Do You Have any Recommendation? An Annotation System for the Seekers' Strategies in Recommendation Dialogues

Martina Di Bratto[1], **Riccardo Orrico**[1], **Ancuta Budeanu**[2],
Marta Maffia[2] **and Loredana Schettino**[3]

1. University of Naples "Federico II", Italy
2. University of Naples "L'Orientale", Italy
3. University of Salerno, Italy

martina.dibratto@unina.it, mmaffia@unior.it, lschettino@unisa.it

Abstract

The development of dialogue systems benefits from the study of the communication strategies used by human speakers. In the context of recommendation dialogue systems some researchers have investigated the sociable recommendation strategies employed by the Recommenders in natural settings to make successful and persuasive recommendations (Hayati et al., 2020, INSPIRED corpus). However, the Seeker's contribution, as well as the Recommender's, shapes the development of the communicative exchange, in that the Seekers may use specific strategies to disclose their preferences and reach their goal. So, modelling the Seeker's communicative strategies along with the ones used by the Recommender may improve the efficiency of recommendation dialogue systems. In this work, we provide a reliable tagset for the Seekers utterances present in the Inspired dataset, defining a set of communicative strategies coherent with the already existing one for the Recommenders.

1 Introduction

Nowadays conversational recommendation systems seem to be acquiring a fundamental role in information seeking and retrieval. In a recent paper, Hayati and her colleagues (Hayati et al., 2020) have argued for the need to study the communication strategies used by human speakers in a natural setting for developing dialogue systems that are able to make successful and persuasive recommendations. The authors have proposed Inspired,

a dataset of recommendation dialogues collected in a realistic setting, enriched with a detailed annotation of the sociable recommendation strategies employed by the Recommender.

However, as in any interaction, these dialogues are the result of the cooperation between the interlocutors, who actively partake in both the construction of meaning and of the relationship among each other (Bazzanella, 2005): the Seeker's contribution, as well as the Recommender's, shapes the development of the communicative exchange, in that the seekers may use specific strategies to disclose their preferences and reach their goal, i.e., to get items that suit their needs. Hence, modelling the Seeker's communicative strategies along with the ones used by the Recommender may improve the efficiency of recommendation dialogue systems.

In this work, we aim to fill this gap proposing a tagset for the Seekers communicative strategies that is coherent with the one previously provided for the Recommenders by Hayati and colleagues. The paper is structured as follows: recommendation dialogue systems are considered in relation to the Argumentation Theory (§ 2) and the Inspired tagset (Hayati et al., 2020) is described (§ 2.1), then the tagset for the Seeker's strategies is presented (§ 3), along with the data proving the reliability of the annotation scheme (§ 3.1) and a preliminary analysis of the interactions (§ 4).[1]

2 Recommendation Dialogue

Recommendation dialogues are characterized by two or more participants who disclose their preference and make recommendation in order to select a certain item that should satisfy the re-

[1]The present study is the result of a collaborative work of all the authors. Paragraphs 2 and 2.1 have been written by Martina Di Bratto, paragraph 3 by Marta Maffia and Ancuta Budeanu, 3.1 and 4 by Riccardo Orrico and, finally, sections 1 and 5 by Loredana Schettino.

