Can Virtual Assistants Produce Recommendations?

Dimitrios Rafailidis
Maastricht University
Maastricht, The Netherlands
dimitrios.rafailidis@maastrichtuniversity.nl

Yannis Manolopoulos
Aristotle University of Thessaloniki
Thessaloniki, Greece
manolopo@csd.auth.gr

ABSTRACT

Virtual assistants, also known as intelligent conversational systems such as Google's Virtual Assistant and Apple's Siri, interact with human-like responses to users' queries and finish specific tasks. Meanwhile, existing recommendation technologies model users' evolving, diverse and multi-aspect preferences to generate recommendations in various domains/applications, aiming to improve the citizens' daily life by making suggestions. The repertoire of actions is no longer limited to the one-shot presentation of recommendation lists, which can be insufficient when the goal is to offer decision support for the user, by quickly adapting to his/her preferences through conversations. Such an interactive mechanism is currently missing from recommendation systems. This article sheds light on the gap between virtual assistants and recommendation systems in terms of different technological aspects. In particular, we try to answer the most fundamental research question, which are the missing technological factors to implement a personalized intelligent conversational agent for producing accurate recommendations while taking into account how users behave under different conditions. The goal is, instead of adapting humans to machines, to actually provide users with better recommendation services so that machines will be adapted to humans in daily life.

CCS CONCEPTS

•Information systems →Collaborative and social computing systems and tools;

KEYWORDS

Virtual assistants; recommendation systems; chatbots; conversational systems

1 INTRODUCTION

Recommendation systems are intelligent agents that elicit the interests and preferences of individuals and make recommendations accordingly [15]. Recommendation systems not only have the potential to narrow down the search space of the information overload, but also to support and improve the quality of the decisions that people make in daily life. With the advent of machine learning strategies, recommendation systems can now intelligently elicit user preferences and capture their complex associations to make suggestions [6]. However, compared to existing machine learning strategies in recommendation systems, in practice there are several opportunities to elicit user information by making the underlying machine learning models more conversational and collaborative [41]. Meanwhile, recent advances in Artificial Intelligence (AI) have enabled new forms of human-computer interaction characterized by greater adaptability and better human-machine symbiosis. To facilitate the development of next generation AI agents that can truly understand and collaborate with humans, it is important that AI agents can understand and adapt to individual differences or personality traits. The AI upsurge allowed us to talk to computers via commands. Intelligent conversational agents (virtual assistants) have allowed us not to just talk to machines, but also accomplish our daily tasks. For example, Google's Virtual Assistant, Apple's Siri, Amazon Alexa, and Microsoft Cortana have revolutionized the way we interact with phones and machines. These virtual assistants are termed as "dialogue systems often endowed with human-like behaviour", and they have started becoming integral parts of people's lives. Although both recommendation systems and virtual assistants are based on various machine learning strategies, there is a large technological gap between them [48]. There are immense problems lying in the field of virtual assistants and recommendation systems to be solved to reach the desired goal, that is the real adaptation of machines to our personal preferences while generating personalized recommendations. Existing solutions in conversational recommendation systems are either based on single round ad-hoc search engines or traditional multi-round dialog systems [7, 8], ignoring users' evolving, diverse and multi-aspect preferences when producing recommendations. The most fundamental question is:

How can we provide people with an AI friend who will talk and give suggestions just like a human friend would have done?

Our knowledge of bridging the gap between virtual assistants and recommendation systems is flawed. There have been many studies of virtual assistants and recommendation systems based on machine learning strategies, but no unified approach that forms a single conversational recommendation system. This article deals with the technological gap between virtual assistants and recommendation systems, shedding light on ways to develop a unified framework, not only to capture users' evolving, diverse and multi-aspect preferences, but also to consider users' interactions with the recommendation system via conversations, and adjust recommendations on-the-fly.

The rest of the paper is organized as follows. In Section 2 we first introduce the functionality of recommendation systems, whereas in Section 3 we detail how virtual assistants are currently used in daily life. In Section 4 we discuss the technological gap between virtual

WIMS2019, Seoul, Republic of Korea

DOI: 10.1145/3326467.3326468

assistants and recommendation systems; Section 5 concludes the study.

