Conditions of an appropriate dataset for movie conversational recommender: The users' point of view

Tayyebeh Saeedi* Ali Mohades† Hossein Zeinali‡

Abstract

Current models of the Conversational Recommender System (CRS) tend to suggest only one item at each turn, which may increase the length of the conversation and lead to user impatience. To overcome this limitation, we conducted a questionnaire with 798 participants in order to assess user preferences for multi-item recommendations within a conversational turn in the movie recommendation domain. The results of the questionnaire clearly show that users have a strong preference for multi-item recommendations within each turn, especially when the interaction is with experts. We also discuss that the datasets from existing CRS may not be adequate to train models to meet users' expectations for multi-item recommendations. This shows it is important that have a dataset that better aligns CRS models with real-world preferences. This work forms the foundation upon which further work on developing datasets of multi-item recommendations will be delivered for better conversational recommender systems.

Keywords: Conversational recommender systems, multi-item recommendations, user preferences, CRS datasets, movie recommendation, and personalized recommendations.

1 Introduction

Conversational Recommendation Systems (CRSs) have emerged as an attractive research area in recent years [15, 34, 35, 40, 42] due to the substantial expansion of intelligent systems, including chatbots and various e-commerce platforms. In contrast to traditional recommender systems, which focus on user-item interaction data to generate recommendation results [1, 21, 38, 44], CRSs aim to dynamically interact with users using natural language, infer their preferences, and provide high-quality recommendations [4, 6, 30]. Moreover, while traditional recommender systems present many items to users, requiring them to choose from the recommended options, CRSs offer a more engaging and personalized experience [30].

As mentioned, CRSs provide more personalized recommendations to users while avoiding confusion users by not presenting them with too many options. However, to the best of our knowledge, most existing CRSs recommend only one item at each turn of conversation [16, 22, 23, 26, 32, 33, 36, 37, 41, 42, 43, 45]. Consequently, the conversation may be too long and users may leave the conversation before receiving a recommendation that aligns with their preferences, cause of a lack the patience for lengthy chats. Therefore, our main research question is: **how many items should be recommended at each turn of the conversation?**

To address the above question, we first focused on the movie conversational recommendation domain, which is the most popular industry. Then, we checked why existing CRSs typically recommend only one movie at each turn of conversation and whether there is a way to offer more items. We found that this limitation comes from the structure of existing CRS datasets in the movie domain, such as ReDial [14] and INSPIRED [9], where most conversations involve recommending a single item at each turn. So, our question evolved into another: **Are the current datasets suitable, or is it necessary to develop another dataset, that recommends more than one item at each turn?** To answer this question, we created an online questionnaire to gather user preferences on how many movies they would like to be recommended at each turn of conversation. In the questionnaire, we focused on three movie recommendation scenarios: when a movie is recommended by a friend, by an expert, or by the users themselves to someone else. You can find our questionnaire in the **Appendix A**.

In this paper, we conducted field research to determine user preferences regarding how many movies should be recommended at each turn of the conversation. Our findings revealed that users prefer more than one movie to be recommended per turn (See Section 4 for more details). This indicates the need for a new dataset based on this condition, enabling conversational recommender system (CRS) models to be trained to recommend more than one movie at each turn.

*Department of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), t.saeedi@aut.ac.ir

†Department of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), mohades@aut.ac.ir

‡Department of Computer Engineering, Amirkabir University of Technology (Tehran Polytechnic), hzeinali@aut.ac.ir

The general structure of this paper is as follows: Section 2 presents a literature review of available datasets in the conversational recommender system field. Section 3 provides the methodology of this research, which is based on a questionnaire. Section 4 presents the results of the questionnaire and discusses the results. Finally, Section 5 provides the conclusion of this research.

