \neq 0\}\right|}, f = 1, \ldots, N,$$  \hfill (4.2)

where $r_{ui}$ is the rating that the user $u$ gave for item $i$ and $I_u$ is the set of items rated by user $u$. By construction, if $w_{uj} > w_{ul}$ then the $j$-th feature is more important than the $l$-th feature according to user $u$. The intuition of equation (4.2) is to take into account the features associated with an item that the user rated together with their frequency. For example, we assume that the ratings given by user $u$ are ranging from 1 to 5 and are known for the following three items: $r_{u,12} = 3$, $r_{u,10} = 1$, and $r_{u,22} = 5$. Let us further suppose that the feature vectors of the considered items are: $x_{12} = (1,1,0)$, $x_{10} = (0,1,0)$, and $x_{22} = (1,0,1)$, respectively. Based on equation (4.2), $w_u$ is computed as follows: $w_{u1} = \frac{3+5}{2} = 4$, $w_{u2} = \frac{3+1}{2} = 2$, $w_{u3} = \frac{5}{1} = 5$. Ultimately, $w_u$ is normalized by dividing it by $\sum_{j=1}^{\infty} w_{uj}$, a normalization factor that is used to satisfy the structural constraints $w_{uj} \geq 0$ and $\sum_{j=1}^{\infty} w_{uj} = 1$, so that it results in $w_u = (0.36, 0.18, 0.46)$.

### Session-based Preferences

During the group discussion, the system derives the session-based preferences of each group member by observing their expressed evaluations of the proposed items. These evaluations have the form of: best choice, like, dislike, or not evaluated\(^5\) (neither best choice, nor like, nor dislike). Thus, for each group member, all items proposed in the group discussion are classified into four sets: BS($u$) (best choice), LS($u$) (like), NS($u$) (not evaluated) and DS($u$) (dislike). As we assume that the user prefers items with larger utility, the following constraints must hold:

$$U(u,o) < U(u,p) < U(u,q) < U(u,s),$$  \hfill (4.3)

for all $o \in DS(u)$, $p \in NS(u)$, $q \in LS(u)$, and $s \in BS(u)$.

It is worth noting that for each individual, the system can collect multiple constraints from their evaluations of items, so we denote with $\phi_u^g$ the set of constraints on the utility function of user $u$ inferred from his or her evaluations given in group discussion $g$ [91]. For instance, let us assume $DS(u) = \{x_9\}$, $LS(u) = \{x_{22}\}$, and $BS(u) = \{x_{11}\}$. We also suppose that $x_9 = (1,0,0,1,0)$, $x_{22} = (0,1,1,0,1)$, and $x_{11} = (0,1,1,1,0)$. According to

\(^5\)We make the simplifying assumption that not evaluated items have an implicit evaluation between disliked and liked.
Building a Chat-based Group Recommender System

(4.3), \( U(u, 11) > U(u, 22) \), so \( \sum_{j=1}^{N} w_{u_j} x_{11j} > \sum_{j=1}^{N} w_{u_j} x_{22j} \). Similar constraints can be deduced from comparing item 22 and item 9. In the end, we have \( \phi_u^g = \{ w_{u2} + w_{u3} + w_{u5} > w_{u1} + w_{u4} ; w_{u4} > w_{u5} \} \).

The objective of the proposed model is to help the group members to find items that match not only their preferences revealed in group discussion \( g \), but also those of the other group members. The underlying idea is that, in addition to the individual specific preferences, in absence of any other preference information the group members are assumed to agree on the importance of several features of the items. For example, the presence of a swimming pool in a hotel is generally seen as valuable. Hence, this is done, for each user \( u \), by searching for a utility vector \( w_u^g \) that not only satisfies the inferred constraints in \( \phi_u^g \), but also maximizes its cosine similarity with the vector \( w_g \), the aggregated utility vector of the group members (i.e., the larger the value, the closer the two vectors). The resulting optimization problem is formulated as follows:

\[
w_u^g = \arg \max_w \cos(w, w_g) \quad \text{subject to } w \text{ satisfies } \phi_u^g.
\] (4.4)

User’s Utility Update and Group Utility Function

After inferring session-based preferences, the system estimates an updated utility function for each group member, i.e., finding a new vector of weights \( w_u \) that takes into account the session-based preferences, which may be influenced by the group setting, and the individual long-term preferences. More precisely, a linear combination of the user’s long-term and session-based utility vector, \( w_u \) and \( w_u^g \) respectively, is assumed to be the updated utility vector of the user:

\[
w_u = \sigma w_u + (1 - \sigma) w_u^g, \sigma \in [0, 1],
\] (4.5)

where \( \sigma \), so-called preference “stability” parameter, is used to weigh the contribution of the long-term and session-based preferences: a larger stability \( \sigma \) implies that the long-term preferences are considered to be more important.

These updated utility vectors are then aggregated to generate a utility vector of the group denoted \( w_g \). It is worth noting that at the beginning of the learning process, the group utility vector \( w_g \) is initialized by using the Average aggregation function since the system has no additional information about the members’ roles in the group. However, as soon as they are in the group discussion and interact with the system, \( w_g \) is updated with Weighted Average aggregation rule:

\[
w_g = \sum_{u \in g} \alpha(g, u) w_u,
\] (4.6)
where $\alpha(g,u)$ is a non-negative coefficient associated with user $u$ in group discussion $g$ and $\sum_{u \in g} \alpha(g,u) = 1$. Motivated by studies in social psychology, which support the importance of the participation rate in the group discussion, i.e., so-called “babble effect” [39], the $\alpha(g,u)$ coefficient of each user is the proportion of the number of user’s actions (POI proposals, POI evaluations and POI comments) over the total number of actions acquired from all group members:

$$
\alpha(g,u) = \frac{\text{# actions performed by } u \text{ in } g}{\text{# actions performed by all group members in } g}
$$

(4.7)

here, the more feedback the user provides, the higher the value of his or her coefficient is.

Finally, when a user requests group recommendations, the full set of items is ranked according to the group utility function, defined as follows:

$$
U(g,i) = \sum_{j=1}^{N} w_{gj} x_{ij}.
$$

(4.8)

The system then suggests the POIs with the highest utility for the group, and therefore the generated recommendations are the same for all group members. The full iterative cycle of the group recommendation model is illustrated in Algorithm 1.

Algorithm 1: User’s utility vector update and recommendation algorithm

1. Initialization
   - For all $u \in g$, compute $w_u$ by equation (4.1)
   - $w_g = \frac{1}{|g|} \sum_{u \in g} w_u$
2. $DS(u) = NS(u) = LS(u) = BS(u) = \emptyset$
3. During the group discussion
   - While there is a new evaluation in the group discussion do
     - Add the evaluated item to the suitable set among $DS(u), NS(u), LS(u),$ and $BS(u)$
     - Infer user’s $u$ constraints $\phi^g_u$ on the utility function of $w^g_u$ as in (4.3)
     - Compute user’s session-based preferences, $w^g_u$, by solving the optimization problem as in (4.4)
     - Update user’s utility vector $w_u$ with the linear combination as in (4.5)
     - Update the group utility vector $w_g$ as in (4.6)
4. End
4.3 Evaluation Procedure

We conducted a user study to evaluate the usability of our system, the perceived quality of the group recommendations and the group choice satisfaction, i.e., the satisfaction of the group members for the POI selected by the group with the system support. Besides, the users’ identification with the group and their acquaintance with the area were also checked during the evaluation. In particular, the user study was carried out in two separate periods. Each implementation followed a three-phase structure: before, during and after a group discussion.

**Before group discussion.** Before starting a group discussion, each participant was requested to rate at least five POIs in order to acquire their individual user preferences. It is important to note that none of them knew the functionality of STSGroup in advance as it was only introduced to the participants when the group meeting took place. Prior to the group discussion, the participants were free to pick up their companions and form their groups. One member in each group was assigned to be the “initiator”, a person who starts the discussion by proposing the first POI to the group.

**During group discussion.** The participants were invited to meet physically, and at the beginning of the group meeting, each user received a mobile device in which STSGroup was installed. The experiment was performed using LG Google Nexus 5 smart-phones running Android 6.0.1. The participants were asked to exchange their STSGroup user name, then send or accept friend requests. Afterwards, the participants were presented with the following task scenario:

> “Imagine that you and your group members have a plan to visit a place in South Tyrol together. According to your own preferences, the system offers you a suggestion list. Your task is first to select one or more places in the list that you think are suitable for your group to experience together and propose them to your group. Afterwards, you and the group members could discuss the proposed options and decide which place your group will choose to visit”.

We explained that each member is able to scroll through and view personal suggestions within STSGroup. Aside from that, it also provides them with the possibility to browse group recommendations computed by the method described in Section 4.2. All these functions could be flexibly used to come up with a group choice. They, and similarly their friends, could select places in the suggestion lists and propose them to their group. Additionally, they could discuss the proposed options in the supported group chat, and eventually
chose one of them to visit. We also requested that the communication with each other was mainly based on the system chat functionality, which was ensured by the presence of an experimenter in the room.

**After group discussion.** The participants filled out a questionnaire composed by four parts: perceived system usability, perceived recommendation quality, choice satisfaction and group identification. For each questionnaire item, users replied on a five-point Likert scale ranging from “strongly disagree” to “strongly agree”.

We used the System Usability Scale (SUS) [9] to assess our system usability, which is one of the most popular post-study standardized questionnaires and allows to measure the usability even with a small sample population [84]. The 10 SUS statements are:

- **S1:** I think that I would like to use this system frequently.
- **S2:** I found the system unnecessarily complex.
- **S3:** I thought the system was easy to use.
- **S4:** I think I would need the support of a technical person to be able to use this system.
- **S5:** I found the various functions in the system were well integrated.
- **S6:** I thought there was too much inconsistency in this system.
- **S7:** I would imagine that most people would learn to use this system very quickly.
- **S8:** I found the system very cumbersome to use.
- **S9:** I felt very confident using the system.
- **S10:** I needed to learn a lot of things before I could get going with this system.

Each SUS item’s score contribution ranges from 0 to 4. For positively phrased statements (odd numbers) the score contribution is the scale position minus 1. For negatively worded statements (even numbers), the contribution is 5 minus the scale position. To get the overall SUS score, the sum of the items’ scores is multiplied by 2.5, so the overall system usability score ranges from 0 to 100. Several benchmarks for SUS across different systems have been published [9], and an average SUS score computed in a benchmark for cell phone applications is around 67. In our user study, we used this value as a baseline to determine whether our application usability exceeds the benchmark.

The next section of the questionnaire is composed of 5 statements about perceived recommendation quality, 3 statements about choice satisfaction, and 3 statements about how
the user identifies themselves with the group. The measurements of recommendation quality and choice satisfaction were adopted from the work of Knijnenburg et al. [51].

- **RecQual1**: I liked the final choice suggested by the system.
- **RecQual2**: The final choice recommended by the system was well-chosen.
- **RecQual3**: I didn’t like the suggested final choice.
- **RecQual4**: The new item recommendations for a group, excluding the proposed items were relevant.
- **RecQual5**: I didn’t like any of the recommended new items.
- **ChoiceSat1**: I was excited about the place that we have chosen.
- **ChoiceSat2**: The chosen place fits my preference.
- **ChoiceSat3**: I didn’t prefer the chosen place, but it was fair.
- **Identification1**: I’m glad to be a member of the group.
- **Identification2**: I feel strong ties with my group.
- **Identification3**: I considered myself similar to the other members in my group in terms of preferences.

Moreover, in the post-study questionnaire, the participants could voluntarily respond to a multiple choice question asking about their familiarity with South Tyrol, e.g., “I’m a citizen of the region”, “I know this region quite well”, or “I don’t know much about the region”.

**Data collection.** In the end, the first implementation of the study involved 15 participants (students and colleagues) formed in 6 groups of two (3) and three (3) users. The second extended study consisted of 27 participants organized in 10 groups of two (4), three (5) and four (1) users. Both study implementations followed the same described structure, but the only distinction here was that in addition to asking a set of subjects to meet and physically use STSGroup, the second data collection process was also conducted through the on-line mobile emulators Appetize.io⁶ (Figure 4.4a) and BrowserStack⁷ (Figure 4.4b) that allowed 4 groups of two (3) and three (1) users to interact remotely with our web-based application.

⁶https://appetize.io
⁷https://www.browserstack.com/app-live/
4.4 Results

(a) Appetize.io

(b) BrowserStack

Fig. 4.4 Screenshots from the on-line mobile emulators that were used to collect data

In total 42 participants to the experiment were engaged in 16 groups of two (7), three (8) and four (1) members. More than 1000 POIs were considered, which belong to a wide range of sightseeing destination categories, such as, natural monuments, historical buildings, castles, and so on. In STSGroup, POIs are represented with Boolean vectors denoting the presence/absence of a set of keywords that are extracted from item categories and descriptions coming from a web-service provided by the Regional Association of South Tyrol’s Tourism Organizations\(^8\). To this end, each POI is modeled by 84 item features. The system collected 328 user’s ratings for 120 POIs before any actual group discussion took place.