quirements retrieved during the communicative exchange. Conversational Recommendation Systems (CoRS), in the same way, aim at finding or recommending the most relevant information (e.g., web pages, answers, movies, products) for users based on textual- or spoken-dialogues, through which users can communicate with the system more efficiently using natural language conversations (Fu et al., 2020). CoRS, thus, can be seen as persuasive social actors since a recommendation can be considered persuasive when it attempts to change people's mind or behavior by employing various persuasive strategies (Shi et al., 2020). A conversation where two or more interlocutors (humans or not) aim to resolve a conflict of opinion, can be considered as a form of persuasion dialogue leveraging on argumentation (i.e., the process of exchanging ideas in order to establish the truth of a statement). CoRSs can be framed in the field of formal argumentation and more specifically, refer to the argumentation-based dialogue. It considers the problems arising from dialogues involving different agents and whose information are shared and distributed among them. This interaction introduces multiple, not necessarily aligned knowledge and, possibly, conflicting goals in the pursuit of a solution to a problem. (Di Maro, 2021). Walton's classification of dialogues (Walton, 1984) is often employed in the study of the argumentation-based dialogue. He distinguished six different categories of dialogue: *persuasion*, *negotiation*, *information seeking*, *deliberation*, *inquiry*, and *quarrel*. The purpose of persuasion dialogues, thus, can be seen as 'pure' argumentation and can be often embedded in other dialogue types (Prakken, 2018). The Recommendation task, indeed, tends to present a pattern structured in two phases, Exploration and Exploitation (E&E), which can be intended as two types of dialogues embedded into each other. According to (Gao et al., 2021, p. 15), with exploration "[. . .] the system takes some risks to collect information about unknown options". On the other hand, during the exploitation phase, "[. . .] the system takes advantage of the best option that is known". Hence, the exploration phase can be associated to the *inquiry* dialogue since the main aim is to achieve the "growth of knowledge and agreement" starting from an initial situation of "general ignorance" (Walton and Krabbe, 1995, p. 66). The exploitation phase, on the other hand,

starts when the Recommender considers the collected information sufficient to move to the phase whose aim is to resolve a conflict of opinion, i.e. *persuasion dialogue*. During the entire conversation, even if the two participants have a distinct role, they seem to actively interact with each other in order to construct the dialogue meaning and achieve the communicative goal. The Recommender, in fact, is seen as a domain expert who participates actively, guiding the conversation throughout the two phases. The Seekers, who do not have a wide domain knowledge, mostly follow the Recommenders' moves during the exploration phase, while in the exploitation phase they provide implicit or explicit feedback that may lead the Recommender to model the dialogue, eventually finding the most suitable recommendation. Indeed, detecting seekers' communicative intentions is a pivotal process to train a conversational recommender system given that Intent Recognition is responsible for understanding the action that the user is requesting (Iovine et al., 2019). Nonetheless, in a recent review of existing approaches to conversational recommendation (Jannach et al., 2021), the author take note of a still scarce effort in investigating and defining relevant user intents, with a few exceptions considering either domain-independent intents (Cai and Chen, 2019; Narducci et al., 2018, a.o.) or restricted specific subsets (Nguyen and Ricci, 2018, e.g.).

2.1   The Inspired Corpus

The Inspired corpus (Hayati et al., 2020)[2] is a recommendation dialogue dataset of two-paired crowd-workers who chat in a natural setting in English. In each conversation, one participant acts as the Recommender, while the other as the movie Seeker. The aim of the Recommenders is to recommend a movie to the Seekers following their preferences and, thus, achieving the conversational goal successfully. The whole dataset consists of 1,001 dialogues where just the Recommender's utterances are manually annotated with the corresponding strategies. The annotation scheme of the Recommender's utterances is composed by a set of persuasive strategies divided in two categories: preference elicitation strategies and sociable strategies.

Also the collected conversations present the two-phase pattern typical of the recommendation

---

task. In the exploration phase preference elicitation strategies are used by the Recommender in order to collect sufficient information regarding the seeker's preferences and tastes about the movie domain. They are divided in *experience inquiry* and *opinion inquiry*.

In the exploitation phase, on the other hand, eight different strategies have been recognized. During this phase, thus, the Recommenders can start the interaction by *offering help* to find the recommendation. They can also express their *personal opinion* or *personal experience* in order to convince the Seekers basing the recommendation on their own experience. Moreover, they can opt for other persuasive strategies such as *credibility*, *similarity*, *encouragement*, *preference confirmation* or *self-modeling* which are mainly used to built rapport with the Seekers, also establishing and improving their role as domain experts.