2 RECOMMENDATION SYSTEMS

The utility of recommendation systems cannot be overstated, given its widespread adoption in many web applications, along with its potential impact to ameliorate many problems related to overchoice. Recommendation systems provide value for people by narrowing down the set of choices and helping them explore the space of available options, or serve as a filtering component in situations of information overload. From the provider perspective, recommendation systems are personalized services that increase users' trust and loyalty, as well as obtain more knowledge about what people are really looking for. Given the explosive growth of information available on the web and Internet of Thing devices, users are often greeted with more than countless products, movies, restaurants, information about healthcare services and so on. As such, personalization is an essential strategy for facilitating user experience. Recommendation systems have been playing a vital and indispensable role in various information access systems to boost business and facilitate decision-making processes, and are pervasive across numerous web domains such as e-commerce, news and media websites. For example, 80% of movies watched on Netflix came from recommendations [9], 60% of video clicks came from home page recommendation on YouTube [4], whereas Amazon announced that 35% of sales comes from recommendation systems. The core mechanisms of recommendation systems are mainly categorized into collaborative filtering, content-based recommendation systems and hybrid recommendation systems based on the types of input data. Collaborative filtering makes recommendations by learning from user-item interactions [15], content-based recommendation is based on comparisons across items' and users' auxiliary information, such as text, images and videos [16], and hybrid models refer to recommendation systems that integrate collaborative and content-based strategies, to solve the data scarcity of user preferences and the cold-start problem with users having a poor history record [19]. In a similar spirit, over the past decade recommendation algorithms for rating prediction and item ranking have steadily matured with matrix factorization and other latent factor models emerging as state-of-the-art algorithms to apply in both existing and new applications/domains.

However, the recommendation systems algorithms are typically applied in relatively straightforward and static scenarios: given information about a user's past item preferences, can we predict whether s/he will like a new item or rank all unseen items based on the predicted interest? In reality, recommendation is often a more complex problem, as the evaluation of a list of recommended items never takes place in a vacuum. With richer user interaction models, more elaborate recommendation systems become possible, which can stimulate, accept and process various types of user input. At the same time the repertoire of actions is no longer limited to the one-shot presentation of recommendation lists, which can be insufficient when the system goal is to offer decision support to the user. State-of-the-art methods of recommendation systems are not applicable to a majority of practical scenarios due to the dynamic change of content, for example, latest news, new products, and so on. Thus, it is highly desirable to quickly adapt to users' preferences on new content through effective interactive mechanisms, such as conversations.

3 VIRTUAL ASSISTANTS IN DAILY LIFE

The growth of the global virtual assistant market is being primarily driven by the penetration of smartphones along with a rapid growth in the social media traffic, which has led to a substantial rise in the consumer awareness about benefits offered by virtual assistants. With virtual assistants, users and systems can interact for multiple semantically coherent rounds on a task through natural language dialog. Thus, it becomes possible for the system to understand the user needs or to help users clarify their needs by asking appropriate questions directly to the users. The system has to be capable of asking aspect-based questions in the right order so as to understand the user needs, while search is conducted during the conversation, and results are provided when the system feels confident [48]. Virtual assistants interact with the user in a simplified dialogue to perform a task, support interfaces that adapt to the user's queries, and personal agents that can proactively support the user, modelling his/her needs. Nowadays, with the advent of virtual assistants, there is a proliferating demand for technology in various applications including banking, financial services and insurance, automotive, IT & telecommunications, retail, healthcare, education and others. Recently, it has been reported that about one in six physicians in the European Union are already using virtual assistants [31]. It is clear that virtual assistants have now entered people's daily life to accomplish tasks.

