TourExplain: A Crowdsourcing Pipeline for Generating Explanations for Groups of Tourists

Öykü Kapcak, Simone Spagnoli, Vincent Robbemond, Soumitri Vadali, Shabnam Najafian, Nava Tintarev
Delft University of Technology, Delft, the Netherlands
{kapcak,s.spagnoli,v.t.robbemond,s.vadali}@student.tudelft.nl
{s.najafian,n.tintarev}@tudelft.nl

ABSTRACT

When a group is traveling together it is challenging to recommend an itinerary consisting of several points of interest (POIs). The preferences of individual group members often diverge, but it is important to keep everyone in the group satisfied during the entire trip. We propose a method to consider the preferences of all the people in the group. Building on this method, we design explanations for groups of people, to help them reach a consensus for places to visit. However, one open question is how to best formulate explanations for such sequences. In this paper, we introduce *TourExplain*, an automated crowdsourcing pipeline to generate and evaluate explanations for groups with the aim of improving our initial proposed explanations by relying on the wisdom of crowds.

CCS CONCEPTS

• Information systems → Crowdsourcing; **Recommender systems**; **• Human-centered computing → Natural language interfaces**; **Empirical studies in HCI**;

KEYWORDS

Explanations; Crowdsourcing; Crowdworking; Group recommendation; Tourism; Sequences

1 INTRODUCTION

Recommender systems are decision support systems which help users to find one or more items in a large space of possible options that best fit their wishes and needs. The main focus of current recommender systems is to propose items to individual users. However, in tourism people often consume several items, and often do so in groups rather than individually.

A group traveling together can be recommended an itinerary consisting of several points of interest (POIs). However, reaching a consensus is difficult, and often compromises need to be made. Such compromises can potentially help users expand their tastes. Mary's preferred POI may become John's new favorite spot! Compromises can also lead to rejection of the recommended items. One way to avoid this is to explain recommendations that are surprising; or even expected to be disliked; by an individual user [12]. In addition, there are many ways to formulate explanations for groups, but few guidelines for generating such explanations. To address these challenges, we present a novel crowdsourcing pipeline for generating and evaluating group explanations.

2 RELATED WORK

This work builds on two strands of research, namely 1) explanations for group recommendations and 2) crowdsourcing for improving the explanation text.

2.1 Explanations

A group traveling together can be recommended an itinerary consisting of several points of interest (POIs). To keep the group satisfied during the entire sequence of recommendations (e.g., POIs), we need to consider the preferences of all the people in the group [5]. This can be challenging when the preferences of individual group members diverge. An explanation in such contexts can assist users reach a consensus for places to visit.

Ardissono et al. [1] developed a handheld recommender system for sightseeing destinations and itineraries for heterogeneous tourist groups. This system supplied explanations based on the properties of items but did not consider the need to support consensus. Moreover, Nguyen and Ricci also combined user preferences generated by the interactions between group members. Although they studied group decision making and consensus, they have not studied explanations [13].

Masthoff et al. [10] suggest several *preference aggregation strategies*. These have as input a set of predictions for all users in a group for a set of items, and have as output a sequence of recommended items. In our previous work, we built on this work and designed explanations for groups of people that helped them reach a consensus [12]. One open question is how to best formulate explanations for such sequences. In this work, we therefore aim to improve our initial proposed explanations by relying on human wisdom using crowdsourcing.

2.2 Crowdsourcing

Crowdsourcing is a practice for solving computationally hard tasks by assigning them to an undefined (and generally large) network of people in the form of an open call, usually through online platforms (Mechanical Turk[1], FigureEight[2], etc.). This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals (crowdworkers) [7]. Crowdsourcing approaches are used for creating content or generating ideas with the contribution of a crowd. The approach proposed in this paper is to use the wisdom of crowds to generate and improve explanation text for end-users. This idea is similar to previous work

* The first to fourth authors contributed equally to this work.