3 Seeker Annotation

Taking into account the Recommender's annotation scheme proposed by Hayati and colleagues (Hayati et al., 2020) and after an inspection of the dialogues included in the Inspired Corpus, an annotation scheme for Seeker's utterances was developed. The established categories, while covering the domain-specific user intents, are in line with some of the relevant domain-independent ones found in the literature (Jannach et al., 2021, 105), e.g., *Initiate Conversation*, to "start a dialogue with the system"; *Chit-chat*, for "utterances unrelated to the recommendation goal" ; *Provide Preferences*, to "share preferences with the system"; *Ask for Recommendation*, to "obtain system suggestions"; *Obtain Explanation*, to "learn more about why something was recommended"; *Feedback on Recommendation*, to "give feedback on the provided recommendation(s)"; *Quit*, to "terminate the conversation".

We divided Seekers' strategies into four categories.[3]. The first category corresponds to a single strategy, labeled as **recommendation_request** and used by the Seeker to generically ask for a candidate item: *ex. Do you have any recommendations?*

---

[3]In this pilot stage of the research, we decided to work on the labelling of communicative strategies used by the Seekers in the above mentioned "user information gathering" and "movie recommendation" phases of dialogues. Other strategies, located at the beginning (*greetings*) and at the end of the dialogues (*intentionality, acceptance, refusal*) were also identified but they will not be discussed in this paper

The second category (henceforth called *get_movie*) includes **global requesting strategies**, by which the Seeker can direct the recommendation process on the basis of specific attributes of the movies. They are divided as follows:

- *get_from_genre*, used to ask for a candidate item according to its genre; *ex. What kind of comedy movies do you have to recommend?*

- *get_by_actor*, used to ask for a candidate item featuring a specific actor/actress; *ex. Do you have another movie with Tom Hanks?*

- *get_similar_to*, used to ask for a candidate item with analogous attributes to another specified item; *I would love to see a remake or something similar to Notting Hill.*

- *get_by_year*, used to ask for a candidate item according to its release date; *Do you know anything more recent?*

The third category corresponds to the **giving preference strategies** usually uttered by the Seeker to reply to the Recommender's inquiries:

- *personal_opinion* used to specify personal preferences over candidate items or one/some of their attributes. Also, it can express a positive or negative value towards them; *ex. I liked the acting and the movie itself; I didn't like that movie.*

- *personal_experience*, used to tell about experiences that could be present or not in the past, thus defining if the Seeker have or have not watched that movie; *ex. I saw the trailer for For v Ferrari; No, I haven't seen it.*

Finally, the *get_info* category includes **local requesting strategies** uttered by the Seeker to require information about a specific, recommended movie. This category includes:

- *get_genre*, used to asks about the value of the attribute "genre" for a specified item; *ex. Is it an action movie?*

- *get_acted_in*, used to ask about the movie's cast; *Do you know who else is in the cast?*

- *get_score*, used to request information about the quality evaluation of the movie; *ex. How about the new Rambo?*

- *get_plot*, used to ask about the storyline of a movie; *ex. Could you tell me what the general plot is?*

In order to test the validity of the annotation system, we proceeded to annotate Seekers' utterances taken from the first 20 dialogues between Recommenders and Seekers (331 utterances produced by Seekers) which were annotated by 5 annotators (the authors of this contribution). Each Seeker's utterance could be given one or two labels: a second label was added in those cases in which two strategies were expressed by the Seeker in the same utterance. In most of these cases the assignment of a first and a second label was facilitated by the sequentiality of information in the utterance (ex.the utterance *I recently watched John Wick 3, very good movie, in my opinion and fully action packed* was given *personal_experience* as first label and *personal_opinion* as second label by all the annotators); on the contrary, other cases could present a higher level of ambiguity (for example, in case an annotator intended the utterance *i like the sci-fi movies* to express both *personal_opinion* and *get_from_genre*. In these cases, there was not a unique criterion to identify which one was the first and which one was the second label). Data about annotators agreement and preliminary results of Seekers' strategies based on our annotation are presented in the following sections.

