Does the User Have A Theory of the Recommender? A Pilot Study

Muheeb Faizan Ghori*
School of Computing
DePaul University
Chicago, Illinois
mghori2@depaul.edu

Arman Dehpanah
School of Computing
DePaul University
Chicago, Illinois
adehpana@depaul.edu

Jonathan Gemmell
School of Computing
DePaul University
Chicago, Illinois
jgemmell@cdm.depaul.edu

Hamed Qahri-Saremi
School of Computing
DePaul University
Chicago, Illinois
hamed.saremi@depaul.edu

Bamshad Mobasher
School of Computing
DePaul University
Chicago, Illinois
mobasher@depaul.edu

ABSTRACT

Recommender systems have become a mainstay of modern internet applications. They help users identify products to purchase on Amazon, movies to watch on Netflix and songs to enjoy on Pandora. Indeed, they have become so commonplace that users, through years of interactions with these systems, have developed an inherent understanding of how recommender systems function, what their objectives are, and how the user might manipulate them. We describe this understanding as the *Theory of the Recommender*. In this pilot study, we design and administer a survey to 25 users familiar with recommender systems. Our detailed analysis of their responses demonstrates that they possess an awareness of how recommender systems profile the user, build representations for items, and ultimately construct recommendations. The success of this pilot study provides support for a larger user study and the development of a grounded theory to describe the user's cognitive model of how recommender systems function.

CCS CONCEPTS

• **Human-centered computing** → **User models**; • **Retrieval tasks and goals** → *Information extraction*; • **Computing methodologies** → *Cognitive science.*

KEYWORDS

recommender systems, cognitive models, qualitative research

ACM Reference Format:
Muheeb Faizan Ghori, Arman Dehpanah, Jonathan Gemmell, Hamed Qahri-Saremi, and Bamshad Mobasher. 2019. Does the User Have A Theory of the Recommender? A Pilot Study. In Proceedings of Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS '19), CEUR-WS.org, 9 pages.

IntRS Workshop, September, 2019, Copenhagen, Dk

1 INTRODUCTION

Recommender systems help users find items in large and complex information spaces such as those found in online retailers or streaming services. These systems have become an integral tool in modern internet applications helping users cope with the increasing complexity of online environments and improving the company's competitive edge. Recommender systems often leverage user information and interactions with the system to provide personalized recommendations that satisfy the needs and preferences of the user.

These systems use several techniques to generate recommendations. Common approaches include collaborative filtering [16], content-based filtering [24], and model-based methods [2].

In the past decade, these technologies have become ubiquitous. Consequently, modern internet users are exposed to recommender systems on a daily basis. They view recommended items, consume items that catch their interest, and perhaps rate or leave feedback about these items. These repeated interactions may have led to an inherent understanding of how recommender systems work.

In this paper, we ask the question: Does the user possess a theory of the recommender? The title of this paper is inspired by Premack and Woodruff's seminal paper [28] asking if chimpanzees understand the goals, perceptions, knowledge, and beliefs of others. In a similar vein, we want to ascertain if users understand the goals, perceptions, knowledge, and beliefs of recommender systems. We hypothesize that as recommender systems have become more commonplace and sophisticated, so too has the user's understanding of the recommenders.

A study into the user's understanding of recommender systems is critical for many reasons, among them: 1) the development of a framework for understanding the user's cognitive model of how recommender systems work, 2) predicting what behaviors such a cognitive model would elicit, and 3) designing systems that can identify and leverage these behaviors, thereby increasing the performance and value of recommender systems.

To test our hypothesis, we design and conduct a user study to identify concepts related to the participant's understanding of recommenders given several common scenarios. The identification of these concepts in this pilot study is a first step toward developing a robust grounded theory, describing the user's cognitive model.

Grounded theory allows the construction of theories that are grounded in empirical observations or data. It uses a constant comparative method where each observation is compared with others to find similarities and differences to generate concepts, hypothesis, and relationships to explain behavior and processes.