[1] https://www.mturk.com/, retrieved July 2018
[2] https://www.figure-eight.com/, retrieved July 2018

which used crowdsourcing to find better formulations for numerical expressions [2]. This previous work used templates to collect simple sentences (perspectives) from workers to make numerical expressions easier to understand. Finally, they evaluated the effectiveness of these perspectives on everyday readers' numerical comprehension.

Similarly, other authors proposed a model to generate personalized natural language explanations in the movie domain [4]. The crowdworkers were provided by quotes extracted from online movie reviews and the user rating history. Compared to our work, these explanations were designed for the movie domain and for individual users rather than group recommendations. Another difference in the design pattern: we specify specific criteria (based on Gricean Maxims [6]) in our all three steps: Find, Fix and Verify steps. In the finding step to give crowdworkers clear guidelines for finding any shortcomings in terms of these criteria; in the fixing step to give them clear guidelines for improving the explanations; in the verification step for validating the explanation.

Bernstein et al. [3] also applied crowd-sourced contributions to help humans write and edit their work. *Soylent* is a language processing interface that uses people to help authors to shorten, proofread, and edit documents.

This paper builds on the Find-Fix-Verify design pattern used in Soylent [3], where a different group of crowdworkers 1) Find errors in a given text (*Find*), 2) Fix them by editing (*Fix*), and finally 3) Verify the modifications (*Verify*).

3 USER INTERACTION

A group of people can use the *TourExplain* system when going on a trip. The group creates a new "Trip" in the system and enter trip parameters (i.e., POIs to be considered, number of participants, and whether the explanations need to be anonymous or not). Following the creation of the trip in the system, each member of the group has to enter their own preferences for each POI (in a private environment). After all of the preferences have been submitted, the system generates an itinerary, or a sequence of POIs, for the group, as well as explanations. Each explanation is then posted to the crowdsourcing system to be improved as described in Section 4.2 "Subsystem 2: Crowdsourcing". Once the crowdsourcing part of the system completes, each user will be received the recommended itinerary and its corresponding explanations.

4 SYSTEM DESIGN

Figure 1 outlines the workflow for the *TourExplain* system. It consists of two subsystems that communicate via an API: (1) *explanation* generation, and (2) *crowdsourcing* to improve the generated explanations. The implementation of our system supports the use of both subsystems, as well as the use of each individual module separately. Besides, this architecture allows us to easily add, exchange, or remove modules in our system.

4.1 Subsystem 1: Explanation

This subsystem consists of two parts: 1) *generate sequences*, and 2) *generate explanations.*

Generate sequences. Here, the system generates a sequence of POIs for the group to visit, according to previously proposed preference aggregation algorithms [12]. A preference aggregation strategy dictates how to combine individual preferences to recommend a sequence. This dictates both whether an item is included, as well as its position in the itinerary. The latter is important to consider since it has previously been found that overall satisfaction with a sequence depends on the order of the items in the sequence [11].

Following are the two above mentioned algorithms that we used to generate itineraries. Readers who wish to get a coherent overview of the proposed algorithms is referred to our previous work [12]:

- A 1: **Least Misery + Most Pleasure + Without Misery**. The plus signs imply chaining three strategies, applying one after the other.
- A 2: **Fairness -> Average**. The arrow implies applying a tie-breaking strategy, i.e., when several items receive an equal score using only Fairness.

Generate explanations. Pure crowdsourcing approaches to explain the preference aggregation strategies used to generate the sequence of recommended items will not succeed because most crowdworkers are not domain or recommendation experts. Even if they are informed about applied algorithms we cannot expect a crowdworker to write an appropriate explanation for the recommended items. Therefore, we provide them with initial explanations in the beginning which they can *improve* based on specific criteria.

Using a template-based natural language generation approach, the system generates explanations for each user according to their preferences in the recommended sequence. These (personal) explanations are based on predefined templates, examples of templates are *"Hello X"*, *"we know you would love to see Y"*, and *"however, others in your group would love to see Z"*.