3.1 Annotation Quality

Since the annotation system accounts for the possibility of having two different strategies within the same utterance, the agreement among the 5 annotators could have 3 possible outcomes: for each utterance there could be i) agreement (A), all 5 annotators agreed on both first and second label (type and presence); ii) partial agreement (PA), at least one annotator disagreed on one strategy, though all 5 agreed on the other (e.g. all annotators agree on the first label, but no agreement is reached on the second); iii) disagreement (D), at least one disagreement for both labels.

In most cases (about 85%) the annotators agreed on at least one of the strategies detected. More specifically, A was reached in about 35% of the utterances, while PA in 50% of the cases. D was registered only for 15% of the utterances. The confusion matrix reported below shows more detailed information about the single strategies. Data reported in the matrix are mean percentages of val-
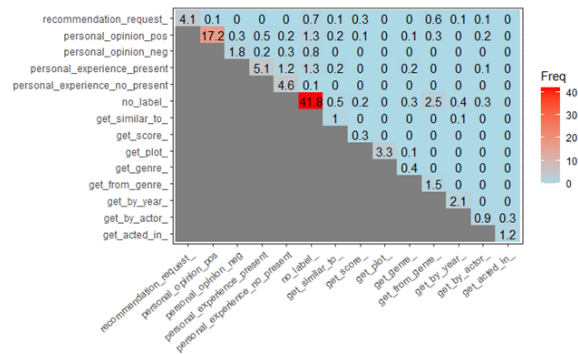

Figure 1: Confusion matrix of the labels assigned by the 5 annotators. The data reported were calculated as mean percentages over the whole dataset of pairwise annotator agreement. No-label refers to the absence of a second strategy; opinion and experience were split into two different strategies according to their evaluation (positive vs negative for experience and present vs not present for experience)

ues of the 10 pairs of annotators: label-by-label agreement was first calculated for each pair of annotators and then mean values for all the pairs were extracted and plotted in the matrix to check for which strategies reported, on average, the highest levels of agreement or disagreement across the annotators (Figure 1).

It is clear from the matrix that most cases of disagreement refer to *get_from_genre*. More generally, the matrix shows that among the cases of disagreement, the annotators failed to agree on the assignment of labels relative to global and local requesting strategies, which were often annotated as not representing a specific strategy at all. A sounder measure for the agreement (Fleiss' Kappa) was calculated for those utterance in which all annotators agreed to assign only one label, which amount to about 1/3 of the total of the utterances[4]. The Fleiss' Kappa value obtained for these annotation is 0.887, indicating an overall high agreement among the 5 annotators. The inspection of the score obtained for each specific label shows that while all strategies were detected with a high level of agreement, low values are registered for the category *get_from_genre* (Kappa = 0.247)

---

[4]The measure was not calculated for the whole data set because of the absence of a stable criterion for ordering strategies in case two were present (see section 3).
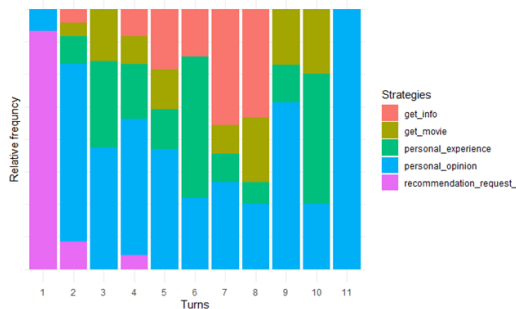
Figure 2: Distribution of the seeker's strategies in each turn of the dialogues.

Figure 3: Distribution of the recommender's strategies in each turn of the dialogues. (Hayati et al., 2020)

4 Retrieved Data

This section presents a description of the strategies employed by Seekers in the subset that we analyzed. The data reported here refer to those utterances in which all 5 annotators agreed on the type of strategy detected.