4.4 Results

In this section, we present the results of the system usability analysis and the outcome of group decision making process supported by the system in terms of perceived quality of the group recommendations, perceived satisfaction with the group choice as well as user identification with the group.

With regards to the system usability, Figure 4.5 shows the SUS score of each participant. Most of these scores are higher than the benchmark of 67. Overall, STSGroup obtained the mean SUS score of 74.88 with the standard deviation of 4.45 (over the 42 users considered in

\(^8\)LTS: http://www.lts.it
We conducted a one-sided test, explicitly checking to see if this mean score is significantly higher than the benchmark of 67: \( t = 11.48 \) and the one-tailed probability associated with this score is \( p\)-value \( = 1.1e - 14 < 0.01 \), hence the confidence that the mean SUS score of STSGroup exceeds the benchmark is higher than 99%. Although the observed groups are composed of students and colleagues, due to the small sample size, we could not check any significant difference in the perceived usability of STSGroup between these two types of groups.

We further computed the average responses for the individual 10 SUS statements. The highest average scores are for S6, S4 and S8. This implies that the participants evaluated STSGroup as neither complex nor difficult to use. They also did not think that the system is inconsistent or cumbersome, and they believed that they are able to use the system without technical help. We further observed that S9, S7 and S5 received instead the lowest scores. This implies that the users were not fully confident of using the system and thought that most people would not learn to use it quickly. They also found some of the functions in the system not well integrated. All these issues could be explained by the fact that in STSGroup, we support two types of recommendations simultaneously: personal context-aware recommendations that are tailored to the individual user’s preferences, and group recommendations that are adapted to the aggregated group utility of the group members. In other words, users could use all the functions that were already available in STS, apart from the new ones designed for groups. Due to the integration of individual and group recommenda-
Table 4.1 Group recommendation quality

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree (%)</th>
<th>Agree (%)</th>
<th>Neither disagree nor agree (%)</th>
<th>Disagree (%)</th>
<th>Strongly disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RecQual1(^a)</td>
<td>16.8</td>
<td>64.3</td>
<td>18.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>RecQual2(^b)</td>
<td>31.0</td>
<td>54.7</td>
<td>14.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>RecQual3(^c)</td>
<td>0.0</td>
<td>0.0</td>
<td>11.9</td>
<td>45.8</td>
<td>42.3</td>
</tr>
<tr>
<td>RecQual4(^d)</td>
<td>8.7</td>
<td>74.6</td>
<td>12.0</td>
<td>4.7</td>
<td>0.0</td>
</tr>
<tr>
<td>RecQual5(^e)</td>
<td>0.0</td>
<td>0.0</td>
<td>7.2</td>
<td>61.5</td>
<td>31.3</td>
</tr>
</tbody>
</table>

\(^a\) I liked the final choice suggested by the system  
\(^b\) The final choice recommended by the system was well-chosen  
\(^c\) I didn’t like the suggested final choice  
\(^d\) The new item recommendations for a group, excluding the proposed items were relevant  
\(^e\) I didn’t like any of the recommended new items  

Regarding choice satisfaction and group identification shown in Table 4.2, we observe that 61.8% of the participants confirmed that “the chosen place fits my preference” (Choic-
Fig. 4.6 The distribution of individuals’ ratings collected before the group discussions

eSat2) and more than half of users did not really prefer the chosen place (ChoiceSat3). The satisfaction with the final choice, however, was still relatively high. Particularly, 85.7% participants indicated that they were excited about the place chosen by their group (ChoiceSat1). This somehow signals that some users were satisfied with the collective choice even though it was not really in accordance with their preferences.

Exploring the results of how users identified themselves with other members in their group, we found that the high choice satisfaction might be related to the strong relationship between the members. Specifically, more than 80% of the participants were happy to be in their group (Identification1), and more than two-third of the users (69%) confirmed that they felt a strong bond with their group (Identification2). The analysis of the responses to the last question (Identification3) shows that only 30.9% of participants considered themselves similar to the other group members in terms of preferences. This somehow implies that although a minority of the group members considered them similar to the others, the majority of participants still were happy with the outcome of the group choice (ChoiceSat1).

4.5 Data Observations and Analysis

In this section, we analyze the data collected in the group decision making processes that have been supported by the system. The analysis of the observed data will be used to motivate the design of a group discussion simulation model introduced later on, in Chapter 6.

As previously described in Section 4.1, with the support from STSGroup, each user, while discussing the appropriateness of the options with the other group members, can exchange messages, propose items to the group, and evaluate other members’ proposals. We collected the logs of their interactions during the group discussions in the following format:
Table 4.2 Group choice satisfaction and group identification

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree (%)</th>
<th>Agree (%)</th>
<th>Neither disagree nor agree (%)</th>
<th>Disagree (%)</th>
<th>Strongly disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChoiceSat1(^a)</td>
<td>9.5</td>
<td>76.2</td>
<td>11.9</td>
<td>2.4</td>
<td>0.0</td>
</tr>
<tr>
<td>ChoiceSat2(^b)</td>
<td>7.1</td>
<td>54.7</td>
<td>20.0</td>
<td>20.0</td>
<td>0.0</td>
</tr>
<tr>
<td>ChoiceSat3(^c)</td>
<td>2.4</td>
<td>9.5</td>
<td>45.2</td>
<td>38.1</td>
<td>4.8</td>
</tr>
<tr>
<td>Identification1(^d)</td>
<td>26.2</td>
<td>61.9</td>
<td>11.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Identification2(^e)</td>
<td>9.5</td>
<td>59.5</td>
<td>23.8</td>
<td>2.4</td>
<td>4.8</td>
</tr>
<tr>
<td>Identification3(^f)</td>
<td>7.1</td>
<td>23.8</td>
<td>26.2</td>
<td>38.1</td>
<td>4.8</td>
</tr>
</tbody>
</table>

\(^a\) *I was excited about the place that we have chosen*

\(^b\) *The chosen place fits my preference*

\(^c\) *I didn’t prefer the chosen place, but it was fair*

\(^d\) *I’m glad to be a member of the group*

\(^e\) *I feel strong ties with my group*

\(^f\) *I considered myself similar to the other members in my group in terms of preferences*

Moreover, for each group discussion we recorded: the discussion duration (minutes) and the number of item evaluations given by the group members. Within the 16 group discussions mentioned above, 108 evaluation samples were collected, i.e., the feedback was revealed in the form of either best choice, like or dislike. On average there were 2.3 interaction cycles for each group discussion. We recall that a cycle is marked by the beginning of a POI proposal made by one group member, followed by possible feedback given by the others. Also, the observation of 16 group discussions showed that the longest group discussion was held for about 20 minutes while the quickest one was approximately 5 minutes long.

When it comes to proposing, we observed that the participants took turns to propose their preferred items, resulting in the number of items proposed by each user being nearly equal, i.e., they made approximately one proposal during the group discussion. When it comes to assessing the items proposed by the other members, we noticed that not all of them were evaluated. In fact, the participants gave feedback only on 67% of the cases. To grasp the distribution of the users’ utilities, we first, for each user, calculated their utilities for the items in the catalog and took the 33rd and 66th percentile utility values, so that the number of items was equally divided into three ranges of utility values. Then we computed the mean of the 33rd and 66th percentile utility values, which is respectively rounded up to 0.18 and 0.37, as visualized in Figure 4.7. Based on that analysis, we defined $Low \ (0, 0.18]$, $Medium \ (0.18, 0.37]$ and $High \ (0.37, 1]$ as the three ranges of the mean utility values.

Interestingly, we observed that most of the time the participants proposed items with a large utility (the ones they thought would be best for the group as a whole), which matches
the utility maximization rule of utility-based choice theory [13]. This observation is important as well as supporting the definition and the usage of the user’s utility function that we have defined.

On the contrary, but not surprisingly, given the results of the observational study presented in Chapter 3 that group preferences are constructed during the group decision making process, we discovered that the participants did not always favor the items with the largest utility when giving evaluations. Particularly, we estimated the probability that an item with a given user’s utility value belonging to a particular range would receive a certain type of feedback (i.e., either best choice, like, or dislike). Based on collected data, for each utility range, this probability is calculated as the proportion of the items that received that feedback among all types of feedback observed in that range. Table 4.3 shows the partial randomness of individuals’ feedback. For instance, let us consider the utility value of user $u$ for item 11, $U(u, 11) = 0.25$. This is regarded as an item of medium utility for the user. In this case the estimated probabilities that the item will receive the best choice, like, and dislike feedback from the user $u$ are 0.28, 0.55 and 0.17, respectively. These observations are supported by the multinomial logit model, a probabilistic choice model, stating that the probability of choosing an alternative increases monotonically with an increase in the utility of that item and decreases with increases in the utility of each of the other items [70].

Further analysis of the logged information revealed that most of the group discussions (i.e., 11 out of 16) ended with at least one item that received most positive evaluations (e.g., best choice or like) from the group members, even though we did not explicitly ask

**Fig. 4.7** The distribution of 33rd and 66th percentile user’s utility values
Table 4.3 Probability of each type of feedback in low, medium, and high utility ranges

<table>
<thead>
<tr>
<th>Feedback</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Best choice</td>
<td>0.16</td>
</tr>
<tr>
<td>Liking</td>
<td>0.31</td>
</tr>
<tr>
<td>Disliking</td>
<td>0.53</td>
</tr>
</tbody>
</table>

the participants about their final group choice. We also observed that there were some discussions that ended up without a clear choice being reached.

4.6 Discussion and Conclusions

In this chapter, we argued that to assist users in making better group decisions, a GRS should support the whole decision process. We thereby have introduced the interaction design of the mobile GRS that is aimed to support the group decision making process by offering a group chat environment wherein a number of recommendation functions are integrated. We also presented the implemented group recommendation model that exploits both individuals’ long-term preferences and the users’ feedback revealed during the group discussion in order to update the system definition of the users’ utility functions.

We have carried out an exploratory user study where the usability of the system, the quality of group recommendations and the choice satisfaction have been measured. From the study, we found evidence that the usability of the proposed system is higher than a standard benchmark, and it also obtains encouraging results for perceived group recommendation quality and choice satisfaction. The system, however, undoubtedly has some limitations which ultimately are linked to the difficulty for the users to understand the exact meaning of certain recommendation functions, such as the difference between individual and group recommendations.

Also in this work, we have analyzed the interaction data collected by our system in the course of group discussions. During the data analysis, we have noted several challenges pertaining to the performed user study:

1. Users were not shared among groups since one user joined in only one group discussion. For that reason, we cannot make any conclusion about the behavior of the same user in different group settings, and accordingly, about the influence of the group on the preferences expressed by users in the group discussion.
2. The low number of participants is an obvious limitation of our experiment, which leads to insufficient data to conduct an exploratory correlation analysis on factors that might influence the group choice such as group sizes, composition of the groups, participants’ knowledge about the discussed POIs, and the relationship among the group members.

3. Due to the limited time allocated to group discussion, each group discussion was at most 30 minutes long, hence asynchronous communication has not been observed, and the amount of interactions is limited to some extent.

Finally, considering the proposed group recommendation model, we acknowledge that it is essential to study how to appropriately combine long-term and session-based preferences. Ideally, the GRS must efficiently make use of all the available user information and quickly adapt to users’ current needs and requirements. We therefore hypothesize that the optimal combination of the two types of preferences is likely to differ depending on specific group settings. This issue will be investigated in a follow-up study presented in the next chapter.
Evaluating an Interactive Group Recommendation Model

In this chapter, we explore the appropriate fusion of long-term and session-based preferences in different group scenarios. We first describe the simulation procedure that is designed to simulate users' behavior in these scenarios. We then, for each considered scenario, analyze the results obtained by comparing the performance of group recommendations generated by alternative combinations of these two types of user's preferences.

5.1 Research Motivation

In the previous chapter, we proposed an interactive group recommendation model that not only relies on individual long-term preferences, which are typically acquired in the form of item ratings, but also exploits users' feedback on items, which is revealed over the course of a group discussion. This immediate feedback, referred to as session-based preferences, reflects the current user's needs, which might deviate from the individual preferences due to the influence of the group setting [39, 61]. The proposed model was implemented in a mobile GRS, STSGroup, that offers a chat environment in which a variety of decision support and recommendation functions are integrated. From the preliminary results of a user study, we provided evidence that the implemented group recommendation approach is able to enhance the perceived group recommendation quality.