2 Literature Review

Conversational recommender systems are pretty new on the research scene, so there aren't many public datasets available yet. Because of this, some researchers in this field primarily worked with traditional recommender system datasets (including Amazon [25], BookCrossing[1], Yelp[2], LastFM [2], MovieLens [8], etc.), adapting them to a Multi-round Conversational Recommendation (MCR) setting, as seen in studies like [3, 17, 26, 27, 29, 33, 37]. Creating datasets for conversational recommendation systems task is hard since these datasets need to have certain features to work well. A good dataset should (1) have conversations and (2) recommend items to the user during turns of conversations. Also, to train a conversation recommendation system model in a certain domain (like movies, books, or music), the dataset needs to focus on that domain, because each domain has its own special items and each item has special attributes.

However, researchers in this field built conversational datasets for this task, some of which are available. Table 1 shows all available conversational datasets in the CRS task. Other than the datasets that are presented in Table 1, other conversational datasets for conversational recommender systems have been proposed in recent research, such as MGConvRex [39], E-ConvRec [10], PEARL [12], which yet are still unpublished. As noted in [13], among datasets mentioned in Table 1, the ReDial dataset is currently the most widely used in research. Nevertheless, in almost all its conversations, just one item is recommended to the seeker by the recommender at each turn. It shows that even a benchmark dataset in the CRS field does not focus on the user request and has approximately 10 turns per dialogue, which is a lot for impatient users.

In this study, we focused on the user's request for how many movie recommendations at each turn of the conversation. For this, we used field research with a questionnaire and asked users about their requests over all the world. In the next section, we discussed this in detail.

[1]https://grouplens.org/datasets/book-crossing/

[2]https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset

3 Methodology

As mentioned in the introduction (Section 1), available datasets in the domain of movie conversational recommendation, almost recommend just only one movie per conversational turn. In this study, we focused on investigating if such recommendations are interesting to users, and for this, a questionnaire was developed to measure potential multi-item recommendations usage in conversations. This questionnaire was made available in both Persian and English languages in order to reach users of different cultures and backgrounds and collect data relevant to the research on the features of conversational recommendations. (For more details, the English version of this questionnaire is available in **Appendix A**, and the Persian one is just translated from that). We made this bi-lingual questionnaire with Google Forms[3] and shared it on social media networks, such as LinkedIn, WhatsApp, Telegram, and Instagram for Iranian and non-Iranian communities.

The questionnaire was filled out by 798 respondents, of whom 749 were identified as Iranian and 49 as non-Iranian. Such a sample would ensure that both the local as well as the international views are more or less represented. This sampling approach was adopted to explore any cultural variations in conversational recommendation expectations and to determine a general preference pattern that could be generalized across different user groups. The respondents field this questionnaire from 27th September 2023 until 4th February 2024. The statistics of respondents are provided in Table 2. As you can see, most of them are women, and the average age of all respondents is 30.18.

As you can see details of the questionnaire in **Appendix A**, is obvious, that the focus of this questionnaire is the number of movie recommendations that users would like to receive in each conversational turn. We aimed to focus on questions that would elicit the situation of recommendation solicitation, either by friends, professionals, or themselves, to see if preferences would differ with the situation.

Before we analyzed the results of the collection data, we did some preprocessing on them. For example, some respondents explained their answers, and we changed their explanation to a number between 1 and 4, or an expression (more than 4), as our questionnaire was designed. Also, some respondents wrote their answers as 1 or 2 movie recommendations; we changed half of them to 1 and others to 2. Fortunately, such answers of the respondents who watched movies in their free time are an even number, so it could be divided by 2.

In the next section, we will discuss about the results of the collection data.