For example, we consider a user *John* who has expressed a liking for seeing the Eiffel Tower because John and a couple of friends are visiting Paris soon. However, John's friends have expressed they preferred seeing the Louvre over the Eiffel Tower. This could lead to a template based sentence: *"Hello John, we know you would love to see the Eiffel Tower, however, others in your group would love to see the Louvre first."*

The system is provided with a number of templates to handle a number of predefined situations considered by the explanation generating algorithm. These automatically generated explanations are then sent to the second part of the system via an API to be reviewed by crowdworkers.

An example scenario for when an explanation may be needed is when a POI that is highly rated by person A is not chosen in the sequence of recommended POIs. The explanation for this person can be: *"Even though you wanted to visit POI X, most of your friends gave a very low rating for that POI. Therefore, we did not include that into the recommended POIs for the group."*

4.2 Subsystem 2: Crowdsourcing

The aim of the *crowdsourcing subsystem* is to improve the aforementioned generated explanations by using the wisdom of crowds.

We employ the Find-Fix-Verify pattern as described by [3] to detect and eradicate errors in the explanations. This approach not only flags up errors in explanations but also improves the explanations.

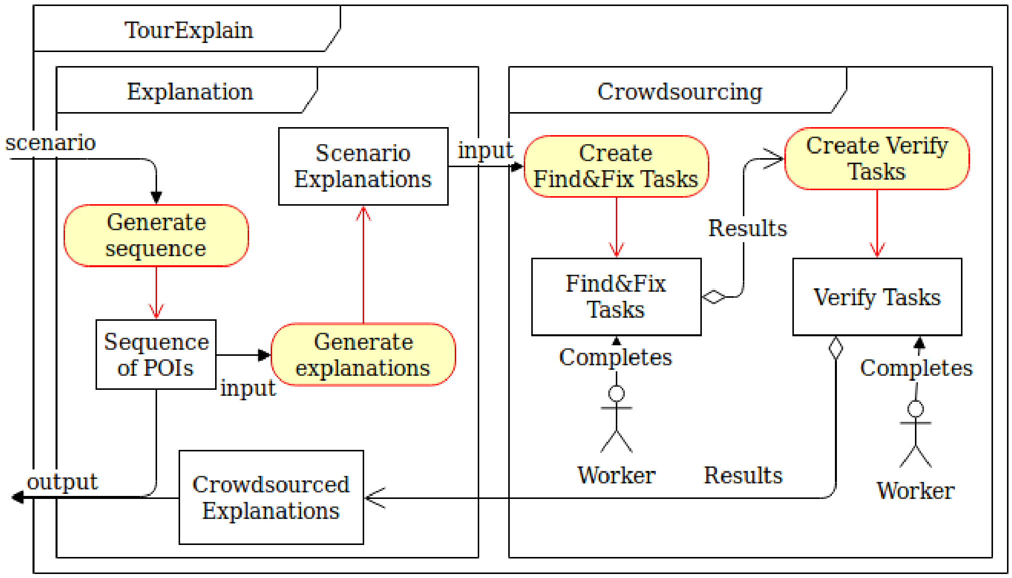

Figure 1: The system consists of two subsystems: 1) Explanation generation; 2) Crowdsourcing to improve the explanations.

For our purpose, we have adapted this approach and combined the Find- and Fix steps. This improves the accuracy of suggestions and is less time-consuming as well. Unlike the case of Soylent [3], the text to improve is short and can be modified efficiently. The same worker can directly suggest an improvement when they find an explanation inappropriate, as opposed to simply passing on that information to another worker, who then has to find an improvement.