While the conducted user study was useful and important to understand whether the interactive recommendation model is effective and well-accepted by users, it was inadequate to fully assess the model performance, which necessitates that the system be examined in different group situations. As pointed out by the studies on group decision making dynamics [39], group members can have several kinds of social responses to group pressure, leading to numerous potential group scenarios. For example, they might express ideas and judgments that are consistent with their personal stance, or they can possibly change their
opinions to accept the influence of others, or they can probably take actions that are opposite to whatever the group suggests. Given a dynamic group situation, the major challenge with the group recommendation model is how to properly exploit a user’s preferences induced by the group session. We argue that the relative importance of the individual long-term and session-based preferences should vary according to certain specific group settings.

Arguably, it is difficult to tell when the system should mainly count on individual preferences and when it should mostly adapt to the ones derived from the group interactions. To this end, among the various dimensions of possible analysis, we break down the appropriate combination of long-term and session-based preferences into three concrete group scenarios, corresponding to three typical kinds of group impacts on a user’s behavior: a) independence - the group has no influence on the user preferences; (b) conversion - the group setting nudges group members to become more similar to each other; and (c) anti-conformity - the group setting pulls the members in divergent directions. Specifically, we take advantage of a simulation approach to simulate users’ interactions in the three considered situations wherein we implement three different combination strategies: (i) when the importance of the long-term and session-based preferences is equal; (ii) when higher importance is given to long-term preferences; and (iii) vice versa, when greater importance is given to session-based preferences. For each scenario, we then evaluate the performance of the three preference combination rules in terms of ranking quality, i.e., we observe the ranking of the recommendations as group size and the amount of feedback elicited during a simulated group discussion grow.

The experimental results show fundamental properties of long-term and session-based preferences fusion in the proposed interactive group recommendation model. In the scenario (a), the model requires less preference information gleaned from the group discussion whereas in the scenarios (b) and (c) it must weigh more the session-based preferences to faster identify the best recommendations for the group. We also demonstrate that the proposed model is able to better capture the changes in user preferences as more feedback is provided. We believe that this research study can lay the groundwork for understanding when it is necessary and how to appropriately leverage each type of preferences, so that a GRS is able to improve its recommendation quality as well as quickly adapt to the users’ needs and requirements.

5.2 Group Discussion Simulation Model

In order to investigate the effect of users’ interactions driven by different group settings on the performance of three preference combination strategies, we have designed a group
discussion simulation procedure wherein we model: (i) groups of users; and (ii) user’s behavior corresponding to each considered scenario, namely independence, conversion, and anti-conformity.

**User Groups**

Intuitively, it is more challenging to generate recommendations for heterogeneous groups whose members have diverse tastes, so we inspect this type of groups within the scope of this work. Specifically, we randomly generate groups of varying sizes by adopting the method introduced by Baltrunas et al. [8], which samples users from standard data sets. We here employ the publicly available STS dataset\(^1\) that contains 2534 ratings for 249 points of interest (POIs) by 325 users [18]. These ratings were entered by users of the STS app\(^2\), a context-aware mobile application providing POI recommendations [19]. In our experiments, we simulate small groups of varying sizes from 2 to 5 users, since this is the target scenario for STSGroup (see Section 4.1). Regarding the POIs (items), each one is represented by 84 features that are extracted from various information sources, like categories (e.g., natural monuments, museums) and a short description of the item. This information is obtained through a web-service provided by the Regional Association of South Tyrol’s Tourism Organizations\(^3\).

Let us recall that before entering into any group, the utility vector \(w_u\) of each group member is initialized based on their group-independent ratings, as in Equation (4.2), which represents the individuals’ long-term preferences (see Section 4.2).

**User’s Behaviors**

To simulate how group members behave in the specific group settings, we create, for each user, a true utility vector denoted with \(v_u\), which represents their newly arising preferences when they are in a group. We assume that the true utility vector is defined as a composite of the individual and group-induced preferences, which is motivated by the research on social norm stating that when in a group, people try to modify their judgments so that they match those of others in their groups, but at the same time, they also attempt to change the group to suit their personal inclination [85]. Therefore, the vector \(v_u\) is defined as follows:

\[ v_u = w_u + \gamma w_g, \]  

\(^{1}\)https://tinyurl.com/sts-dataset  
\(^{2}\)https://play.google.com/store/apps/details?id=it.unibz.sts.android  
\(^{3}\)http://www.lts.it
where \( w_g \) is the centroid of the pre-discussion utility vectors, computed by averaging the individual utility vectors \( w_u, \forall u \in g \), since the members’ roles in the group are assumed to be equal. In Equation (5.1), we can see that user’s interactions in the group setting originate from their own preferences captured by \( w_u \) and from those of other group members characterized by \( w_g \). The parameter \( \gamma \) is used to control the degree of influence from the group setting on users’ behaviors, which can be either positive or zero or negative. The true utility vector \( v_u \) is finally normalized to sum to 1. In particular, the value of \( \gamma \) is set differently according to the certain group scenarios:

- **Independence**: user’s preferences are not influenced by the group setting, and hence \( \gamma = 0 \Rightarrow v_u = w_u \).

- **Conversion**: the group setting pushes users to be in agreement, hence the true user’s utility \( v_u \) veers closer to the centroid vector \( w_g \), i.e., \( \gamma > 0 \).

- **Anti-conformity**: the group setting causes users to diverge from the average opinion during the group discussion. Hence, the true utility vector \( v_u \) is now obtained by moving \( w_u \) in the opposite direction to the centroid vector \( w_g \), which means \( \gamma < 0 \).

The flow of the simulation process mimics the group interaction between users and our chat-based system STSGroup, in which each user through a discussion with other group members can propose items to the group, and evaluate other members’ proposals in the form of: best choice, like, dislike or not evaluated at all. The simulated user’s behavior, thereby, all comes down to two actions: which items are proposed to the group discussion, and how the user evaluates the items proposed by the other members. Concretely, we base the individuals’ actions on the defined true utility vector.

**Propose items to the group.** The items obtaining the highest true user’s utility \( v_u \) are assumed to be offered to every new discussion cycle, aka group iteration. We note that if the item is already proposed to the group then the one with the next highest utility is selected.

**Evaluate the proposed items.** Regarding user’s evaluations of the items proposed by the other members, we rank all the proposed items in order of decreasing true user utility function, i.e., \( U^v(u,i) = \sum_{j=1}^{N} v_{uj}x_{ij} \), where \( N \) is the number of item features and \( x_i = (x_{i1}, \ldots, x_{iN}) \) is a feature vector of item \( i \) (see Equation (4.1)). Then, the proposed items are
5.2 Group Discussion Simulation Model

- **Group Simulator**
  1. Randomly form a group $g$
  2. For each scenario, create true utility vector $v_u$, $\forall u \in g$ as in (5.1)
  3. Propose the item with the highest true utility $v_u$ at iteration $t$
  4. Evaluate items proposed by other members by classifying them into: $BS(u)$, $LS(u)$, $NS(u)$ and $DS(u)$ as in (5.2)

- **Group Recommender**
  5. Infer the user’s constraints $\phi_u$ derived from $BS(u)$, $LS(u)$, $NS(u)$ and $DS(u)$ as in (4.3)
  6. Find $w_u$ by solving the optimization problem (4.4)
  7. Update $w_u$ with the linear combination as in (4.5)
  8. Compute $w_u = \frac{1}{|g|} \sum_{u \in g} w_u$
  9. Rank items according to estimated group utility $U(g, i)$ as in (4.8), resulting in group recommendation list

![Fig. 5.1 The computation sequence in the simulated group discussion](image)

Classified into the four sets $BS(u)$, $LS(u)$, $NS(u)$ and $DS(u)$ using the following conditions:

$$i \in \begin{cases} 
    BS(u) & \text{if } U^v(u, i) \geq u_1(u) \\
    LS(u) & \text{if } u_2(u) \leq U^v(u, i) < u_1(u) \\
    NS(u) & \text{if } u_3(u) \leq U^v(u, i) < u_2(u) \\
    DS(u) & \text{if } U^v(u, i) < u_3(u). 
\end{cases}$$

(5.2)

For each user $u$, the threshold values $u_1(u)$, $u_2(u)$ and $u_3(u)$ are determined in a way that the full set of items evaluated by the user is equally divided into the four subsets, i.e., four subsets of equal cardinality.
Figure 5.1 illustrates the sequence of computations for a simulated group discussion. As we can see in the illustration, there are two logical concepts: Simulator and Recommender. Each group discussion runs for \( t \) iterations, and the blue box specifies associated steps in each group iteration that starts with a proposal made by one group member and ends with feedback on that proposed item given by the others. In essence, the Simulator is responsible for generating user-related actions, i.e., forming a group (Step 1) and defining true utility vector \( v_u \) for each user in the different group scenarios (Step 2). The Simulator then employs the true user’s utility vector to identify what items are proposed to the group discussion (Step 3) and how the user evaluates the items proposed by other group members (Step 4). It is important to note that the true utility vectors are unknown to the Recommender. In fact, the Recommender only takes the simulated users’ feedback as the inputs (Step 5), from which it learns the session-based preferences \( w^g_u \) (Step 6) and then updates the utility vector \( w_u \) for the user (Step 7). In other words, given the simulated user’s evaluations, its goal is to recover the hidden true utility vector \( v_u \) for each user in the specific group settings, so the closer the estimated utility vector \( w_u \) to the true one \( v_u \), the better performance it achieves. Finally, the Simulator and Recommender generate the group utility vectors, respectively denoted with \( v_g \) (Step 8a) and \( w_g \) (Step 8b). The assumed group choice is the item obtaining the largest true group utility function \( U^v(g,i) \) (Step 9a), which is later used to evaluate the quality of generated group recommendations (Step 9b).

5.3 Experiment Setup

The research goal is, for each simulated scenario, to compare the efficacy of the ranked recommendations generated by the interactive group model implemented with three combination strategies of long-term and session-based preferences. In the experimental evaluation, we examine how the ranking quality is dependent on group sizes and on the number of interaction cycles in which users’ feedback on the proposed items is elicited.

Evaluation Metric

The efficacy of the ranked list of group recommendations is measured by Mean Reciprocal Rank (MRR) [65], which considers the rank of the first relevant result, and it is therefore well suited to GRSs that matter the first group choice. The metric is defined as follows:

\[
MRR = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{rank_k},
\]
where \( n \) is the number of running trials, which is 100 in our experiment, and \( \text{rank}_k \) refers to the rank position in the recommendation list of the assumed group choice (i.e., the item having the largest true group utility) for the \( k \)-th trial. The value of MRR ranges from 0 to 1 and the higher value is the indication of the better performance.

In fact, aside from MRR, the quality of the group recommendations was also measured by other metrics. Particularly, we computed the cosine similarity between the assumed group choice and the top recommended item (i.e., the first-ranked item in the recommendation list), together with observing how the true group utility of the top recommended item changes, in comparison to the true group utility of the assumed group choice. The analysis of these results, however, did not reveal additional insights into the characteristics of the recommendation model, so we do not report them here, and we would rather focus on the analysis of the MRR measurement.

**Parameters**

To simulate the impact of the group on users’ behavior in the conversion and anti-conformity scenario, the value of parameter \( \gamma \) is set to 0.2 and \(-0.2\), respectively. We choose these relatively small values since in the scope of this work, we primarily focus on situations where group members are rather slightly influenced by the average group opinion.

As previously mentioned, for each simulated situation we compare the performance of three model variants operating with different values of the stability parameter \( \sigma \), which is used to balance the importance that the system gives to the long-term and session-based preferences. Obviously, many cases can be analyzed in principle, but here we restrict ourselves to three rather prototypical cases: when the model strongly relies on the long-term users’ preferences (\( \sigma \) is close to 1, i.e., \( \sigma = 0.9 \)), when an equal weight is given to both types of preferences (\( \sigma = 0.5 \)), and when the model strongly weighs the session-based preferences derived from the group session (\( \sigma \) is close to 0, i.e., \( \sigma = 0.1 \)).

For each group discussion, the number of interaction cycles varies from \( t = 1 \) to \( t = 10 \). The larger the value of \( t \), the larger the amount of users’ feedback is acquired by the system. We inspect groups of different size from 2 to 5 users, and for each group size, 100 groups are randomly generated (i.e., 100 running trials). The average MRR of these 100 groups give the final result. The results obtained from the different group sizes, however, turn out to be qualitatively similar, we hence only report the results for groups of size 2 and 5.
5.4 Results

We have so far described how to simulate users’ behavior in the three group scenarios, i.e., independence, conversion and anti-conformity. In each scenario, we aim to investigate their effect on the quality of recommendations generated by the three preference combination rules using the different values of parameter $\sigma$ to control the relative importance of long-term and session-based preferences ($\sigma = 0.1, 0.5,$ and $0.9$). The larger value of $\sigma$ indicates that the long-term preferences are considered to be more important. The experimental analysis is conducted with an increasing number of interaction cycles ($t = 1 \ldots 10$). If the group recommendation model is sound, then as more preferences are elicited, i.e., more interaction cycles, it should perform better.
The results illustrated in Figure 5.2a and 5.2b support our initial intuition: the average effectiveness (MRR) of the model is growing when the number of interaction cycles grows. This occurs because, as \( t \) increases, more constraints on the users’ session-based preferences are inferred, hence the model better learns the new utility function of each group member. We observe that the system performance is significantly superior for smaller groups: the MRR is higher and grows faster than that obtained by the larger groups with the same number of interaction cycles. This is partially unexpected, since according to our simulation procedure in a larger group size a greater number of elicited preferences is obtained as a user evaluates all the proposals of the other group members. The explanation could be due to the diverse users’ preferences, which is somewhat confirmed by other experiments showing that when the group size increases, the effectiveness of the group recommendations tends to decrease, if the groups are composed of heterogeneous users [8]. Overall, we see that it is easier to generate good recommendations for groups of size 2 than for groups of size 5, and this holds for all the considered scenarios and preference combination strategies.