[3]https://docs.google.com/forms

Table 1: Available conversational datasets for CRS task

Dataset	#Dialogue	#Utterance	Domain	Language	Published Year	URL
ReDial [14]	10,006	182,150	Movie	English	2018	`https://redialdata.github.io/website/`
CCPE-M [31]	502	11,972	Movie	English	2019	`https://github.com/google-research-datasets/ccpe`
GoRecDial [11]	9,125	170,904	Movie	English	2019	`https://drive.google.com/drive/folders/1nilk6FUktW2VjNlATdM0VMehzSOPIvJ0`
OpenDialKG [28]	15,673	91,209	Movie Book Sport Music	English	2019	`https://github.com/facebookresearch/opendialkg`
DuRecDial [20]	15,673	91,209	Movie Music Movie star Food Restaurant News Weather	Chinese	2020	`https://baidu-nlp.bj.bcebos.com/DuRecDial.zip`
TG-ReDial [46]	10,000	129,392	Movie	Chinese	2020	`https://github.com/RUCAIBox/TG-ReDial`
INSPIRED [9]	1,001	35,811	Movie	English	2020	`https://github.com/sweetpeach/Inspired`
COOKIE [5]	–	11.6 M	E-commerce	English	2020	`https://github.com/zuohuif/COOKIE`
DuRecDial 2.0 [19]	16,482	255,346	Movie Music Movie star Food Restaurant Weather	English Chinese	2021	`https://github.com/liuzeming01/DuRecDial`
U-NEED [18]	7,698	53,712	E-commerce	Chinese	2023	`https://github.com/LeeeeoLiu/U-NEED`
E-ReDial [7]	756	12,003	Movie	English	2023	`https://github.com/Superbooming/E-Redial`
MobileConvRec [24]	12.2 K	156 K	All 45 Categories on Google Play†	English	2024	`https://huggingface.co/datasets/recmeapp/MobileConvRec`

† Including food & drink, news & magazines, music, shopping, social, sports, weather, etc.

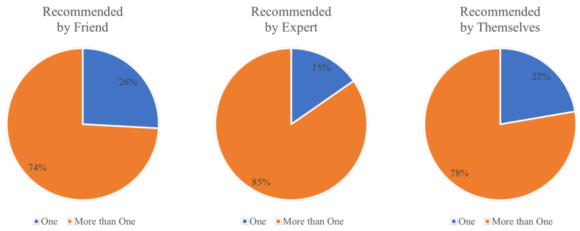
Table 2: The statistics of respondents

Nationality	Iranian	749
	Non-Iranian	49
Gender	Female	546
	Male	248
	Other	4
Age	Min.	13
	Max.	67
	Avg.	30.18
Watch Movie	Yes	704
	No	94

4 Results and Discussion

In this section, for a better understanding of the results of the requests of respondents, we visualized results in the pie charts. As shown in Table 2, 704 respondents watched movies in their free time, which is 88% of all respondents. The following results are based on these respondents. The analysis of the questionnaire responses showed a notable preference among users for receiving more than one movie recommendation at each turn of conversational. Figure 1 shows this result for all three scenarios (when a movie is recommended by a friend, by an expert, or by the users themselves to someone else) in the percentage. As shown in Figure 1, 74%, 85%, and 78% of respondents would like to be recommended more than one movie at each turn of conversation by a friend, by an expert, and by themselves, respectively.

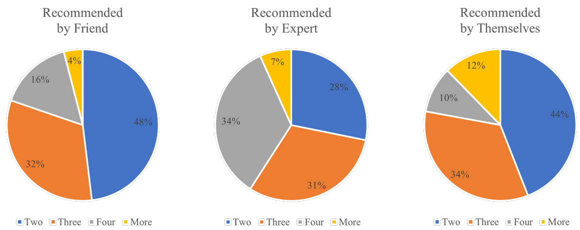
In the next step, we analyzed how much of them would like to be recommended by 2, 3, 4, or more than 4 movies per conversational turn. These results are shown in Figure 2. As shown in Figure 2, when movies would like to be recommended by a friend, most of the respondents (44% of them) want two movies recommended at each turn. Furthermore, when one of the people in the



(a) (b) (c)

Figure 1: (a) The percentage of the number of movies recommended by friends. (b) The percentage of the number of movies recommended by expert. (c) The percentage of the number of movies recommended by themselves to other.

conversation is an expert, most persons (34% of them) would like to be recommended four movies per turn. In the end, when users recommended movies to others, most of the respondents (44% of them) would like to recommend two movies.