Based on the product, the virtual assistant market has been segmented into Chatbots and smart speakers. A Chatbot is a computer program that carries out a conversation through, whereas smart speakers are a type of wireless speakers and voice command devices. Virtual assistants try to interact with human-like responses that are reasonable or interesting [41]. Informational virtual assistants try to help users find information or directly answer user questions. Task oriented virtual assistants try to help users finish a specific task, such as booking a flight or cancelling a trip. Virtual assistants are usually built for a specific domain, such as music, books, movies, and so on. A recent report shows how virtual assistants are currently used by their owners in the US, UK, France and Germany; in particular, 82% of the virtual assistant owners in these countries use virtual assistants to seek assorted information such as news, weather, recipes, appointments, advice, offers and so on [32]. The Google's Virtual Assistant is primarily available on mobile and smart home devices, and can engage in two-way conversations. Users primarily interact with the Google's Virtual Assistant through natural voice, though keyboard input is also supported. The Google's Virtual Assistant is able to search the Internet, schedule events and alarms, adjust hardware settings on the user's device, and show information from the user's Google account. Google has recently announced that the Google's Virtual Assistant will be able to identify objects and gather visual information through the device's camera, and support purchasing products and sending money, as well as identifying songs. In a similar spirit, Apple's Siri is a virtual assistant which uses voice queries and a natural-language user interface to answer questions, and performs actions by delegating requests to a set of Internet services. The software adapts to users' individual language usages, searches, and preferences, with continuing use. Finally, the returned results are individualized. Alexa is a virtual assistant developed by Amazon. It is capable of voice interaction, music playback, making to-do lists, setting alarms, streaming podcasts, playing audiobooks, and providing weather, traffic, sports, and other real-time information, such as news. Microsoft Cortana is a virtual assistant that can set reminders, recognize natural voice without the requirement of keyboard input, and answer questions using information from the Bing search engine. Microsoft recently reported that Cortana has 133 million monthly users [33], estimating that 325.8 million people per month will use any type of virtual assistants worldwide [34, 35]. However, compared to recommendation systems virtual assistants are designed to complete certain tasks and do not capture users' personal, evolving and multi-aspect preferences.

In addition, health virtual assistants have also been designed, such as PocketSkills which supports dialectical behavioural therapy, aiming to decrease depression and anxiety trough conversations [39]. In a similar spirit, Woebot is a Facebook chatbot developed by Stanford University researchers that offers interactive cognitive behavioral therapy, helping people with dipressia [36]. Chatbots are usually programs that are meant to have conversations with users via text or speech methods. They are meant for specific tasks in various companies and sometimes for general chit-chat purposes. They are subset or parts of AI bots rather than being complete virtual assistants [11]. Compared to Chatbots, virtual assistants are built based on complex algorithms of Natural Language Processing (NLP), Machine Learning, and Artificial Neural Networks (ANNs), learning throughout their usage and have better performance, whereas Chatbots are based on fixed rules which cannot be further modified.

With the emerging of various conversational devices, and the progress of deep learning and neural NLP research, especially on natural language dialog systems, virtual assistants based on direct user-system dialoguing has gained attention by the academia as well [5, 38, 42, 49]. For example, Zhao et al. [49] was among the first works of building an end-to-end conversational system. Dhingra et al. [5] built a goal oriented information access system based on reinforcement learning, trying to select related items with certain attribute values. Spina and Trippas [40, 42] studied the ways of presenting search results over speech-only channels and transcribing the spoken search recordings to support conversational search via deep learning. Finally, Kang et al. [14] explored the initial and follow-up queries users tend to issue to virtual assistants. However, most of those deep learning strategies for virtual assistants focus on NLP challenges instead of recommendation systems. They neither focus on recommendation problems nor do they model and utilize users' preferences to generate recommendations via user-system conversations.

4 THE TECHNOLOGICAL GAP BETWEEN VIRTUAL ASSISTANTS AND RECOMMENDATION SYSTEMS

Academic research in recommendation systems is largely focused on algorithmic approaches for item selection and ranking, trying to predict the ratings or generate ranked lists. However, presenting an ordered list of recommendations might not be the most suitable mechanism to support users in a decision-making problem, for example, when the user needs to clarify and refine his/her preferences. To achieve this, more interactive and possibly complex systems are required, so that users can fine-tune their profiles to provide the system with a richer repertoire of "conversational moves". Virtual assistants could solve this problem as users discuss with the system and enable more interactive recommendation systems without complex interfaces while at the same time providing more accurate recommendations. For instance, you are considering to watch a movie but you are not sure you would enjoy it; then, you would ask your friends for advice. Alternatively, imagine that an acquaintance recommends a movie that you do not think you would enjoy. In the latter case, you would be the one willing to provide information to help your friend make better recommendations in the future. Current recommendation systems do not allow this type of interactive process to occur between the system and its users, whereas virtual assistants are typically oriented towards executing standalone commands rather than complex conversations. On the one hand, a plethora of personal virtual assistants have started to arise in a variety of products across domains, ranging from entertainment or retail bots to health virtual assistants. However, virtual assistants are powered by recent advances in natural language understanding and focus on conversations, not on recommendations. On the other hand, conversations in recommendation systems have to focus on balancing the user's explore-exploit trade-off.