Comparing the performance of the three preference combination rules, i.e., using three values of \( \sigma \) in the group model, under the three group scenarios, we observe that the best combination strategy varies depending on the group setting. We additionally perform a two-way ANOVA analysis to test whether the differences in MRR are significant, considering the variation in the combination strategies of long-term and session-based preferences and the group scenarios as independent variables. As shown in Table 5.1, the results obtained from groups of size 5 lead to the conclusion that the combination value \( \sigma \) as well as the group scenarios significantly impact the ranking quality of group recommendations in terms of MRR. The differences are also statistically significant for the other group sizes.

| Table 5.1 Table of ANOVA2 analysis for groups of size 5 |
|-----------------|--------|----------------|--------|--------------------|-----------------|
|                 | Df     | Sum Sq        | Mean Sq | F value          | Pr(>F)          |
| \( \sigma \)    | 2      | 11.17         | 5.59    | 5492             | <2e-11 ***      |
| scenario         | 2      | 41.09         | 20.55   | 20202            | <2e-11 ***      |
| \( \sigma \):scenario | 4      | 11.82         | 2.96    | 2906             | <2e-11 ***      |
| Residuals        | 891    | 0.91          | 0.001   |                   |                  |

**Independence scenario.** The results in Figure 5.2a and 5.2b demonstrate that when the group has no impact on user’s preferences, the model, with \( \sigma = 0.9 \), after a few interactions (\( t = 4 \) and \( t = 6 \)), can correctly rank the group choice in the top position while with \( \sigma = 0.5 \) and \( \sigma = 0.1 \), the system needs more than 10 interaction cycles. Since in this setting, the users do not change their preferences during the group discussion, the system that highly
favors the stable preferences can immediately identify the true utility of users after a few interactions. However, it is noteworthy that adapting to the session-based preferences incurs an initial cost even in this independence scenario: the recommendation quality in the very first interaction cycles (i.e., when \( t \leq 4 \) and \( t \leq 6 \) in the groups of size 2 and 5 respectively) is still much inferior to that obtained by the model that totally ignores them.

**Conversion scenario.** In this scenario, group members tend to align their preferences, hence one could conjecture that in this case it is easier for the recommender to identify the group choice. This hypothesis is partially supported as both Figure 5.2a and 5.2b illustrate that the MRR of the group recommendations generated when \( \sigma = 0.5 \) and 0.1 is better than the MRR obtained in the independence scenario, but this does not hold for \( \sigma = 0.9 \). Particularly, in the groups of size 2 and when \( \sigma = 0.9 \), the recommender requires at least 9 interaction cycles to place the group choice at top rank (MRR \( \approx 1 \)) whereas in the first scenario only 4 cycles are needed. On the other hand, with \( \sigma = 0.5 \) the system gains a higher MRR sooner (\( t = 5 \) for MRR \( \geq 0.75 \)) while in the independence scenario, 8 interaction cycles are required to reach that value. Similarly, when \( \sigma = 0.1 \), the system requires fewer interaction cycles than in the previous scenario, to obtain a similar performance level.

These results can be explained by observing that in this scenario the users’ session-based preferences deviate from the long-term ones, hence mostly relying on the individual preferences (\( \sigma = 0.9 \)) is no longer the ideal approach for the recommender. It turns out that the recommender in this case benefits from keeping an equal balance between long-term and session-based preferences.

**Anti-conformity scenario.** In this last situation, group members are diverging during the group discussion, so our intuition says that it should be harder for the learning process to infer the true utility vector \( v^{(u)} \). The results shown in Figure 5.2a and 5.2b meet our expectation as we can see that the MRR in this setting is much smaller than in the independence and conversion scenarios, and it is also growing much slower. The results provide empirical evidence that it is more difficult to find an item that satisfies all group members whose preferences are diverse. Interestingly, as opposed to the prior simulations, the MRR of group recommendations generated by the lower value of \( \sigma \) is higher than the MRR obtained by the larger value of \( \sigma \). This signals that in this case the system should weigh more the preferences inferred from observing the discussion.
5.5 Discussion and Conclusions

In this chapter, we argue that the relative importance of the long-term and session-based preferences should depend on the specific group settings. To verify this claim, we have proposed a generic simulation procedure that enables us to compare the performance of the three preference combination strategies in the different group scenarios where users are likely to behave under the group influence. The simulated users’ behavior are specifically motivated by different forms of social responses with respect to conformity (conversion) and nonconformity (independence and anti-conformity). The analysis of the interactive group recommendation model has been extensively conducted across the amount of elicited feedback and group sizes.

The empirical evidence has shed light on some interesting behavior of the proposed group recommendation model and supports our claim that the appropriate exploitation of long-term and the session-based preferences is dependent on the specific group scenario. In particular, we have observed that the combination scheme that weighs more the long-term preferences can better serve the group when the group discussion has no impact on group members’ preferences, but when the group context pushes members to have either more or less similar preferences, then users benefit from a GRS that weighs more the session-based preferences. The analysis further shows that it is by far not an easy task to dynamically and optimally adapt the system-generated recommendations to the unknown effect of a group on the evolving users’ preferences, i.e., when it is unclear which users’ interactions originate from their own preferences and which are influenced by the group setting.

In addition, the experimental results support the goodness of the proposed model in correctly capturing the changes of the user’s needs as more feedback is provided, but having said that, the preference learning model calls for the necessity of identifying an automatic optimization mechanism. This goal is even more challenging considering other group factors related to a group decision making problem, such as: the specific relationships between users, and their effects on preferences, or the users’ personal attitudes towards expressing their opinions on the discussed items, or even the specific roles that can be played by particular users (e.g., the presence of dominant members).

Besides, our simulation experiments show that there is still much room for improvement. First of all, how users interact in a group setting so far has been simulated in a deterministic way, i.e., it works on the assumption that they invariably prefer the items with the largest utility. This is, however, not always the case, since according to the analysis of the logged data collected in group decision making processes that have been supported by our system STSGroup, the individuals’ feedback is non-deterministic (see Section 4.5). It is, therefore,
essential that the simulation model fits the real data observation and take into account the uncertainty in users’ responses that might depend not only on their preferences, but also on their personality. Moreover, we have investigated only the case when group members’ attitudes towards evaluating the discussed items are uniform (i.e., either independence, conversion or anti-conformity) and they are all willing to give feedback, which is not a very realistic case. Therefore, the next chapter is dedicated to enhancing the group discussion simulation model, so that we can explore the characteristics of the interactive group recommendation model in more diverse simulation settings and with various group compositions.
CHAPTER 6

Simulating Group Discussions in Conflict Situations

In this chapter, we study, by simulations, the effect of an individual’s conflict resolution style on a group decision making outcome supported by a recommendation process. We first present a group discussion simulation process wherein we model users’ behavior along with their different conflict resolution styles. We then describe our experimental setup. Finally we discuss the results obtained from the simulated group discussions, as well as from an analysis of the quality of the group choice in real groups as a function of the conflict resolution style of the group members.

6.1 Research Motivation

A group decision making process, by its very nature, embraces conflict as an unavoidable consequence since different people may have diverse outlooks and preferences, which could pull them in different directions. In conflict situations, group members could either take into account the interests of the others or outright ignore them. This kind of behaviour, as conceptualized in the TKI model [49, 88], can be characterized by two basic dimensions, namely assertiveness and cooperativeness, that are, in essence, the extent to which the individual attempts to satisfy their own and other person’s preferences, respectively. These two dimensions of a user’s behavior are used to identify four possible conflict resolution styles: accommodating, competing, avoiding, and collaborating.

The previous research on group recommendations, unfortunately, has not analyzed the impact of the users’ conflict resolution styles on the outcome of the decision making process supported by a GRS. Moreover, the effect of the conflict resolution styles could be interconnected with various group factors. It is not clear, for example, whether the length of group interactions and group size relate to that effect. Also, it is unknown what the quality of the outcome will be if group members tend to compete against each other even when they share
similar preferences, or conversely, if they have diverse tastes but they also take into account the others’ preferences when they interact with the other group members. Inspecting the impact of these interlinked group factors on the quality of group recommendations, however, is hard to be done by the analysis of real users due to the large variety of group situation possibilities that would need to be extensively evaluated. Moreover, for an interactive GRS, the evaluation is even more problematic as it demands a high degree of interactivity.

To this end, we demonstrate that simulation-based experiments can be effectively exploited to address the aforementioned issues. Specifically, we propose a group discussion simulation procedure, in which we model the dynamics of individuals’ actions in alternative group discussion situations characterized by four group factors, namely, conflict resolution style, inner-group similarity, length of group interaction and group size. The interaction flow of the simulated group discussion is influenced by the observations of real users’ actions while interacting with our system called STSGroup that we have implemented and introduced in Chapter 4. The goal is to retain as much as possible the essential features of a realistic system-mediated group discussion. The effect of these group factors on the recommendation quality is measured by observing the individual’s loss in utility for choosing a recommended group choice rather than the choice that the individual would make, if not constrained by the group context, which is motivated by the principle of process losses in social psychology. This focuses on measuring the reductions in performance efficacy caused by group settings that hinder individuals from reaching their full potential [39]. We also measure the difference between the maximum and minimum utility, received by the users in the group when they take the group choice. This measure describes the fairness of the group recommendation identified by the system.

The experimental results show, among other findings, that if group members have similar tastes then groups composed of users with competing conflict resolution style obtain the largest utility loss, compared to groups whose members adopt cooperative styles, and yet, whatever their conflict resolution style is, there is no distinct difference in their utility for the group choice because they are treated equally. Conversely, when group members have diverse preferences, the average utility loss of competing members for the group choice is still the largest, but the differences in their utility is the lowest, since they all get a similar but lower utility. The findings of our simulation experiments are also supported by an analysis of real group discussions derived from the observational study on group decision making process, which has been presented in Chapter 3. We are convinced that the methods and insights gained from this work are beneficial for researchers and practitioners in the field of GRSs. In fact, they can help to formulate new conjectures and test various properties of a group discussion before conducting a user study.
6.2 Group Discussion Simulation Model

To find out how the considered four group factors, i.e., *conflict resolution style, inner-group similarity, length of group interaction* and *group size*, influence the quality of the group recommendation process, we have designed a group discussion simulation model which is justified and informed by the analyses of the real users’ interactions with our system STSGroup (see Section 4.5). We will illustrate the simulation model by specifying the two main steps of the procedure: (i) how individuals behave and respond in conflict situations, and (ii) how a group discussion is proceeding. On top of that, prior to the description of a simulated group discussion process, we will clarify the differences between the simulation models proposed in this work and in Section 5.2.

6.2.1 Individual’s Conflict Resolution Style

By reusing the TKI model [88], we identify different conflict resolution styles that group members can adopt. These styles are defined on the grounds of two fundamental dimensions: assertiveness and cooperativeness. In the simulation model, we implement the assertiveness as the intensity with which the simulated user expresses their own opinions, and cooperativeness as the extent to which the user tries to satisfy the other members’ concerns. According to TKI, the four styles of dealing with conflicts are:

- *Accommodating* is unassertive and cooperative: an individual is inactive in pursuing their own concerns and tries to satisfy those of the other members.

- *Competing* is assertive and uncooperative: an individual pursues their own concerns in an active way and refuses to accept proposals made by other members.

- *Avoiding* is unassertive and uncooperative: an individual is inactive in pursuing their own concerns, but also does not attempt to gratify those of the others.

- *Collaborating* is assertive and cooperative: an individual is active in following their own concerns and, at the same time, tries to satisfy those of the other members.

In fact, the TKI model has a fifth style called *Compromising* that is moderate in assertiveness and cooperativeness but, for the sake of simplicity, we do not include this style, rather we examine a *Baseline* case, i.e., a simulated user who behaves like an average user in the analyzed STSGroup observations.