The most frequent strategy is the expression of personal opinions, which alone accounts for almost 50% of the total of the strategies. Of these, the great majority (around 90%) is represented by the strategy 'personal_opinion_pos'. The strategy 'personal_experience' is also quite frequent, amounting to around 20% of the strategies; among these, the expression of absence of experience (*ex. No, I haven't seen that movie*) is more frequent, accounting for more that 60%. Recommendation requests account for 10% of the strategies, while the remainder is made up of those strategies aiming at either collecting specific information about a movie (i.e., get_info) or eliciting a title given a specific preference (i.e., get_movie). Of the former set of strategies, the information that is more frequently asked concerns the plot of the movie, while for the latter, Seekers appear to be most interested in the release date. Although annotated data about the Seekers' turns are referred to a small subset of the whole corpus, it is possible to draw some preliminary strategies on the co-construction of the dialogue by the two participants, by considering the by-turn distribution of the strategies in both participants. As for the Seeker, the different strategies employed are not evenly distributed across the dialogue turns, as shown in Fig.2. The plot shows that recommendation requests are almost the only strategy employed at the beginning of the dialogue, after which their occurrence drops dramatically. On the contrary, the occurrence of get_info and get_movie
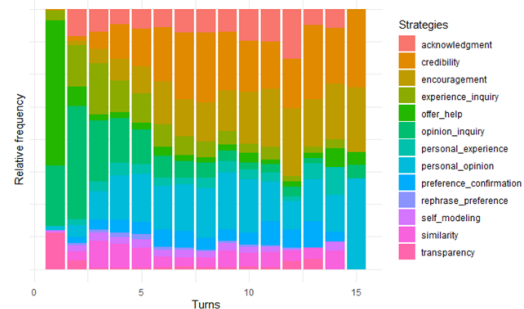
increase as the dialogue unfolds. Personal opinion and experience, on the other hand, are more evenly distributed, with a drop of their occurrence in the median turns. As for the Recommender, the by-turn distribution of strategies is shown in Fig.3.

The plot shows that, on the Recommender side, the use of the strategy offer_help mirrors the use by the Seeker of a request for recommendation, being employed almost exclusively in the first turn. More generally, the first part of the dialogue is characterized by inquiries, by the Recommender to the Seeker, about his/her opinions and experiences. While the use of these strategies decreases as the dialogue unfolds, strategies aimed at overcoming conflicts (e.g. preference confirmation) or persuading/informing (e.g. encouragement or similarity) are more frequent in the second half of the dialogue. This is mirrored, on the Seeker side, by the use of strategies linked to personal opinions/experiences and global and local requesting strategies.

5 Discussion and Conclusions

This work is supported by the idea that studying communication strategies used by human speakers is fundamental to improve the performances of dialogue systems. This was already supported by Hayati and her colleagues (2020) who analyzed the Recommenders' sociable strategies in recommendation dialogues to develop successful and persuasive recommendation dialogue systems. However, considering the cooperative nature of dialogues, we argue that annotating the Seeker's move may be pivotal in the training phase of recommendation dialogue systems. Hence, we propose an annotation scheme for the Seeker's utterances that is coherent with the annotation of Rec-

ommender's utterances. Considering the Seeker's role and main moves, we have drawn four categories: *recommendation requests*, *global requesting strategies*, *giving preference strategies* and *local requesting strategies*. Results on the reliability of the annotation scheme show that the agreement between the 5 annotators ranges from substantial to almost perfect (Landis and Koch, 1977) for most strategies but one, i.e., the strategy used to ask for movies of a specific genre. Similarly, observing the other cases of disagreement, we find that they mostly concern the identification of global and local requesting strategies. We showed that in most of these cases annotators failed to agree on whether an utterance contained a second strategy (manly a specific title request). In this cases, some annotators assigned a second label believing that the more specific request was generated as a conversational implicature stemming from the Seeker's mention of a certain movie title or attribute and the expression of his/her own opinions and experiences. The fact that most of the cases of disagreement fall within this situation might also explain why we registered high levels of disagreement for the *get_from_genre* label. Observing the confusion matrix (Figure 1), what can be noticed is that this category has been frequently confused with the no_label one. An explanation of this phenomenon could be found in utterances like "I love sci-fi movie" to which only the first label as personal_opinion_pos has been assigned. Nonetheless, other annotators also added get_from_genre as second label, for the reason explained above. We believe that this does not specifically depend on the strategy per se, but simply on the fact that genre is the feature of a movie that most frequently was mentioned by the Seekers (30% of the total features, as opposed to i.e. actors and directors, occurring respectively, in 20% and 4% of the cases), therefore more frequently led the annotators to assign different strategies. A finer analysis of the turn by turn strategies of the two participants on a larger number of dialogues would be informative about the extent to which Recomemenders make the inference (and act on it). This would help understand how to treat these cases.