4.1 Information Need

The central difference of virtual assistants and recommendation systems is the representation of the information need: while virtual assistants, as Information Retrieval systems, typically use an explicit query prompted by the user, recommendation systems exploit user's data in an implicit manner [44]. In contrast to existing virtual assistants, recommendation systems have not only to generate accurate recommendations, but novel ones to surprise users and trigger their interest, covering users' diverse tastes and making it easier for them to understand which alternatives exist [10].

4.2 Scarcity of Users' Preferences

When virtual assistants seek information, they rely on a large amount of labelled data, which may not be available in real-world applications, such as users' preferences in recommendation systems [20, 22]. The scarcity of users' preferences has a negative impact on the quality of recommendations of collaborative filtering models, a mainstay strategy in recommendation systems that has been widely adopted in the past. On the contrary, virtual assistants do not account the user data scarcity. Recently, in recommendation systems several deep learning strategies have been introduced to solve the data scarcity of users' preferences [27]. In addition, to overcome the scarcity problem of users' preferences, several models exploit the selections of social friends to generate trust-based recommendations [23, 24, 26, 28]. However, the challenge is that we have to learn both about users' preferences and trust degrees, as friends do not have necessarily the same preferences, a key factor that is currently ignored by virtual assistants to produce accurate recommendations.

4.3 Adaptation to Evolving Preferences

In recommendation systems users shift their preferences over time, depending on different factors [29, 30]. For example, curiosity leads users to explore new items contrary to their ordinary choices and/or users interact with a bias based on popularity irrespective to their history record. If a user has a pleasant experience in the past, then probably s/he will choose the same or a similar interaction in the future, or users become more and more familiar with items they interact and thus gain meaningful experience, which means that they become skeptical and ungenerous while interacting with similar items to the ones that they did in the past. In addition, accounting for the order of users' selections, sequential recommendation systems, also known as session-based recommendation systems, model and learn the sequence of users' selections/sessions to generate next-item recommendation [21]. Despite that users' dynamic preferences and sequential selections are important for producing quality recommendations, users' evolving and sequential preferences are not yet considered by virtual assistants.

4.4 Adjustment to Cross-domain Recommendation Tasks

In recommendation systems, users can express their preference on items from different domains, such as books and music, or the same users express their opinion on different social media platforms such as Facebook and Twitter. While virtual assistants are designed to complete specific tasks in users' daily life, the goal of recommendation systems is also to transfer the knowledge of users across different domains/tasks, also known as cross-domain recommendation systems. The challenge in cross-domain recommendation systems is to capture users' multi-aspect behaviours when transferring knowledge and generating recommendations for various domains [1, 25]. In particular, we have to carefully transfer the knowledge of user preferences from one domain to another by handling (weighting) their different behaviors and making suggestions accordingly. Adjustment to different domains based on users' preferences is a key factor that is currently missing from virtual assistants which focus only on a specific domain.

4.5 Transparency and Explainability

Another important difference between recommendation systems and virtual assistants is that to build trust between recommendation systems and users, it has become important to complement recommendations with explanations so that users can understand why a particular item has been suggested [17]. Transparent and explainable explanations help convincing users that the system knows them very well and makes custom-made recommendations for them. In fact, when users understand the recommendation logic, they can even be empowered to correct the system's proposals. This means that recommendations without context lack motivation for a user to pay attention to them. Adding an associated explanation for a recommendation increases user satisfaction and the persuasiveness of recommendations [18]. Nonetheless, until now, virtual assistants do not provide explanations to users.