In our simulation, we have decided to model assertiveness with the probability that group members propose items related to their concern. In that sense, that probability is increased if
Table 6.1 Summary description of the adopted design for simulating the TKI conflict resolution styles

<table>
<thead>
<tr>
<th>Style</th>
<th>Assertiveness</th>
<th>Cooperativeness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Propose</td>
<td>Evaluate positively</td>
</tr>
<tr>
<td>Competing</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Accommodating</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Avoiding</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Collaborating</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Baseline</td>
<td>Use the probability derived from the STSGroup observations</td>
<td></td>
</tr>
</tbody>
</table>

+ (increase) / − (decrease) the probability

the members are assertive and decreased otherwise. Conversely, cooperativeness is modeled by the probability of giving positive or negative evaluations. Users with a cooperative style have a higher probability of giving positive feedback (i.e., best choice or like) and a lower probability of giving negative feedback (i.e., dislike). The opposite holds for the uncooperative users. A summary of the model that we design for simulating the conflict resolution styles is provided in Table 6.1. On the basis of the model, we simulate the group discussion procedure described later.

**Assertiveness Dimension**

To simulate the intensity of user involvement in defending their own concern, each simulated user is characterized by a probability $p(u)$: the probability that, in an iteration of a group discussion, $u$ will propose one of their favorite items to the group. More assertive users are supposed to have larger $p(u)$ probabilities. We have estimated the probability that a generic user of our user study made a proposal to their group as a baseline point. According to our data analysis (Section 4.5), the number of items proposed by participants is about equal. The probability that one group member makes a proposal in a group discussion iteration is therefore, in baseline case, the same for all members $\forall u \in g, p_p(u) = \frac{1}{|g|}$.

Then, to simulate more or less assertive users, that base probability, in the simulation, is increased or decreased to some extent: $p_p(u) \pm \gamma, \gamma > 0$. The new probabilities are re-normalized accordingly, by dividing them by $\sum_{u \in g} p_p(u)$, so that the sum of the probabilities of all group members is equal to 1.

**Cooperativeness Dimension**

The cooperativeness level of each user is modeled in the simulation by increasing (with respect to the baseline probability) the probability that they will give a positive evaluation
6.2 Group Discussion Simulation Model

Fig. 6.1 An illustration of how the probabilities that a user will give positive and negative evaluations can be transformed with respect to the level of cooperativeness

(i.e., best choice or like) and simultaneously diminishing the probability of giving a negative feedback (i.e. dislike) to an item proposed by other users. For uncooperative users, the simulator applies an opposite strategy.

To transform the probabilities of giving a particular feedback given the estimated user utility for an item, we use an exponential cumulative distribution function (see Equation 6.1) that is illustrated in Figure 6.1a and 6.1b.

\[
f(x) = \frac{b^x - 1}{b - 1}, \quad b \neq 1
\]  

Here \( p_D, p_L \) and \( p_B \) are respectively the baseline probabilities that an item is assessed as dislike, like and best choice, given its utility. We set \( x_1 = p_D, x_2 = p_D + p_L, \) and \( x_3 = p_D + p_L + p_B = 1. \) The transformed values of \( p_D, p_L \) and \( p_B \) are denoted by \( q_D, q_L \) and \( q_B \) respectively (\( q_D = f(x_1), q_L = f(x_2) - f(x_1), \) and \( q_B = 1 - f(x_2), \) so that \( q_D + q_L + q_B = 1). \)

As we can see in the uncooperative case with \( b < 1, \) the function grows \( q_D \) (i.e., from \( p_D = 0.41 \) to \( q_D = 0.68 \)) and concurrently reduces \( q_L \) and \( q_B \) (i.e., from \( p_L = 0.38 \) to \( q_L = 0.25 \) and from \( p_B = 0.21 \) to \( q_B = 0.07 \)). In the cooperative case with \( b > 1, \) by contrast, it decreases \( q_D \) (i.e., from \( p_D = 0.41 \) to \( q_D = 0.17 \)) and simultaneously increases \( q_L \) as well as \( q_B \) (i.e., from \( p_L = 0.38 \) to \( q_L = 0.4 \) and from \( p_B = 0.21 \) to \( q_B = 0.43 \)).

Originating from the observed data indicating that people give feedback on approximately 67% of the proposed items, each simulated member is represented by the probability that they will evaluate an item proposed by the others, denoted with \( p_f(u) = 0.67. \)
6.2.2 Group Discussion Procedure

The overall structure of a simulated group discussion procedure is shown in Figure 6.2, which comprises two logical components: Group Simulator and Group Recommender. The blue box further indicates steps associated with each discussion cycle, aka iteration.

Differences in Simulation Procedure

In general, the group discussion simulation process proposed in this work shares the common structure with the one previously described in Section 5.2. Having said that, however, they basically differ from each other in how Group Simulator and Group Recommender operate. We hereby highlight the differences while the details of the proposed procedure will be explained afterwards.

With respect to Group Simulator component

- **User Groups.** In the prior work, users’ profiles were built by using the real ratings given by real users while interacting with STS app, and then, virtual groups were randomly created by sampling from these users’ profiles. However, due to the fact that
STS is a CARS dedicated to individuals, we could not access any information about the interaction of these real users in group settings. Hence, we analyzed data logs collected through the user study of our system STSGroup, in which real users’ interactions during their group discussions were automatically recorded (see Section 4.3), but still, the shortcoming of using STSGroup data is the limited number of observed users. To this end, in this simulation procedure, users’ profiles are synthetically constructed by drawing from a distribution that is backed up by the collected real data from the system. Also, we examine two types of groups in terms of the inner-group similarity, i.e., groups of users with similar and diverse tastes.

- **Item proposals.** The previous process was based on the assumption that for each iteration of the simulated group discussion, every single member always proposes the items obtaining their largest “true” utility, i.e., the one that is assumed to reflect the true preferences of each member in a particular group setting, and it is predefined before the start of the discussion. The new procedure, conversely, selects who will make an item proposal based on the extent of their assertiveness, which is modeled by the probability \( p_p(u) \) (see Section 6.2.1). Here only one member is selected at a time, and the selected member proposes an item having the highest user’s utility that is revised after each interaction cycle.

- **Item evaluations.** In the previous process, all the simulated members were evaluating the items offered by the others in a deterministic way, i.e., they invariably favor the ones having the largest value according to their true utility. In this new model, members who will give feedback are chosen by the probability \( p_f(u) = 0.67 \). The value is defined by the proportion of people who provided feedback on the items proposed by the other members in the observed group discussions (see Section 4.5). Then, how the members evaluate the proposals is based on their level of cooperativeness, characterized by the probabilistic model (see Section 6.2.1).

**With respect to Group Recommender component**

- **User’s utility update.** The prior group recommendation model is motivated by a prosocial orientation concept coming from social psychology, in which group members strive to maximize their own outcomes and the others’ outcomes as well [39]. For that reason, each member’s session-based preferences, evolved during the simulated discussion, are inferred by finding the user utility vector that not only satisfies their individual preferences but also aligns as much as possible with those of the oth-
Finally, the updated utility vector is the linear combination of the long-term and session-based preferences (see Section 4.2).

In the new model users are driven by an individualistic orientation and they tend to maximize their own outcomes [39]. This modification is motivated by the goal of this new study, which puts the focus on investigating how the individual utility function of group members is affected by the group setting when they adopt different conflict resolution styles. Thus, the updated utility vector of each member during the simulated group discussion is assumed to be the one that satisfies the constraints derived from their evaluations given in the interaction session and is as close as possible to their original utility vector, which reflects their individual long-term preferences. The resulting optimization problem is formulated as follows:

$$w^g_u = \arg \max_w \cos(w, w_u) \text{ subject to } w \text{ satisfies } \phi^g_u,$$

(6.2)

where $w_u$ models the individual preferences for the $N$ item features, $\phi^g_u$ is the set of constraints on the utility function of $u$ inferred from the evaluations given by $u$ in the simulated group discussion $g$, and $w^g_u$ is the updated utility vector of $u$ in $g$. The utility score of $u$ for each item $i$ in the group discussion, thereby, is re-calculated:

$$U^g(u,i) = \sum_{j=1}^{N} w^g_{uj}x_{ij},$$

(6.3)

where $x_i = (x_{i1}, \ldots, x_{iN})$ represents each item $i$.

As we aim to quantify the "price" (i.e., utility loss) one has to pay in selecting the collective choice instead of the personal choice, together with the distinction between the maximum and minimum achievable utility of the group choice, the evaluation must be performed in relation to the group-independent preferences of the members, i.e., the individual user utility function, i.e., $U(u,i) = \sum_{j=1}^{N} w_{uj}x_{ij}$ (see Section 6.3.2). In other words, we do not employ the updated utility function $U^g(u,i)$ for the evaluation as it is already adjusted to the changing behavior of the members in the group setting.

- **Group utility function.** In principle, after having updated the user's utility function, the group utility score for an item, denoted by $U(g,i)$, can be generated by applying several individual preference aggregation strategies. In the previous group recommendation model, we considered only the Average preference aggregation strategy while in the current model, we have implemented three preference aggregation rules, i.e., Average, Borda count and Multiplicative.
Average: The updated utility functions of the group members are averaged:

\[ U(g,i) = \frac{1}{|g|} \sum_{u \in g} U^g(u,i). \] (6.4)

Borda count. Items are independently ranked by the updated utility functions of the members \( U^g(u,i) \). Hence each item will have a user-dependent rank denoted as \( \text{rank}_{ui} \), where the top rank, 1, is taken by the item with the largest user utility. The group score for each item is calculated as the sum of the reciprocal of the individual rank:

\[ U(g,i) = \frac{1}{|g|} \sum_{u \in g} 1/\text{rank}_{ui}. \] (6.5)

Multiplicative. The updated utility scores of group members are multiplied to generate the group utility for an item:

\[ U(g,i) = \prod_{u \in g} U^g(u,i). \] (6.6)

Procedure Description

As illustrated in Figure 6.2, to simulate the process of a group discussion, the responsibility of the Group Simulator boils down to the following sequence of steps:

1. **Forming a group.** We have decided to generate synthetic users’ profiles, which, however are similar to the profiles of the real users who participated in the STSGroup evaluation. More precisely, with the users’ ratings elicited before any actual group discussions taking place, the simulator initially computes the users’ utilities using Equation (4.2). Then, it applies \( k \)-means clustering on the sample set of the real users’ utilities, where the number of clusters, \( k \), is determined by the maximum group size we considered, which is 5 in our experiment. The synthetic users’ utilities, thereby, are drawn from a distribution with mean and variance of each cluster. This allows the simulation experiment to build groups composed of users with similar and diverse tastes. For the former, the simulator arbitrarily selects a cluster, and then generates each group member’s utility vector from this selected cluster distribution. For the latter, the simulator derives the members’ utilities from different clusters.

2. **Making a proposal.** After a group is composed, the simulator randomly selects a member and decides if this member makes a proposal with probability \( p_p(u) \), whose
value varies depending on the degree of assertiveness. If not, the selection is started again. Afterwards, the simulator identifies what the member will propose to the group discussion by considering the items with the highest user utility. It is noteworthy that the proposed items by each simulated member are selected according to their updated utility $w_u^g$, i.e., the best estimate of the session-dependent user utility.

3. **Giving feedback.** The simulator then uses the probability $p_f(v), \forall v \in g \land v \neq u$ as a basis for selecting who among the remaining group members will give feedback on the item proposed by $u$. The given evaluation is determined by the cooperativeness of the specific conflict resolution style modeled in Section 6.2.1.

4. **Holding a group discussion.** As part of our study is to understand how the number of interaction cycles relates to the impact of conflict resolution styles on the group recommendation performance, the simulator considers two different settings: (i) the discussion runs for 10 iterations, and (ii) the discussion stops as soon as the first group choice is found.

At each iteration, the outcome of the simulator, the simulated users’ evaluations, is used as input to the *Group Recommender*. The recommender then revises the members’ utility models, which are derived from their feedback revealed during the simulated group discussion. It then combines individual utilities to generate the group utility score $U(g,i)$ that is used to rank the discussed items. The simulation assumes that the collective choice is the item, among those discussed in the group, that has received a *best choice* feedback from all of the group members. If none of the proposed items obtains this type of feedback, then there is no group choice, and we assume that this produces no utility to the group members. Hence, in that case, the average individual loss, is taken as the mean utilities of the top individual choices (since the group choice has zero utility). When we analyze the group choice performance at various iterations (i.e., $t = 1 \ldots 10$), we assume that the group choice at iteration $t$ must be ranked at least equal to or higher than the previous choice at iteration $t - 1$, according to the group utility score $U(g,i)$. Otherwise a rational group should keep the preceding one. Simply put, the group choice at a successive iteration must always improve a previous choice since no rational group should replace a previously made choice with another having inferior utility.