(a) (b) (c)

Figure 2: (a) The percentage of the number of movies recommended more than two by friends. (b) The percentage of the number of movies recommended more than two by expert. (c) The percentage of the number of movies recommended more than two by themselves to other.

These results give us two insights: (1) Users' expectations for recommendation volume differ depending on the perceived expertise of the recommendation source, and (2) a conversational recommender system is more user-friendly and better aligns with real-world user expectations when the chatbot recommends more than one movie at each turn of conversation. The second insight proves that it is necessary to build or collect a conversation dataset for the CRS task, which recommends at least two items per conversational turn. The building of such a dataset will remain for future work since it is a hard task and needs at least 900 people for crowd-work gathering data.

5 Conclusion

This paper indicates the limitation of the current conversational recommendation systems, where most datasets—including ReDial—usually limit each conversational turn to just one recommendation. This field research was conducted as a bilingual questionnaire in which respondents showed a fantastic preference for multiple movie suggestions per turn. This preference suggests that current CRS models are still restricted to datasets of single-item recommendation datasets and thus fail to capture the expectations of the user, making the conversation sometimes lengthy and not so intrinsically interesting.

Our findings support the fact that there is indeed a need for the development of a CRS dataset that can support multi-item recommendations since this will align better with real-world user interactions. Such a dataset could improve user satisfaction in CRS applications, making them more responsive to user preferences. This would call for further work resulting in datasets containing these multi-item recommendation preferences to support enhanced CRS capabilities for general and expert-driven recommendation scenarios.

References

[1] M. O. Ayemowa, R. Ibrahim, and M. M. Khan. Analysis of recommender system using generative artificial intelligence: A systematic literature review. *IEEE Access*, 2024.

[2] Ò. Celma Herrada et al. *Music recommendation and discovery in the long tail*. Universitat Pompeu Fabra, 2009.

[3] Z. Chu, H. Wang, Y. Xiao, B. Long, and L. Wu. Meta policy learning for cold-start conversational recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 222–230, 2023.

[4] T. Di Noia, F. M. Donini, D. Jannach, F. Narducci, and C. Pomo. Conversational recommendation: Theoretical model and complexity analysis. *Information Sciences*, 614:325–347, 2022.

[5] Z. Fu, Y. Xian, Y. Zhu, Y. Zhang, and G. de Melo. Cookie: A dataset for conversational recommendation over knowledge graphs in e-commerce. *arXiv preprint arXiv:2008.09237*, 2020.

[6] C. Gao, W. Lei, X. He, M. de Rijke, and T.-S. Chua. Advances and challenges in conversational recommender systems: A survey. *AI open*, 2:100–126, 2021.

[7] S. Guo, S. Zhang, W. Sun, P. Ren, Z. Chen, and Z. Ren. Towards explainable conversational recommender systems. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2786–2795, 2023.

[8] F. M. Harper and J. A. Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015.

[9] S. A. Hayati, D. Kang, Q. Zhu, W. Shi, and Z. Yu. Inspired: Toward sociable recommendation dialog systems. In *Proceedings of the 2020 Conference on Empiri-*

cal Methods in Natural Language Processing (EMNLP), pages 8142–8152, 2020.

[10] M. Jia, R. Liu, P. Wang, Y. Song, Z. Xi, H. Li, X. Shen, M. Chen, J. Pang, and X. He. E-convrec: a large-scale conversational recommendation dataset for e-commerce customer service. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5787–5796, 2022.

[11] D. Kang, A. Balakrishnan, P. Shah, P. A. Crook, Y.-L. Boureau, and J. Weston. Recommendation as a communication game: Self-supervised bot-play for goal-oriented dialogue. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1951–1961, 2019.

[12] M. Kim, M. Kim, H. Kim, B.-w. Kwak, S. Chun, H. Kim, S. Kang, Y. Yu, J. Yeo, and D. Lee. Pearl: A review-driven persona-knowledge grounded conversational recommendation dataset. *arXiv preprint arXiv:2403.04460*, 2024.