4.6 User and Item Biases

In recommendation engines, users' preferences can be significantly distorted by the systems' predicted items that are displayed to users [3]. Such distorted preferences are subsequently submitted as users' feedback to recommendation systems, which can potentially lead to a biased view of consumer preferences and several potential problems: biases can contaminate the recommendation system's inputs, weakening the system's ability to provide high-quality recommendations in subsequent iterations. In addition, biases can artificially pull consumers' preferences towards displayed system recommendations, providing a distorted view of the system's performance. Biases can also lead to a distorted view of items from the users' perspectives. Thus, when using recommendation systems, biases can be harmful to system's use and value, and the removal of biases from consumer preferences constitutes an important and highly practical research problem. Moreover, user preferences are known to be exposed to external and internal influence and bias factors, such as mass media, marketing, opinion management, or social conformity. Notably, user and item biases are not considered in current implementations of virtual assistants. An open question remains how to consider user and item biases in recommendation systems, when users interact with them via conversations?

4.7 Preference Elicitation

Recommendation systems focus on capturing users' feedback in different ways, whereas virtual assistants pay attention to different conversational strategies. This section focus on the main different technological aspects to elicit user preferences and conversation strategies in recommendation systems and virtual assistants, respectively.

4.7.1 Capture users' multi-channel feedback in recommendation systems. In recommendation systems, there are several ways to state user preferences, without involving conversations. For example, users are requested either to rate the items on a predefined scale, as well as to add comments. Other recommendation systems limit the feedback scale to "thumbs up/down" or positive only "like" statements. Users might be also requested to name a few favorite artists, movies, books, social events, and points-of-interest, to specify their interests in different categories such as "Entertainment", "Politics", "Sports", and so on. While these types of user feedback are expressed in an absolute manner, relevant studies point out that pairwise preferences are important in recommendation systems, as pairwise preferences naturally arise and are expressed by users in many decision making scenarios [12]. In everyday life, there are situations where rating alternative options is not the most natural mechanism for expressing preferences and making decisions, for instance, we do not rate sweaters when we want to buy one [13]. Pairwise preferences are a pivotal issue in designing effective recommendation systems, as they can lead to larger system usability compared to absolute preferences. Although user-system conversations are the main type of user feedback when using virtual assistants, the multi-channel user feedback in recommendation systems is currently missing from virtual assistants to better model and capture human behavior and generate accurate recommendations.

4.7.2 Users' conversation strategies in virtual assistants. In virtual assistants to initiate users' conversations with the system, we have to design a Conversation Manager. Users in the speaking condition start a dialogue with the system by speaking at their computer or device, whereas users being in the typing condition by typing into an input box. For the speaking interface, we have to support a voice-to-speech service such as [37], to convert the audio to text. In addition, we have also to allow users to view the results and to retry or edit the results manually if the transcription results in errors, or if their microphone is not working. In general, instead of asking users to provide all requirements in one step, the Conversation Manager usually guides users through an interactive dialog, following different strategies of follow-up queries based on users' satisfaction of the results. Therefore, it is required to investigate various Conversation Manager policies to select what questions to ask and how ask, to minimize the human effort, that is the length of questioning-answering response, and emphasize on capturing user preferences based on the personality and interests. By actively asking questions, conversational recommendation has the advantage of understanding user needs more efficiently. However, this adds challenges related to asking the right question at the right time, as well as inferring user preferences from unstructured responses automatically, issues that current recommendation strategies have still to consider. Also, in the context of conversation strategies, a crucial role is the right choice of the medium that users interact with the system via conversations. For example, Yang et al. [47] performed an online experiment to study users' interactions with podcast virtual assistants with recommendations delivered visually to those with recommendations delivered vocally, where they found that when recommendations are vocally conveyed, users consume more slowly, explore less, and choose fewer long-tail items. This means that indeed the medium affects users' satisfaction and the recommendation engine has to be designed accordingly to produce recommendations via user-system conversations with less human effort.