This decision about how to simulate group choices is informed by our observations of the discussion made in the user study wherein, most of the time, the group discussion ended, after some iterations, when one item received positive evaluations (e.g., *best choice* or *like*) from all the group members.
6.3 Experiment Setup

The experiment was conducted by running simulated group discussions and generating data describing the user behavior in simulated situations. We analyzed a number of alternative group settings defined by the independent variables: inner-group similarity, conflict resolution style, length of group interaction, group size and preference aggregation strategy. The dependent variable is the quality of group recommendations measured by two metrics: Mean Individual Loss — MIL, i.e., the average of the extent to which the user’s utility of the collective choice differs from that of the personal choice; and Max Min Difference — MMD, i.e., the difference between the maximum and minimum user’s utility.

6.3.1 Independent Variables

To obtain a clear picture of the impact of each analyzed factor, we employed a factorial design, as summarized in Table 6.2, which shows the possible combinations of the values of the independent variables.

We ran simulations with two alternative levels of inner-group similarity, namely homogeneous and heterogeneous groups where individuals have similar and diverse tastes respectively. As previously explained, when describing how the groups were composed, groups of like-minded users are sampled from the same cluster while groups of members with varied interests are obtained from different clusters.

We investigated the group discussion outcome in groups characterized by having a uniform conflict resolution style, i.e., where all group members have the same conflict resolution style, and also mixed groups, where users with a competitive attitude are mixed with users with another conflict resolution style. The five uniform groups are composed of users with conflict resolution styles: competing, accommodating, avoiding, collaborating and baseline. For the mixed-style groups, potentially there are many possible combinations, but as pointed out by studies in social psychology [39], competitive situations are likely to intensify conflict between individuals. We thereby focus on the performance involving competition. We specifically considered four mixed combinations in which competing simulated users are respectively paired with accommodating, with avoiding, with collaborating, and with baseline simulated users.

Regarding the length of group interaction, we examined two scenarios when the group discussion has an increasing number of iterations, up to 10, and when the group interaction is terminated as soon as a group choice is determined, i.e., when a discussed item is getting best choice feedback from all group members.
Table 6.2 Overview of the employed independent variables

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Number of levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner-group similarity</td>
<td>2</td>
<td>Homogeneous and heterogeneous groups</td>
</tr>
<tr>
<td>Conflict resolution style</td>
<td>9</td>
<td>5 uniform and 4 mixed styles</td>
</tr>
<tr>
<td>Interaction length</td>
<td>2</td>
<td>Varying interaction length and stopping</td>
</tr>
<tr>
<td>Group size</td>
<td>4</td>
<td>Groups of size 2, 3, 4 and 5</td>
</tr>
<tr>
<td>Preference aggregation strategy</td>
<td>3</td>
<td>Average, Borda count and Multiplicative</td>
</tr>
</tbody>
</table>

With respect to the group sizes and preference aggregation strategies, we studied groups of different size from 2 to 5 users and implemented three preference aggregation strategies: Average, Borda count and Multiplicative.

Each of the considered group situations was evaluated in a series of 200 running trials. On each trial, the discussion process described in Section 6.2.2 was performed. We report the average results of these 200 trials.

6.3.2 Dependent Variables

The main goal of our research is to assess and understand the potential effect of the dynamic changes in users’ responses prompted by different conflict resolution styles on the group recommendation performance in various group situations. We, therefore, have inspected the quality of a decision making outcome supported by the recommendation process by measuring: i) the variation in individual’s utility loss (i.e., Mean Individual Loss — MIL), and ii) the difference between the utility of a winner and a loser (i.e., Max-Min Difference — MMD), where winners and losers are defined as the simulated users that receive the maximum and minimum utility for the collective choice respectively. This makes sense in the context of GRSs, as group recommendations that aim to simultaneously suit the preferences of all group members are hardly as good as those specifically tailored to individuals. Moreover, the occurrence of conflict between user’s interests during the decision making process often ends up with one side branded the winner and the other side the loser, so this motivates our decision to observe the difference in utility between the winners and the losers that evolve during the simulated discussion.

Mean Individual Loss (MIL)

This metric measures how much an individual loses in utility for a joint group choice with respect to their personal (optimal) choice. In particular, for each simulated user we estimate
their individual loss by comparing their utility for the collective group choice $i_g$ to their utility for the optimal individual choice $i_u$ (the utility that a user can obtain by choosing the item with the highest utility). We take the average of all the group members’ utility losses defined as in Equation 6.7:

$$MIL(g, i_g) = \frac{1}{|g|} \sum_{u \in g} U(u, i_u) - U(u, i_g). \quad (6.7)$$

It is important to recall that the personal choice is the item having the highest utility with respect to the original user utility vector $w_u$, the one that is built directly from the individual preferences independent of group settings. As previously explained, by studying the behavior of this metric, we can judge the deviation from the personal optimal choice caused by the changes in users’ behavior that are invoked by their conflict resolution style.

**Max-Min Difference (MMD)**

This metric assesses the discrepancy between the winner’s and the loser’s utility for the collective choice. It is inspired by the definition of “satisfied winner” and “dissatisfied loser” coming from the prior research on individual satisfaction with group decisions in which the experience of winning (satisfied) or losing (dissatisfied) is defined by the match or mismatch between a group choice and the users’ initial preferences [34]. Here in our case, the winner and the loser of a group decision making process are defined as those whose utility for the group choice is the highest and lowest, accordingly. Similarly to the first metric, driven by individual oriented perspective, the comparison is made with respect to the individual user’s utility function defined on the independent user’s preferences, acquired without any influence from a group.

$$MMD(g, i_g) = \max_{u \in g} U(u, i_g) - \min_{u \in g} U(u, i_g). \quad (6.8)$$

### 6.4 Results

We hereby report the dynamics of the group recommendation performance measured by the metrics MIL and MMD for the considered group situations. For the sake of brevity, users that are, and groups composed by, *accommodating* or *collaborating* individuals are referred to as cooperative users and groups, respectively; while users that are, and groups composed by, either *avoiding* or *competing* individuals are called uncooperative users and groups, respectively.
6.4.1 Experiments with homogeneous groups

Uniform conflict resolution styles

In this section, we examine groups whose members have similar preferences and conflict resolution styles.

Varying the interaction length. Figure 6.3 displays the MIL for each conflict resolution style along $t = 10$ interaction cycles for the three preference aggregation strategies (Average, Borda count and Multiplicative) that we have considered. We observe that the average loss in cooperative groups is much lower than in uncooperative ones. In small cooperative groups (i.e., groups of size 2 and 3), group members are also able to find an ideal group choice, the item with MIL = 0, after few interactions, whereas this does not hold for uncooperative groups. Moreover, we see that large uncooperative groups have high MIL and the group interactions do not help in reducing it.

These results can be explained by the fact that accommodating and collaborating users are likely to give their best choice evaluation to the items proposed by the other members, plus their tastes are similar in this case. Quite the opposite, there is a high chance that avoiding or competing users do not like the items suggested by the others even when these suggestions are aligned to their preferences (i.e., they have rather high utility for them). As it turns out, they get the greatest utility loss, and their MIL hardly goes to zero at successive iterations. This is true especially in large group sizes (i.e., groups of size 4 and 5).

The simulation results obtained by using Borda count, Multiplicative and Average preference aggregation strategies are qualitatively similar. We thereby only report the obtained results for the Average preference aggregation rule from now on.

Figure 6.4 illustrates the MMD evolution; all these curves are constant and overlapping. This means that there is no difference in the utility of the winner and the loser, regardless of the user conflict resolution style, group size and number of interactions. This is quite clear because, as long as the group members have the same preferences they must not have different utilities for the same group choice.

To test whether or not the differences in MIL are statistically significant, we used a non-parametric Kruskal-Wallis test after the normality and homogeneity of variance assumption, respectively checked with Shapiro-Wilk and Levene’s test, were not satisfied to perform one-way ANOVA test. The results show the significant differences between the conflict resolution styles in each group size ($p < 2.2e − 16$). From the output of the Kruskal-Wallis test, we continued to calculate Wilcoxon rank sum test with Bonferroni correction for pairwise comparisons between the five conflict resolution styles. The tests performed in all group
6.4 Results

Fig. 6.3 MIL of groups with similar interests and uniform conflict resolution styles
sizes indicate that most of the comparisons are significant \( p < 2.2e^{-16} \) except for the two comparisons between accommodating and collaborating \( p = 1 \) and between competing and avoiding \( p = 0.78 \). With respect to MMD, the statistical tests also confirm no significant difference between conflict resolution styles.

**Stopping the discussion at the first group choice.** In this setting, we analyze the quality of the first acceptable group choice. This means that the simulated group discussion stops as soon as the first choice is found.

Compared to cooperative groups, uncooperative groups, as expected, require more interaction cycles to find a collective choice and their average loss in terms of MIL is also higher. These results are illustrated in Figure 6.5a and 6.5b, accordingly. We also notice that the larger the group size is, the more iterations are needed to find the first group choice.

The Kruskal-Wallis test, again, confirmed that the difference in MIL between the considered conflict resolution styles was significant \( p < 2.2e^{-16} \). The results obtained from the pairwise comparison using the Wilcoxon test also lead us to the same conclusions as the previous setting (see Table 6.3 for the results of groups of size 2).

**Mixed conflict resolution styles**

We now examine the case when groups are composed of users with similar preferences, yet different conflict resolution styles.
Similar preferences – Equal conflict resolution style
Average preference aggregation method

(a) Mean number of iterations

(b) MIL

Fig. 6.5 Performance of groups with similar interests and uniform conflict resolution styles, stopping the discussion at the first group choice.
Simulating Group Discussions in Conflict Situations

Similar preferences – Unequal conflict resolution style
Average preference aggregation method

(a) MIL

(b) MMD

Fig. 6.6 MIL and MMD of groups with similar interests and mixed conflict resolution styles
6.4 Results

**Similar preferences – Unequal conflict resolution style**

**Average preference aggregation method**

![Graph showing mean number of iterations for different conflict resolution styles and group sizes.](image)

(a) Mean number of iterations of the mixed combinations

**Similar preferences – Unequal conflict resolution style**

**Average preference aggregation method**

![Graph showing MIL for different conflict resolution styles and group sizes.](image)

(b) MIL of the mixed combinations

Fig. 6.7 Performance of groups with similar interests and mixed conflict resolution styles, stopping the discussion at the first group choice
Table 6.3 Pairwise comparisons in terms of MIL for groups of size 2 whose members have similar interests and uniform conflict resolution styles, stopping at the first group choice

<table>
<thead>
<tr>
<th></th>
<th>compete</th>
<th>accommodate</th>
<th>avoid</th>
<th>collaborate</th>
</tr>
</thead>
<tbody>
<tr>
<td>accommodate</td>
<td>1.4e-8</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>avoid</td>
<td>1.00</td>
<td>5.5e-8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>collaborate</td>
<td>2.2e-10</td>
<td>1.00</td>
<td>8.4e-10</td>
<td>-</td>
</tr>
<tr>
<td>baseline</td>
<td>0.0031</td>
<td>0.0417</td>
<td>0.0126</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Varying the interaction length. We find out that mixed groups of competing and accommodating members have the lowest loss while the highest loss is scored by mixed groups of competing - avoiding users (see Figure 6.6a).

Regarding MMD, Figure 6.6b shows that the difference between the maximum and minimum utility of the members in a group is constant despite the varying combination of conflict resolution styles, group sizes and number of iterations. This result has the same explanation that we formulated for the previous setting.

The Kruskal-Wallis test results show that the difference in MIL of group members from mixtures of conflict resolution styles are significant ($p < 2.2e-16$). The Wilcoxon pairwise comparisons corroborate the significant effect of mixed groups in all the group sizes ($p < 2.2e-16$).

Stopping the discussion at the first group choice. Figure 6.7a shows that mixed groups of competing and avoiding need more interaction cycles, compared to the other combinations, in order to find the first group choice. Unsurprisingly, they also get the largest average loss, particularly in large group sizes (see Figure 6.7b).

In this setting, the significant differences in terms of MIL are found with $p = 2.4e-8$, $p = 7.6e-11$, $p = 1.6e-8$, and $p = 1.2e-12$ for groups of size 2, 3, 4 and 5, respectively. Comparing MIL of mixed groups of size 2 in pairs, we observe that there are no significant differences between Mix 1 (competing and accommodating) and Mix 3 (competing and collaborating) as well as between Mix 2 (competing and avoiding) and Mix 4 (competing and baseline) (see Table 6.4). Similarly, the difference between Mix 1 and Mix 3 is not significant in groups of size 4 and 5, but the difference is significant for groups of size 3 ($p = .000$). We also find out that there is no evidence that the MIL for Mix 3 (competing and collaborating) is statistically different from the MIL of Mix 4 (competing and baseline) in groups of size 3, 4 and 5.
Table 6.4 Pairwise comparisons in terms of MIL for groups of size 2 whose members have similar interests and mixed conflict resolution styles, stopping at the first group choice

<table>
<thead>
<tr>
<th></th>
<th>Mix 1</th>
<th>Mix 2</th>
<th>Mix 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mix 2</td>
<td>2e-7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mix 3</td>
<td>0.53</td>
<td>0.001</td>
<td>-</td>
</tr>
<tr>
<td>Mix 4</td>
<td>1.9e-5</td>
<td>1.00</td>
<td>0.028</td>
</tr>
</tbody>
</table>

6.4.2 Experiments with heterogeneous groups

Uniform conflict resolution styles

In this section, we report the simulation results when the groups are composed of users with diverse preferences but identical conflict resolution style.