[13] C. Li, H. Hu, Y. Zhang, M.-Y. Kan, and H. Li. A conversation is worth a thousand recommendations: A survey of holistic conversational recommender systems. In *KaRS Workshop @ RecSys*, 2023.

[14] R. Li, S. Ebrahimi Kahou, H. Schulz, V. Michalski, L. Charlin, and C. Pal. Towards deep conversational recommendations. *Advances in neural information processing systems*, 31, 2018.

[15] S. Li, W. Lei, Q. Wu, X. He, P. Jiang, and T.-S. Chua. Seamlessly unifying attributes and items: Conversational recommendation for cold-start users. *ACM Transactions on Information Systems (TOIS)*, 39(4):1–29, 2021.

[16] D. Lin, J. Wang, and W. Li. Cola: Improving conversational recommender systems by collaborative augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4462–4470, 2023.

[17] H. Liu, Y. Zhang, P. Li, C. Qian, P. Zhao, and X. Wu. Deepcpr: Deep path reasoning using sequence of user-preferred attributes for conversational recommendation. *ACM Transactions on Knowledge Discovery from Data*, 18(1):1–22, 2023.

[18] Y. Liu, W. Zhang, B. Dong, Y. Fan, H. Wang, F. Feng, Y. Chen, Z. Zhuang, H. Cui, Y. Li, et al. U-need: A fine-grained dataset for user needs-centric e-commerce conversational recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2723–2732, 2023.

[19] Z. Liu, H. Wang, Z.-Y. Niu, H. Wu, and W. Che. DuRecDial 2.0: A bilingual parallel corpus for conversational recommendation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4335–4347, Online and Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics.

[20] Z. Liu, H. Wang, Z.-Y. Niu, H. Wu, W. Che, and T. Liu. Towards conversational recommendation over multi-type dialogs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1036–1049, 2020.

[21] L. Long, F. Huang, Y. Yin, and Y. Xu. Multi-task learning for collaborative filtering. *International Journal of Machine Learning and Cybernetics*, pages 1–14, 2022.

[22] Y. Lu, J. Bao, Y. Song, Z. Ma, S. Cui, Y. Wu, and X. He. Revcore: Review-augmented conversational recommendation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1161–1173, 2021.

[23] W. Ma, R. Takanobu, and M. Huang. Cr-walker: Tree-structured graph reasoning and dialog acts for conversational recommendation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1839–1851, 2021.

[24] S. Maji, M. Fereidouni, V. Chhetri, U. Farooq, and A. Siddique. Mobileconvrec: A conversational dataset for mobile apps recommendations. *arXiv preprint arXiv:2405.17740*, 2024.

[25] J. McAuley, C. Targett, Q. Shi, and A. Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 43–52, 2015.

[26] A. Montazeralghaem and J. Allan. Learning relevant questions for conversational product search using deep reinforcement learning. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 746–754, 2022.

[27] A. Montazeralghaem, J. Allan, and P. S. Thomas. Large-scale interactive conversational recommendation system using actor-critic framework. In *Proceedings of the 15th ACM conference on recommender systems*, pages 220–229, 2021.

[28] S. Moon, P. Shah, A. Kumar, and R. Subba. Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 845–854, 2019.

[29] Y. Ni, Y. Xia, H. Fang, C. Long, X. Kong, D. Li, D. Yang, and J. Zhang. Meta-crs: A dynamic meta-learning approach for effective conversational recommender system. *ACM Transactions on Information Systems*, 42(1):1–27, 2023.

[30] D. Pramod and P. Bafna. Conversational recommender systems techniques, tools, acceptance, and adoption: A state of the art review. *Expert Systems with Applications*, 203:117539, 2022.

[31] F. Radlinski, K. Balog, B. Byrne, and K. Krishnamoorthi. Coached conversational preference elicitation: A case study in understanding movie preferences. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 353–360, Stockholm, Swe-

den, Sept. 2019. Association for Computational Linguistics.