4.8 Quality Metrics

Quality Information Retrieval metrics that proved successful in recommendation systems are weak indicators to evaluate the recommendation performance of a system with conversations in real-time. The quality of conversational recommendations should be evaluated in terms of number of recommendations that a user will choose expressing the level of a participant's satisfaction; the ranking performance of the recommendation mechanism such as Normalized Discounting Cumulative Gain, Precision and Recall; the dwelling time, that is the amount of time that a participant spends before choosing a recommendation [45]. Also the goal must be to minimize the number of questions asked to obtain users' selections and consequently minimize the user time and effort. In particular, we have to determine which aspect to ask at each time with a carefully trained strategy, so that the system can always ask the most important question to improve its confidence about user needs and search results, thus keeping the conversation as short as possible, and satisfy the user needs as soon as possible. Furthermore, measuring the semantic coherence of conversations of users is also a key performance indicator. To be able to provide intelligent responses, the system must correctly model the structure and semantics of a conversation, as it is also pointed out at [43]. Thus, it is required to design numeric scores that indicate more coherent parts of a conversation and provide a signal for topic drift. For example, one possible way to achieve this is by applying the Word2Vec strategy to create the textual embeddings of the user-system conversations and measure their coherence based on the respective embeddings' similarities [2]. Alternatively, collaborative topic modeling strategies could be followed [46]. Note that making coherent human-system conversations, requires both to minimize the human effort and maximize the recommendation accuracy. A key factor to evaluate a conversational recommendation system is to perform A/B online testing, allowing the comparison of different conversation policies to achieve the best balance between high recommendation accuracy and less human effort.

5 CONCLUSION

Summarizing, there are still many technological gaps between recommendation systems and virtual assistants that researchers have to account when designing a conversational system, that is learning users' evolving, diverse and multi-aspect preferences via human-computer conversations. The dream of having a really artificial friend to make suggestions is not far anymore.

However, to fulfill this dream, researchers have to answer the following fundamental questions:

- *How can user preferences via conversations be modelled into machine learning models in recommendation systems?*
- *Which are the right junctions to perform cross-domain recommendation with machine learning models?*
- *To what extent can we provide explainable recommendations via conversations?*
- *How can we consider user and item biases in conversational recommendation systems?*
- *Which are reliable indicators to evaluate the quality of recommendation systems via conversations in real-time?*

Filling the technological gap between recommendation systems and virtual assistants could help in building a system that is more conversational to allow users to "work together" (collaborate) to improve the quality of recommendations and user experience.

REFERENCES

[1] Mohammad Aliannejadi, Dimitrios Rafailidis, and Fabio Crestani. 2017. Personalized Keyword Boosting for Venue Suggestion Based on Multiple LBSNs. In *Proceedings of 39th European Conference on IR Research, ECIR 2017, Aberdeen, UK.* 291–303.

[2] Tomas Mikolov andIlya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings 27th Annual Conference on Neural Information Processing Systems.* Lake Tahoe, NV, 3111–3119.

[3] Rocío Cañamares and Pablo Castells. 2018. Should I Follow the Crowd?: A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems. In *Proceedings 41st ACM International Conference on Research & Development in Information Retrieval (SIGIR).* Ann Arbor, MI, 415–424.

[4] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. 2010. The YouTube Video Recommendation System. In *Proceedings 4th ACM Conference on Recommender Systems (RecSys).* Barcelona, Spain, 293–296.

[5] Bhuwan Dhingra, Lihong Li, Xiujun Li, Jianfeng Gao, Yun-Nung Chen, Faisal Ahmed, and Li Deng. 2017. Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access. In *Proceedings 55th Annual Meeting of the Association for Computational Linguistics (ACL),* Vol. I. Vancouver, Canada, 484–495.

[6] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. In *Proceedings 24th International Conference on World Wide Web (WWW)*. Florence, Italy, 278–288.

[7] Mehmet H. Göker, Pat Langley, and Cynthia A. Thompson. 2011. A Personalized System for Conversational Recommendations. *CoRR* abs/1107.0029 (2011).

[8] Mehmet H. Göker and Cynthia A. Thompson. 2000. Personalized Conversational Case-Based Recommendation. In *Proceedings 5th European Workshop on Advances in Case-Based Reasoning (EWCBR)*. Trento, Italy, 99–111.

[9] Carlos A. Gomez-Uribe and Neil Hunt. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems* 6, 4 (2016), 13:1–13:19.

[10] Asela Gunawardana and Guy Shani. 2015. Evaluating Recommender Systems. In *Recommender Systems Handbook*. Springer, 265–308.

[11] Mahipal Jadeja and Neelanshi Varia. 2017. Perspectives for Evaluating Conversational AI. *CoRR* abs/1709.04734 (2017). arXiv:1709.04734 http://arxiv.org/abs/1709.04734

[12] Saikishore Kalloori, Francesco Ricci, and Rosella Gennari. 2018. Eliciting Pairwise Preferences in Recommender Systems. In *Proceedings 12th ACM Conference on Recommender Systems (RecSys)*. Vancouver, Canada, 329–337.