Concerning the general distribution of the Seekers' strategies, positive personal opinion and non-present personal experience seem to be more frequent than the global and local requesting strate-gies. The strategies distribution along with the dialogue turns, on the other hand, shows that the first turns are mainly characterized by the occurrence of recommendation requests, reflecting the Recommender's strategy of offering help. In the middle of the conversation, requests for getting information or movie titles increase together with personal opinion and personal experience, even if the latter seems to be more equally distributed. This distribution could reflect the fundamental role of the Seeker in modelling the conversation. In the first phase of exploration the Seekers' personal opinions are explicitly elicited by the Recommenders' inquiries. Instead, in the exploitation phase, the Seeker could also provide *soft* evidence of their preferences, which may be used by the Recommender to help the Seeker find a suitable item. This attitude is very common in human-human dialogue with respect to the human-machine interaction, since it follows the principles of cooperative dialogue (Grice, 1975). For this reason, Recommender systems that adopt a proactive behaviour and take the initiative to provide a piece of information that is not explicitly requested, should be able to better achieve the user needs and fulfil the goal of the dialogue (Balaraman and Magnini, 2020).

Acknowledgments

References

Vevake Balaraman and Bernardo Magnini. 2020. Proactive systems and influenceable users: Simulating proactivity in task-oriented dialogues. In *The 24th Workshop on the Semantics and Pragmatics of Dialogue (WatchDial'20)*.

Carla Bazzanella. 2005. Linguistica e pragmatica del linguaggio. un'introduzione.

Wanling Cai and Li Chen. 2019. Towards a taxonomy of user feedback intents for conversational recommendations. In *RecSys (Late-Breaking Results)*, pages 51–55.

Maria Di Maro. 2021. *"Shouldn't I use a polar question?" Proper Question Forms Disentangling Inconsistencies in Dialogue Systems*. Ph.D. thesis, Mind, Gender and Language, University of Naples Federico II.

Zuohui Fu, Yikun Xian, Yongfeng Zhang, and Yi Zhang. 2020. Tutorial on conversational recommendation systems. In *Fourteenth ACM Conference on Recommender Systems*, pages 751–753.

Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. Advances and challenges in conversational recommender systems: A survey. *arXiv preprint arXiv:2101.09459*.

Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. Inspired: Toward sociable recommendation dialog systems. *arXiv preprint arXiv:2009.14306*.

Andrea Iovine, Fedelucio Narducci, and Marco de Gemmis. 2019. A dataset of real dialogues for conversational recommender systems. In *CLiC-it*.

Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)*, 54(5):1–36.

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Fedelucio Narducci, Pierpaolo Basile, Andrea Iovine, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2018. A domain-independent framework for building conversational recommender systems. In *KaRS@ RecSys*, pages 29–34.

Thuy Ngoc Nguyen and Francesco Ricci. 2018. A chat-based group recommender system for tourism. *Information Technology & Tourism*, 18(1):5–28.

Henry Prakken. 2018. *Historical overview of formal argumentation*, volume 1. College Publications.

Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of persuasive dialogues: testing bot identities and inquiry strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Douglas N Walton and Erik CW Krabbe. 1995. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press.

Douglas N Walton. 1984. Logical dialogue-games and fallacies.