[32] X. Ren, T. Chen, Q. V. H. Nguyen, L. Cui, Z. Huang, and H. Yin. Explicit knowledge graph reasoning for conversational recommendation. *ACM Transactions on Intelligent Systems and Technology*, 2023.

[33] X. Ren, H. Yin, T. Chen, H. Wang, Z. Huang, and K. Zheng. Learning to ask appropriate questions in conversational recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 808–817, 2021.

[34] X. Ren, H. Yin, T. Chen, H. Wang, N. Q. V. Hung, Z. Huang, and X. Zhang. Crsal: Conversational recommender systems with adversarial learning. *ACM Transactions on Information Systems (TOIS)*, 38(4):1–40, 2020.

[35] Y. Sun and Y. Zhang. Conversational recommender system. In *The 41st international acm sigir conference on research & development in information retrieval*, pages 235–244, 2018.

[36] X. Wang, K. Zhou, J.-R. Wen, and W. X. Zhao. Towards unified conversational recommender systems via knowledge-enhanced prompt learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1929–1937, 2022.

[37] Y. Wang, X. Chen, J. Fang, Z. Meng, and S. Liang. Enhancing conversational recommendation systems with representation fusion. *ACM Transactions on the Web*, 17(1):1–34, 2023.

[38] T. Wei and J. He. Comprehensive fair meta-learned recommender system. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1989–1999, 2022.

[39] H. Xu, S. Moon, H. Liu, B. Liu, P. Shah, and S. Y. Philip. User memory reasoning for conversational recommendation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5288–5308, 2020.

[40] H. Yang, H. Won, Y. Ahn, and K.-H. Lee. Click: Contrastive learning for injecting contextual knowledge to conversational recommender system. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1875–1885, 2023.

[41] C. Zhang, X. Huang, and J. An. Macr: Multi-information augmented conversational recommender. *Expert Systems with Applications*, 213:118981, 2023.

[42] C. Zhang, X. Huang, J. An, and S. Zou. Improving conversational recommender systems via multi-preference modelling and knowledge-enhanced. *Knowledge-Based Systems*, 286:111361, 2024.

[43] T. Zhang, Y. Liu, B. Li, P. Zhong, C. Zhang, H. Wang, and C. Miao. Toward knowledge-enriched conversational recommendation systems. In *Proceedings of the 4th Workshop on NLP for Conversational AI*, pages 212–217, 2022.

[44] Y. Zhao, S. Wang, Y. Wang, and H. Liu. Mbsrs: A multi-behavior streaming recommender system. *Information Sciences*, 631:145–163, 2023.

[45] K. Zhou, W. X. Zhao, S. Bian, Y. Zhou, J.-R. Wen, and J. Yu. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1006–1014, 2020.

[46] K. Zhou, Y. Zhou, W. X. Zhao, X. Wang, and J.-R. Wen. Towards topic-guided conversational recommender system. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4128–4139, 2020.

Appendix A

Our questionnaire was designed to be concise, with the following questions, to encourage more people to complete it. It was created using Google Forms.

1. Gender:
   - ◯ Female
   - ◯ Male
   - ◯ Other: – – – – – – – – – – – –
2. Age: – – – – – – – – – –
3. Do you watch movies in your free time?
   - ◯ Yes
   - ◯ No
4. When you talk to your friend and you want them to recommend you some movies, how many movies do you prefer to be recommended to you **at each turn of the conversation**?
   - ◯ 1
   - ◯ 2
   - ◯ 3
   - ◯ 4
   - ◯ Other: – – – – – – – – – – – –
5. When talking to an expert about movies, how many movies would you prefer to be recommended by the expert **at each turn of the conversation**?
   - ◯ 1
   - ◯ 2
   - ◯ 3
   - ◯ 4
   - ◯ Other: – – – – – – – – – – – –
6. If you were to recommend movies to others, how many movies would you recommend **at each turn of the conversation**? (Please enter a number)
   – – – – – – – – – – – – –