[13] Saikishore Kalloori, Francesco Ricci, and Marko Tkalcic. 2016. Pairwise Preferences Based Matrix Factorization and Nearest Neighbor Recommendation Techniques. In *Proceedings 10th ACM Conference on Recommender Systems (RecSys)*. Boston, MA, 143–146.

[14] Jie Kang, Kyle Condiff, Shuo Chang, Joseph A. Konstan, Loren G. Terveen, and F. Maxwell Harper. 2017. Understanding How People Use Natural Language to Ask for Recommendations. In *Proceedings 11th ACM Conference on Recommender Systems (RecSys)*. Como, Italy, 229–237.

[15] Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer* 42, 8 (2009), 30–37.

[16] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2011. Content-based Recommender Systems: State of the Art and Trends. In *Recommender Systems Handbook*. Springer, 73–105.

[17] Julian J. McAuley and Jure Leskovec. 2013. Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text. In *Proceedings 7th ACM Conference on Recommender Systems (RecSys)*. Hong Kong, China, 165–172.

[18] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. 2018. Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits. In *Proceedings 12th ACM Conference on Recommender Systems (RecSys)*. Vancouver, Canada, 31–39.

[19] Alexandros Nanopoulos, Dimitrios Rafailidis, Panagiotis Symeonidis, and Yannis Manolopoulos. 2010. MusicBox: Personalized Music Recommendation Based on Cubic Analysis of Social Tags. *IEEE Transactions on Audio, Speech & Language Processing* 18, 2 (2010), 407–412.

[20] Minghui Qiu, Liu Yang, Feng Ji, Wei Zhou, Jun Huang, Haiqing Chen, W. Bruce Croft, and Wei Lin. 2018. Transfer Learning for Context-aware Question Matching in Information-seeking Conversations in E-commerce. In *Proceedings 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vol. II. Melbourne, Australia, 208–213.

[21] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks. In *Proceedings 11th ACM Conference on Recommender Systems (RecSys)*. Como, Italy, 130–137.

[22] Dimitrios Rafailidis, Apostolos Axenopoulos, Jonas Etzold, Stavroula Manolopoulou, and Petros Daras. 2014. Content-based tag propagation and tensor factorization for personalized item recommendation based on social tagging. *TiiS* 3, 4 (2014), 26:1–26:27.

[23] Dimitrios Rafailidis and Fabio Crestani. 2016. Collaborative Ranking with Social Relationships for Top-N Recommendations. In *Proceedings of 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy*. 785–788.

[24] Dimitrios Rafailidis and Fabio Crestani. 2016. Joint Collaborative Ranking with Social Relationships in Top-N Recommendation. In *Proceedings of ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA*. 1393–1402.

[25] Dimitrios Rafailidis and Fabio Crestani. 2016. Top-N Recommendation via Joint Cross-domain User Clustering and Similarity Learning. In *Proceedings European Conference on Machine Learning & Knowledge Discovery in Databases (ECML/PKDD)*, Vol. II. Riva del Garda, Italy, 426–441.

[26] Dimitrios Rafailidis and Fabio Crestani. 2017. Learning to Rank with Trust and Distrust in Recommender Systems. In *Proceedings of the 11th ACM Conference on Recommender Systems, RecSys 2017, Como, Italy*. 5–13.

[27] Dimitrios Rafailidis and Fabio Crestani. 2017. Recommendation with Social Relationships via Deep Learning. In *Proceedings 7th ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR)*. Amsterdam, The Netherlands, 151–158.

[28] Dimitrios Rafailidis and Fabio Crestani. 2017. A Regularization Method with Inference of Trust and Distrust in Recommender Systems. In *Proceedings European Conference on Machine Learning & Knowledge Discovery in Databases (ECML/PKDD)*, Vol. II. Skopje, FYROM, 253–268.

[29] Dimitrios Rafailidis, Pavlos Kefalas, and Yannis Manolopoulos. 2017. Preference dynamics with multimodal user-item interactions in social media recommendation. *Expert Syst. Appl.* 74 (2017), 11–18.