Guidelines for Find-Fix-Verify. We give our workers three main criteria (based on Gricean Maxims [6]) to look for in the tasks:

- **Quantity:** Is the explanation informative? Does it provide all the information necessary and no more?
- **Quality:** Is the explanation truthful? Does it provide no information which is false?
- **Relevance:** Is the explanation relevant to the given scenario? It should not mention any irrelevant information.

The crowdsourcing pipeline contains two tasks:

Find & Fix tasks. A crowd-worker (worker henceforth) is given an explanation and asked to find any shortcomings in terms of the criteria mentioned above. After that, the worker is asked to make a suggestion to improve (fix) the sentence.

Verify task. A worker is given an explanation that is fixed by another worker in the find-fix step to evaluate in terms of the criteria mentioned above. The worker is asked to verify each criterion on a binary scale, giving their approval or disapproval for the particular metric. When the majority of the workers approve at least *two* criteria for a given explanation, it is considered a satisfactory explanation. Based on the number of approval/disapproval ratings, these satisfactory explanations are ranked from best to worst.

A vital part of our system is that the workers who do the *Find-Fix* versus *Verify* steps are independent of each other. This ensures there is no bias in picking a particular explanation. The tasks are created and launched using the Figure Eight API [3].

Unlike the previous generation of explanations, the crowdsourcing part cannot be done in real-time, but it requires some more time to be done. This is due to the fact that is not possible to know when the tasks will be performed by the workers. This time is subject to multiple factors as the monetary reward for each task, or the number of workers that perform the same task. In fact even though is possible to estimate the time to perform a given task by a worker, it becomes complex to estimate when a launched task will be picked up by a worker, also this time is directly related with the monetary reward for the task.

To ensure data quality, we only select workers that are native English speakers. When it was possible we randomized the order of questions and answers to avoid possible bias. Furthermore, to limit the introduction of error by the workers we performed each step by multiple workers. The number of workers that perform the same step can be dynamically chosen.

5 NEXT RESEARCH STEPS

We plan to use this pipeline as the basis of doctoral work investigating how to best generate explanations for itineraries (sequences of POIs) for groups of users. For this purpose, as suggested by Kim et al. [9], we are going to let crowdworkers form groups and collaborate to accomplish determined tasks.

In the following sections we describe future research avenues that will be pursued in this project. We introduce the notion of group dynamics, which consider the relationship between people within a group. We also consider the influence of interaction design on the requirements for explanations.

[3] https://www.figure-eight.com/, retrieved June 2018

5.1 Group Dynamics

Existing group recommendation techniques usually focus on merely aggregating individual preferences and thus do not take into account social interactions and relationships among the group members. Previous work has found that it is not the case, i.e., group members are influenced in their evaluations by the combination of the group and the interaction between and social relationships among group members [5].

In order to personalized the preference aggregation algorithms and their corresponding explanations as well as make our recommendations group-aware, we plan to use the Thomas-Kilmann Conflict Style Model (TKI model) [8] as a personality model. The advantage of this model is that it focuses on the interaction between group members rather than the characteristics of individual users, as in the Big Five factor model [5].

Another important group aspect that we aim to consider is the types of relationships within groups (c.f., [11]).

- Communal Sharing: Somebody you share everything with
- Authority Ranking: Somebody you respect highly
- Equality Matching: Somebody you are on equal footing with
- Market Pricing: Somebody you do deals with / compete with

We can employ the group types in both aggregating preferences algorithms as well as designing explanations. For instance you might feel comfortable to reveal your preferences to somebody you are on equal footing with (such as your friends) but not with somebody you respect (such as your boss).

5.2 Interaction Design

Individual vs Group Explanations. In this work we tried to improve automatically generated explanations by using the wisdom of crowds for a *single user*. However, we did not evaluate the final result with real groups of users. In our next steps, we will evaluate these explanations by presenting them to groups and compare the results with individual personalized explanations (for each group member). One can expect to find a trade-off between explanations that are suitable for the whole group, compared to personal explanations for each group member. For example, one benefit of group explanation is that we can present it on a common device viewed by the whole group. On the other hand, personal explanation can supply individual users with more personalized information about why that item is recommended to them.