Varying the interaction length. Similarly to the experiment with groups having similar preferences, we also observe in this case (Figure 6.8a) that the average loss, in terms of MIL, of cooperative groups is the lowest, while the highest loss is obtained by uncooperative groups. The reason is because members who work together cooperatively are inclined to accept the proposals of each other in a reciprocal way even though they might differ from their personal choice. Conversely, uncooperative users who are mostly focusing on their own concerns and reject alternatives proposed by other members, often end up with a group choice that is neither compatible with the others nor with themselves.

Unlike the previous case wherein group members have similar preferences and the MIL of cooperative users tends to zero very quickly, in this case, the convergence to zero is a bit slower, and evidently, in the 10 observed iterations it never goes to zero. As expected, it takes more time to build acceptable recommendations for groups composed of individuals having diverse preferences, and reaching a consensus requires even more time, with uncooperative members or with larger group size.

Interestingly, as opposed to the homogeneous case wherein there is no difference in MMD irrespective of the user’s conflict resolution style, Figure 6.8b shows that, in this case, the MMD of uncooperative groups, i.e., the difference between the winner (maximum utility) and the loser (minimum utility), is the lowest. The uncooperative members who refuse to accept items proposed by the other group members lose more than cooperative users, but they all lose the same as each other (small MMD). Conversely, cooperative groups, which are composed of users that tend to accept proposals made by other users, even if they do not comply perfectly to their own preferences, tend to collectively lose less utility but there are larger differences in utility between winners and losers (larger MMD).
Fig. 6.8 MIL and MMD of groups with diverse interests but similar conflict resolution styles
The Kruskal-Wallis test confirms that the results obtained from different conflict resolution styles in terms of MIL and MMD are significantly different ($p < 2.2e-16$). With respect to MIL, it can be seen from the results of groups of size 2 shown in Table 6.5 that the conflict resolution styles have significantly different performances, with Bonferroni adjustment $p < 0.05$ (similarly to the other group sizes). Regarding MMD, however, we do not find the significant difference between \textit{competing} and \textit{avoiding} as well as between \textit{accommodating} and \textit{collaborating} in each group size (see Table 6.6 for groups of size 2).

Table 6.5 Pairwise comparisons in terms of MIL for groups of size 2 whose members have diverse interests but similar conflict resolution styles

<table>
<thead>
<tr>
<th></th>
<th>compete</th>
<th>accommodate</th>
<th>avoid</th>
<th>collaborate</th>
</tr>
</thead>
<tbody>
<tr>
<td>accommodate</td>
<td>$&lt;2e-16$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>avoid</td>
<td>0.004</td>
<td>$&lt;2e-16$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>collaborate</td>
<td>$&lt;2e-16$</td>
<td>0.013</td>
<td>$&lt;2e-16$</td>
<td>-</td>
</tr>
<tr>
<td>baseline</td>
<td>$&lt;2e-16$</td>
<td>$&lt;2e-16$</td>
<td>$&lt;2e-16$</td>
<td>$&lt;2e-16$</td>
</tr>
</tbody>
</table>

Table 6.6 Pairwise comparisons in terms of MMD for groups of size 2 whose members have diverse interests but similar conflict resolution styles

<table>
<thead>
<tr>
<th></th>
<th>compete</th>
<th>accommodate</th>
<th>avoid</th>
<th>collaborate</th>
</tr>
</thead>
<tbody>
<tr>
<td>accommodate</td>
<td>$&lt;2e-16$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>avoid</td>
<td>0.31</td>
<td>$&lt;2e-16$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>collaborate</td>
<td>$&lt;2e-16$</td>
<td>0.56</td>
<td>$&lt;2e-16$</td>
<td>-</td>
</tr>
<tr>
<td>baseline</td>
<td>$&lt;2e-16$</td>
<td>$&lt;2e-16$</td>
<td>$&lt;2e-16$</td>
<td>$&lt;2e-16$</td>
</tr>
</tbody>
</table>
Fig. 6.9 Performance of groups with diverse interests but similar conflict resolution styles, stopping the discussion at the first group choice.
and cooperative groups, particularly in large groups. The obtained results explain why in real situations groups of uncooperative people often fail to reach any consensus.

In this case, we have measured Kruskal-Wallis \( p = 0.026 \), \( p = 1.3e-7 \), \( p = 8.5e-6 \) and \( p = 1.1e-11 \) for groups of size 2, 3, 4 and 5, respectively. As the values of \( p \) are all less than the significance level 0.05, we can conclude that there are significant differences between the conflict resolution styles.

With the multiple pairwise comparisons, we find out that only competing and accommodating produce significantly different results in groups of size 2 (\( p = 0.026 \)). In groups of size 3, collaborating has a significantly different MIL than competing (\( p = 0.004 \)), avoiding (\( p = 4e-6 \)) and baseline (\( p = 0.001 \)). As reported in Table 6.7 the significant comparisons found in groups of size 4 comprise accommodating and competing, avoiding and accommodating, collaborating and competing, collaborating and avoiding, and baseline and avoiding. This holds for groups of size 5 as well.

Table 6.7 Pairwise comparisons in terms of MIL for groups of size 4 whose members have diverse interests but similar conflict resolution styles, stopping at the first group choice

<table>
<thead>
<tr>
<th></th>
<th>compete</th>
<th>accommodate</th>
<th>avoid</th>
<th>collaborate</th>
</tr>
</thead>
<tbody>
<tr>
<td>compete</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>avoid</td>
<td>1.00</td>
<td>1.6e-7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>collaborate</td>
<td>0.00</td>
<td>1.00</td>
<td>8.1e-9</td>
<td>-</td>
</tr>
<tr>
<td>baseline</td>
<td>0.12</td>
<td>0.25</td>
<td>0.00</td>
<td>0.09</td>
</tr>
</tbody>
</table>

**Mixed conflict resolution styles**

We now take into consideration groups of users with different preferences and mixed conflict resolution styles.

**Varying the interaction length.** Figure 6.10a shows that the combination of competing and cooperative users (see Mix 1 and Mix 3) leads to the lowest loss in terms of MIL. However, when the competing users are mixed with avoiding (Mix 2) or with baseline members (Mix 4), the average loss becomes the largest and second largest, respectively.

Figure 6.10b visualizes the utility difference between the winner and loser in different combinations. Despite having the largest average loss, the difference between the winner and loser’s utility in mixed groups of competing and avoiding users is the lowest. MIL and MMD in mixed groups of uncooperative users are also moderately stable in larger groups (i.e., groups of size 4 and 5). Moreover, we have observed in the experimental data that
Fig. 6.10 Performance of groups with diverse interests and mixed conflict resolution styles
most of the time the *competing* members are the winners when they are paired with either *accommodating* or *collaborating* users.

With respect to MIL and MMD, the significant differences between the diverse mixtures of conflict resolution styles are confirmed by the Kruskal-Wallis test. We also find that the pairwise differences between mixtures are significant in each group size.

**Stopping the discussion at the first group choice.** Figure 6.11a shows the average number of interaction cycles required for identifying the group choice in each situation. In line with the previous group situations, when *competing* and *avoiding* users are paired with each other, more interaction cycles are needed to find the first group choice. They also suffer the largest loss with respect to MIL, which is illustrated in Figure 6.11b. In groups of 4 and 5 users, the mixed styles groups, overall, arrive at the group choice relatively late, i.e., on average, they need 5 and 10 iterations to find the collective choice. From this observation, we can say that group size does matter for groups consisting of competitive members, hence in realistic settings they will be unlikely to reach a group decision.

Considering a 0.05 level, we find statistically significant differences of MIL in groups of size 3 ($p = 0.03$), size 4 ($p = 0.02$) and size 5 ($1.02 e - 11$), but this difference is not found for groups of size 2 ($p = 0.24$). The mixed groups of *competing* and *accommodating* have MIL that are significantly different from the mixed ones of *competing* and *avoiding* in groups size 3, 4 and 5. Besides, there are statistical differences between the groups where *competing* members are paired with *avoiding* people and the groups where they are respectively mixed with *collaborating* and *baseline* members in groups of size 4 and 5.

### 6.5 Analysis of Real Groups

The goal of this analysis is to show that the previously illustrated simulations of users’ behavior in group discussions and the outcome of these simulations, measured by *Mean Individual Loss* (MIL) can also be partially observed in real group discussions. The following analysis somewhat supports the validity of the proposed simulation approach, and illustrates the main advantage of a simulation procedure, i.e., that group behavior can be explored in more detail without the restrictions imposed by the uncontrolled variables of real groups.

#### 6.5.1 Methodology

The analyzed data of real groups and their discussions was collected from the observation of group decision making processes, which has been described in Chapter 3. Let us recall
Fig. 6.11 Performance of groups with diverse interests and mixed conflict resolution styles, stopping the discussion at the first group choice.
that in this observational study, groups of users were introduced to the task of deciding on a destination to visit, jointly as a group. The study was implemented in three phases, in which groups were observed before, during and after their discussion. In this analysis, we specifically employed the data obtained from the analysis of group activities that were audio recorded and observed based on the template of Bales’s IPA [7]. We note that the recordings of the group interactions were part of the implementation at TU Wien only, where each group contained two members who observed their own group discussions, referred as observers, and the others who participated in group discussions, as decision-makers. The data collection process resulted in 27 participants organized in 8 groups of two (2), three (1), and four (5) decision-makers.

The first step in the analysis was to define the MIL variable, and to assign each group member to an appropriate conflict resolution style. We have to note that the user model captured by the observational study differs from the one used in the simulation analysis, since the preferences were expressed as ranked lists. Therefore, to measure MIL, we had to adopt a different approach, but we tried to identify a metric that captures essentially the same signal as that is measured by MIL in the simulations. First, for each group member we calculated the Spearman Footrule distance between their preferences and the group choice:

$$\text{DIST}(u,g) = \text{rank}_u(\text{group\_choice}(g)) - 1.$$  

More precisely, the distance is calculated as the difference between the individual and group rank of the group choice (i.e., group rank is clearly 1, as it is the preferred group option). Then, MIL is simply calculated as the mean value of the previously defined individual group members’ distances:

$$\text{MIL}(g) = \frac{1}{|g|} \sum_{u \in g} \text{DIST}(u,g).$$

To assign group members, and thereafter groups, to their “appropriate” TKI conflict resolution styles, we used the observational data, collected during the group discussions. Here, it was crucial to have a clearly defined procedure as in the simulation analysis. To this end, only five of the total twelve behavioral categories were used, i.e., Friendly, Agree, Disagree, Opinion give, and Information give. The four conflict resolution styles, plus the baseline category, were defined upon two dimensions, assertiveness and cooperativeness. Assertive participants were defined as those with a high number of repetitions recorded for the Opinion give or Information give categories, while unassertive participants were simply those who were not identified as assertive. Cooperative participants were defined as those with a high number of repetitions recorded for the Friendly or Agree categories, and uncooperative participants as those with a high number of repetitions recorded for the Disagree category.
The number of recorded repetitions was considered to be high when the individual score was higher than the mean number of times the action was observed over the whole dataset.

Finally, a group was assigned to a specific conflict resolution style when the majority of its group members belong to that particular conflict resolution style. At the end, this resulted in eight groups of users with the same equal conflict resolution style: five belonged to baseline and three to the collaborative style. Table 6.8 shows repetitions of selected behavioral actions at the group level, together with the assigned TKI conflict resolution style.