[30] Dimitrios Rafailidis and Alexandros Nanopoulos. 2015. Repeat Consumption Recommendation Based on Users Preference Dynamics and Side Information. In *Companion Proceedings 24th International Conference on World Wide Web (WWW)*. Florence, Italy, 99–100.

[31] Rep1 2018. [Online]. Available: https://www.drgdigital.com/drg-digital-innovation-blog/hallo-siri-voice-assistants/. (2018).

[32] Rep2 2018. [Online]. Available: https://www.statista.com/statistics. (2018).

[33] Rep3 2018. [Online]. Available: https://www.techradar.com/. (2018).

[34] Rep4 2018. [Online]. Available: https://aytm.com/blog/ai-personal-assistants/. (2018).

[35] Rep5 2018. [Online]. Available: http://www.internetlivestats.com/internet-users/. (2018).

[36] Rep6 2018. [Online]. Available: https://www.woebot.io/. (2018).

[37] Rep6 2018. [Online]. Available: https://wit.ai/. (2018).

[38] Lina Maria Rojas-Barahona, Milica Gasic, Nikola Mrksic, Pei-Hao Su, Stefan Ultes, Tsung-Hsien Wen, Steve J. Young, and David Vandyke. 2017. A Network-based End-to-End Trainable Task-oriented Dialogue System. In *Proceedings 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Vol. I. Valencia, Spain, 438–449.

[39] Jessica Schroeder, Chelsey Wilkes, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M. Linehan. 2018. Pocket Skills: A Conversational Mobile Web App to Support Dialectical Behavioral Therapy. In *Proceedings Conference on Human Factors in Computing Systems (CHI)*. Montreal, Canada, 398.

[40] Damiano Spina, Johanne R. Trippas, Lawrence Cavedon, and Mark Sanderson. 2017. Extracting Audio Summaries to Support Effective Spoken Document Search. *Journal of the Association for Information Science & Technology* 68, 9 (2017), 2101–2115.

[41] Yueming Sun and Yi Zhang. 2018. Conversational Recommender System. In *Proceedings 41st International ACM Conference on Research & Development in Information Retrieval (SIGIR)*. Ann Arbor, MI, 235–244.

[42] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. 2017. How do People Interact in Conversational Speech-only Search Tasks: A Preliminary Analysis. In *Proceedings Conference on Conference Human Information Interaction & Retrieval (CHIIR)*. Oslo, Norway, 325–328.

[43] Svitlana Vakulenko, Maarten de Rijke, Michael Cochez, Vadim Savenkov, and Axel Polleres. 2018. Measuring Semantic Coherence of a Conversation. In *Proceedings 17th International Semantic Web Conference (ISWC)*, Vol. I. Monterey, CA, 634–651.

[44] Daniel Valcarce, Alejandro Bellogín, Javier Parapar, and Pablo Castells. 2018. On the Robustness and Discriminative Power of Information Retrieval Metrics for Top-N Recommendation. In *Proceedings 12th ACM Conference on Recommender Systems (RecSys)*. Vancouver, Canada, 260–268.

[45] Marisa Vasconcelos, Heloisa Candello, Claudio S. Pinhanez, and Thiago dos Santos. 2017. Bottester: Testing Conversational Systems with Simulated Users. In *Proceedings XVI Brazilian Symposium on Human Factors in Computing Systems (IHC)*. Joinville, Brazil, 73:1–73:4.

[46] Chong Wang and David M. Blei. 2011. Collaborative Topic Modeling for Recommending Scientific Articles. In *Proceedings 17th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*. San Diego, CA, 448–456.

[47] Longqi Yang, Michael Sobolev, Christina Tsangouri, and Deborah Estrin. 2018. Understanding User Interactions with Podcast Recommendations Delivered via Voice. In *Proceedings 12th ACM Conference on Recommender Systems (RecSys)*. Vancouver, Canada, 190–194.

[48] Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. 2018. Towards Conversational Search and Recommendation: System Ask, User Respond. In *Proceedings 27th ACM International Conference on Information & Knowledge Management (CIKM)*. Torino, Italy, 177–186.

[49] Tiancheng Zhao and Maxine Eskénazi. 2016. Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning. *CoRR* abs/1606.02560 (2016).