Transparency vs Privacy Preserving. The requirements on explanations are also likely to be influenced by group versus individual preferences. For instance, there is a trade-off between having a high transparency while not violating the users' privacy. So users might demand to conceal their preferences for other group members or they feel comfortable to reveal their preference depends on different types of groups or their personalities.

Single Item vs Sequence Explanations. In this work we provided an explanation for each *single item*. However it might not be convenient always depending on several things e.g., domain. For example, in the cinema domain, users would want a recommendation for a specific movie instead of a sequence whereas in the tourism domain, a sequence of POIs would be more appropriate. In our future work, we will design explanations for each POI, and compare them with an explanation for the whole itinerary.

6 CONCLUSION

In this paper we introduce an automated crowdsourcing pipeline to generate and evaluate explanations for groups. The proposed solution is suitable for domains where items are a) consumed in groups, and b) in a sequence. This particularly useful for the recommendation of itineraries in tourism.

Additionally, it is likely that the approach is extendable to other domains, however there is a constraint for domains which require immediate and real-time explanations. For tourism, where trips can be planned in advance of a visit, this limitation may be less severe.

While simple, the proposed approach can be extended to answer different research questions. In this position paper we highlighted two significant and planned extensions:

- **Group dynamics**. How can explanations be improved by taking in account group dynamics such as conflict style or relationships within groups?
- **Interaction Design**. How should we adapt the explanations to the way they are consumed, e.g., for an individual item or for a sequence? Or for a single user versus for the group?

REFERENCES

[1] Liliana Ardissono, Anna Goy, Giovanna Petrone, Marino Segnan, and Pietro Torasso. 2003. Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Applied artificial intelligence* 17, 8-9 (2003), 687–714.

[2] Pablo J. Barrio, Daniel G. Goldstein, and Jake M. Hofman. 2016. Improving Comprehension of Numbers in the News. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2729–2739. https://doi.org/10.1145/2858036.2858510

[3] Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. 2010. Soylent: A Word Processor with a Crowd Inside. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 313–322. https://doi.org/10.1145/1866029.1866078

[4] Shuo Chang, F Maxwell Harper, and Loren Gilbert Terveen. 2016. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 175–182.

[5] Alexander Felfernig, Ludovico Boratto, Martin Stettinger, and Marko Tkalčič. 2018. *Group Recommender Systems: An Introduction*. Springer.

[6] H Paul Grice. 1975. Logic and conversation. *1975* (1975), 41–58.

[7] Jeff Howe. 2006. Crowdsourcing: A definition. (2006).

[8] Ralph H Kilmann and Kenneth W Thomas. 1977. Developing a forced-choice measure of conflict-handling behavior: The "MODE" instrument. *Educational and psychological measurement* 37, 2 (1977), 309–325.

[9] Joy Kim, Sarah Sterman, Allegra Argent Beal Cohen, and Michael S. Bernstein. 2016. Mechanical Novel: Crowdsourcing Complex Work through Reflection and Revision. *CoRR* abs/1611.02682 (2016). arXiv:1611.02682

[10] Judith Masthoff. 2004. Group modeling: Selecting a sequence of television items to suit a group of viewers. In *Personalized digital television*. Springer, 93–141.

[11] Judith Masthoff. 2015. Group recommender systems: aggregation, satisfaction and group attributes. In *Recommender Systems Handbook*. Springer, 743–776.

[12] Shabnam Najafian and Nava Tintarev. 2018. Generating consensus explanations for group recommendations. In *UMAP Latebreaking results*.

[13] Thuy Ngoc Nguyen and Francesco Ricci. 2018. Situation-Dependent Combination of Long-Term and Session-Based Preferences in Group Recommendations: An Experimental Analysis. In *Symposium on Applied Computing (SAC)*. 1366–1373.