<table>
<thead>
<tr>
<th>Group</th>
<th>Friendly</th>
<th>Agree</th>
<th>Disagree</th>
<th>Info give</th>
<th>Opinion give</th>
<th>TK style</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0</td>
<td>3.0</td>
<td>2.5</td>
<td>2.0</td>
<td>4.0</td>
<td>baseline</td>
</tr>
<tr>
<td>2</td>
<td>8.0</td>
<td>1.5</td>
<td>10.5</td>
<td>3.5</td>
<td>6.5</td>
<td>baseline</td>
</tr>
<tr>
<td>3</td>
<td>19.5</td>
<td>14.5</td>
<td>9.5</td>
<td>27.5</td>
<td>20.5</td>
<td>baseline</td>
</tr>
<tr>
<td>4</td>
<td>7.5</td>
<td>0.5</td>
<td>1.0</td>
<td>14.5</td>
<td>28.5</td>
<td>baseline</td>
</tr>
<tr>
<td>5</td>
<td>11.0</td>
<td>5.5</td>
<td>3.0</td>
<td>10.0</td>
<td>13.0</td>
<td>baseline</td>
</tr>
<tr>
<td>6</td>
<td>7.0</td>
<td>1.0</td>
<td>4.0</td>
<td>14.0</td>
<td>18.0</td>
<td>collaborative</td>
</tr>
<tr>
<td>7</td>
<td>9.0</td>
<td>1.0</td>
<td>5.0</td>
<td>10.5</td>
<td>8.5</td>
<td>collaborative</td>
</tr>
<tr>
<td>8</td>
<td>36.0</td>
<td>1.0</td>
<td>0.0</td>
<td>21.5</td>
<td>33.0</td>
<td>collaborative</td>
</tr>
</tbody>
</table>

### 6.5.2 Results

The results obtained by this analysis should be compared with the simulation results obtained from the simulations of group behavior with diverse preferences and discussions that stopped with the first group choice. The reason behind this is that the diversity of group members’ preferences was not controlled during the group formation phase. Moreover, the groups did not receive any instructions about their discussion, therefore we can conclude that the process was finished as soon as the group found an acceptable group choice.

Hereby, we focus on the order of MIL for the two conflict resolution style categories that we obtained in the collected dataset (i.e., all together seven groups, five belong to baseline and two to collaborative categories). Figure 6.12 illustrates MIL scores for the two conflict resolution style categories, having $MIL(baseline) > MIL(collaborative)$. Hence, even if, we did not manage to capture group behavior across all five conflict resolution categories, the two categories that we do have follow the same pattern as in the simulation analysis.

---

1We used a different rule for assigning the group’s conflict resolution style in the simulated and real scenarios because of the lack of observed data in the condition where group members’ resolution styles are all equal.
In conclusion, the results obtained by the analysis of real groups tend to support some of the findings derived from the simulation experiment. However, this evidently indicates all the difficulties, issues, and limitations that emerge in the analysis of real groups. First, the data collection procedure is complex and time consuming. Second, finding participants willing to complete the three phases, and observers capable of recording group behavior, is challenging. Finally, even when overcoming the first two issues, it is difficult to collect data that contains various conditions for conducting a thorough analysis. Having said that, however, even with all its limitations, analysis of real groups is crucial to support the validity of the results obtained by simulations.

6.6 Discussion and Conclusions

In this chapter, we have introduced a group discussion simulation model that can be used to analyze the effect of different group compositions. In the model the users’ behavior is determined by the user utility function and conflict resolution style. To assure that the model can predict realistic outcomes of the simulated discussions, it is derived from the real observations of how users use our interactive GRS. The analysis of a range of simulated group discussions has shed light on how the outcome of the group decision making process supported by the GRS is influenced by the conflict resolution styles interlinked with the inner-group similarity, interaction length and group size.

Based on the simulation results, we can derive that, when group members have similar tastes, no matter what the conflict resolution style is, there is no difference in their utility for the collective choice. However, groups whose members have competing or avoiding conflict resolution style obtain a lower utility compared to those who adopt accommodating and collaborating styles. On the other hand, when the members have diverse preferences, we
have found an interesting tension between the average individual’s utility loss and the fairness of the system-generated group recommendations. In particular, groups of competing or avoiding users often choose a recommendation that gives to all the group members a similar loss in utility (compared to their individual best choice), but it also makes them suffer the greatest loss on average. We have observed the opposite trend in groups of accommodating and collaborating users. When it comes to groups of mixed conflict resolution styles, we noticed that when a group is formed by a combination of competing and either accommodating or collaborating users, the average individual’s loss is the smallest even though the discrepancy in their utility is the largest.

In addition, the analysis conducted on data collected from the observational study of real group discussions bolsters some of the findings of the proposed simulation model. The results of real groups illustrate that the average loss, measured in terms of MIL, of groups composed of collaborating styles is lower than that of groups with baseline style, which matches the result observed in the simulation experiment.

Finally, the evidence gained from the simulation suggests that the similarity between group members, the mean individuals’ utility loss (MIL), the difference between maximum and minimum utility (MMD), and the number of proposals can be leveraged as “circumstantial features” to estimate a group setting in terms of conflict resolution styles at a group level. Apparently, at an individual level, one might conjecture the conflict resolution style of each user by simply observing the frequency of the proposal and their reaction to the proposals made by the others, and then double check if the observed behavior matches the prescribed model. However, at the end of the day, in order to conclude about the users’ conflict resolution styles further analysis is necessary, and it still is essential to determine the
group situation (e.g., the majority rule can be applied). This task is particularly challenging when the estimated conflict resolutions of group members are not uniform.

Therefore, the findings of this simulation further support the need for estimating the group situation. For instance, the schema illustrated in Figure 6.13 offers a tentative set of if-then-else decision rules with the “circumstantial features” represented as nodes and branches corresponding to the value of each feature leading to the estimated group situation. Given an observed group and its members’ actions (e.g., users’ ratings, how they react to the others’ options) and assuming that we precisely know the individual user utility, the group recommendation model can first return the MIL and MMD for the group. Then, these values can be put into the prescribed schema in Figure 6.13 to predict the group situation: these metrics are considered to be “high” or “increased” if they are greater than those of the baseline, and to be “low” or “decreased” otherwise. The idea of using “decision tree model” is inspired by the work of Vroom and Jago, which recommends appropriate decision schemes for group leaders [95].

We leave the task of materializing the “group situation assessment” concept to future work, where we aim to develop a situation-adaptive GRS that adapts group member’s utility models to the specific group situation estimated from tracing users’ interactions, so that the generated recommendations are more effective and meaningful. Further implications of advancing simulations for GRSs will be discussed in the next chapter.
CHAPTER 7

Conclusions and Future Work

In this thesis, we have explored how an interactive group recommendation approach can be leveraged to support a group decision making process. We set the focus on three aspects, in particular, facilitating the group discussion with the recommendation functionality, modeling the dynamics of users’ preferences while they are evolving in the group setting, and studying the impact of different group compositions and group situations on the quality of the recommendations. In this last chapter, we summarize the main findings, reflect on the work carried out in this thesis and give an outline of future research directions.

Summary

Starting with the objective of laying the groundwork for designing an effective GRS that can support the group decision making process, we have conducted an observational analysis of how groups of users, in reality, interact and make a group decision in a travel and tourism scenario. The results demonstrate that group preferences are built over a series of group interactions, and individuals’ characteristics stand out as the influential factor to interpret their satisfaction with the final group choice (Chapter 3).

Grounded on the findings of the observation, we have developed a chat-based GRS called STSGroup, in which the group recommendation process is seamlessly integrated into chat conversations to support the group communications and decision making step. To generate the group recommendations, we have proposed an interactive group preference model that not only exploits historical knowledge about users’ behavior when they act as an individual (long-term preferences), but also takes into account the dynamic of users’ interactions as a member of the group discussion (session-based preferences). We have performed an exploratory user study to assess the system usability along with the perceived recommendation quality and got promising results (Chapter 4).
In an attempt to evaluate the effect of different combinations of long-term and session-based preferences on the performance of the proposed interactive recommendation model, we have designed a procedure that simulates users’ behavior in different group settings motivated by three scenarios of conformity in group dynamics. The experimental results show that the relative importance of the long-term and session-based preferences varies according to the specific group scenarios, and provide evidence for the efficacy of the proposed model in capturing the changes in the user’s needs in a group decision making process (Chapter 5).

Finally, to understand how individuals’ behavior in conflict situations affects the group decision making output supported by the interactive group recommendation process, we have enhanced the prior group discussion simulation model to simulate users’ responses as a function of their preferences and their conflict resolution styles. We have carried out extensive experiments to measure the dynamics of the group recommendation performance in various group situations characterized by alternative conflict resolution styles. The obtained results provide important and novel insights into the effect of the user’s conflict resolution style on the outcome of a group decision making process supported by a GRS (Chapter 6).

**Perspectives**

Designing a conversational GRS still has so much room for improvements, and requires multiple components to be integrated. Therefore, a number of research avenues that have not yet been explored in the course of this work may attract diverse lines of follow-up research. Apart from the implications of GRS research previously discussed in Section 3.4, in the following, we restrict ourselves to highlighting a few promising directions that can advance and deepen some research topics initially addressed in this thesis.

**User-System Interaction Design**

The developed system introduced in Chapter 4, so far, has operated in a passive mode as it only provides recommendations and choice support when group members request it. Thus, one of the future strands of research is to augment the system with a proactive component that is responsible for actively pushing recommendations and preference information requests to the members when the group setting seems appropriate. Developing proactive functionality for GRSs is unequivocally an under-explored avenue due to the major issue of determining when the system should take an action. The answer to this question clearly depends on many factors that are related to the individuals as well as to the group decision making process. For instance, in situations where competition is likely to bring conflict, the system should intervene to help the group to avoid jeopardizing the decision making quality.
In that respect, it is crucial to assess what is likely to happen in the group setting by unobtrusively tracking users’ interactions. As already mentioned in Section 3.4, understanding the group situation will lay the foundation for developing such a proactive system.

Regarding the user interface design, recent research work on computer-supported collaboration has pointed out that the integration of sense-making techniques such as tagging and summarizing chat messages can address the problem of chat representation [98]. In particular, often important and unimportant messages are intertwined within the conversation threads, so this burdens users with sifting through the chat logs to find needed information. Following this line of thought, we can possibly enrich the representation of chat discussions by exploiting markup added by users, such as emoji reactions to items or item-tag notes left by users to re-structure the original chat stream into a readable and accessible summary view within the chat interface.

User Modeling and Group Modeling

The proposed group recommendation model, so far, is based on the assumption that all group members are equally important, therefore, as a default and starting point, it is identified by an average preference aggregation rule. Nevertheless, it is often the case that some of the members may be more or less influential in the process of making a collective choice than the others, depending on their role the group. In that respect, the proposed group model can be further improved by automatically determining the importance weight of each member. Arguably, if we view each individual’s profile as a classifier (e.g., best choice, likes, or dislikes) then one possible solution to aggregate a set of classifiers is ensemble-based methods which are commonly considered for hybridization techniques in the RSs literature, especially in context-aware recommendations, i.e., to combine several contextual models into one model at the same time [2, 3]. Likewise, these ensemble techniques can be applied to automatically learn the appropriate weights for the long-term and session-based preferences in a particular group setting, which still remains an open question in this work.

Furthermore, in this thesis, we mainly focus on a user model represented in terms of a utility function, which means the quality of the model heavily relies on the representation of item features. As mentioned in Chapter 4, our current recommendation model still uses a relatively high number of item features - close to a hundred features. This might account for the fact that the process of solving the optimization problem under inferred constraints is rather slow. One way to overcome this issue is by implementing feature selection techniques at a pre-processing step, which can help to reduce the number of ineffectively used features [45], so that we can obtain a more effective and simpler user model. Alternatively, it is also worth exploring other models that do not require any knowledge of item features. For
example, another way that we can possibly leverage is to allow group members to articulate their preferences in the form of pairwise comparisons of alternative options, such as, item \( i \) is preferred to item \( j \), and then, apply pairwise preference learning techniques [16, 48] to generate group recommendations.

Up until now, the user profile is constructed on the basis of users’ feedback expressed in an explicit way (e.g., through giving ratings, and marking best choice, likes, or dislikes). That means, the preference information revealed in the form of textual chat logs has not been analyzed and considered. Thus, in the future one could adopt Natural Language Processing techniques to extract users’ opinions about discussed items within chat messages, and exploit a larger number of preference statements for enriching the current user model.

**Simulation Analysis**

It is safe to say that in order to unravel the key question of how a GRS can effectively facilitate a group decision making process, one could face additional questions, such as, when the system is supposed to take action, what type of support it should offer, or how we can measure its effectiveness. In Chapter 5 and 6 of this thesis, simulation models were proven to be useful for evaluating the group recommendation performance in various artificial group scenarios. We, therefore, are convinced that a simulation analysis is a starting point that enables us to approach the problem of supporting the group decision making process from several perspectives. For instance, it might be useful to utilize simulations to build what-if scenario analysis for a group decision support system. In particular, the analysis will empower users to try out different personal choices or item evaluations, and then, analyze and visualize the impact of those changes on the final group choice with respect to metrics that the users care about, e.g., individual utility loss or fairness. With this sort of human in the loop simulation, the system can possibly be better fine-tuned while the users might grasp how the system exhibits and, ultimately, make more informed decisions.

In the simulation design presented in Chapter 6, we have grounded the simulator by using real observations, but clearly, the simulated users’ behavior that is obtained from a set of initial parameters cannot perfectly replicate the real dynamics. In fact, overcoming the discrepancy between simulated models and the real-world environment has drawn wide interest in the field of learning robotic manipulation policies. Following the line of work on closing the reality gap, one of the approaches, which somehow can further develop, is to iteratively learn parameters of a simulation model, so that simulated information produced by this trained model is closer to the observations of the real world [24].