As a first step toward developing a rich grounded theory, we design and administer a survey instrument. The instrument consists of several questions probing the subject's understanding of recommender systems. Users are presented with scenarios typically encountered at online retailers, streaming services and news aggregators. They are then asked to answer questions based on their knowledge and personal experience. Our primary measure of user perception is the response to questions such as "Explain how you think the system recommended this item for the user?"

We administered the survey instrument to 25 participants in a pilot study through Amazon's Mechanical Turk[1]. Our results show that the participants appear to possess a cognitive model of recommender systems. For example, while the participants did not use terms such as "user representation" or "collaborative filtering", they often describe a system that "keeps track of the user's purchases" and "recommends items similar to those she had purchased before."

The results of this pilot study will play an important role in improving the survey instrument and evaluating the design, feasibility, cost and time of conducting a larger more thorough study with the goal of using grounded theory to describe the user's cognitive model of recommender systems. Future work will address the limitations revealed by this study and will explore how users modify their behavior based on their understanding of recommender systems, what impact such behavior might have on these systems, and how recommender systems might be designed to cope with, or even leverage, these behaviors.

The rest of this paper is organized as follows. In Section 2 we present our related work. A brief summary of common recommendation algorithms is offered, and established techniques for conducting surveys are provided. In Section 3 our methodology is described in detail. Section 4 presents the survey instrument. The results of the survey are described in Section 5. Here we provide reoccurring patterns observed in the survey responses, a deliberation of important limitations and directions for future work before turning our attention to a discussion of the concepts expressed by the survey participants. Finally, in Section 6 we offer our conclusions.

2 RELATED WORK

Traditional approaches for designing a recommender system include collaborative filtering, content-based filtering, and model-based methods. Collaborative filtering is often divided into two separate approaches. User-based collaborative filtering [16] identifies similar users to the target user and recommends items that those neighbors have consumed. Item-based collaborative filtering generates recommendations by finding items similar to those previously selected by the user [32]. In general, these recommenders are based on the idea that users who agreed in the past are likely to agree in the future [30].

[1]www.mturk.com

On the other hand, content-based filtering systems [6, 24] learn a profile for each user based on features of the items previously consumed, such as the genre of a book or actor in a movie. Similarly, item profiles are created by characterizing the items based on their attributes and features. The system generates recommendations by comparing the description of new items to the user's preference profile. For example, it may recommend a new sci-fi movie to a user that had been identified as a fan of science fiction.

In contrast, model-based methods [2] train a model for each user based on prior user preferences. Several model-based approaches exist, but in general, the goal is to predict the likelihood of new items to be of interest to the target user. Examples include Matrix Factorization [19] and Singular Value Decomposition [3].

We speculate that as recommender systems have become more commonplace and popular, users may have developed a basic understanding of how they work. We liken this understanding to the Theory of Mind [28]. Theory of Mind is the cognitive capacity to perceive and predict other people's behavior in terms of their mental states.

Frith et al. [13] explained that the behavior of people can be understood on the basis of their minds: their knowledge, beliefs, desires, and intentions. Moreover, people engaged in social life attribute various knowledge, beliefs, desires, and intentions to others [15]. Such attributions are useful in analyzing, judging, and inferring other person's behavior. This ability is a fundamental aspect of social cognition that guides an individual's behavior in a society.

In order to understand the user's perception of how recommender systems work, what we refer to as a Theory of the Recommender, we are proposing the development of a grounded theory. Developed by Glaser and Strauss, 'grounded theory is a general methodology for developing a theory that is grounded in data systematically gathered and analyzed' [14]. This approach uses a constant comparative method where each observation is compared with others to find similarities and differences to generate concepts, hypothesis, and relationships that best explains behavior and processes [7].

Grounded theory methods allow elaboration and modification of existing theories or generation of new theories from the collected data [10]. This method is iterative. The emerging theory is incrementally refined based on recurring data collection and analysis. This approach of theory generation on collected data is well suited for this work since our aim is to model the user's perception and beliefs about recommender systems.

Previous work at the intersection of recommender systems and human-computer interaction has focused mainly on enhancing the quality of user experience and interaction with the recommender system. Pu and Chen [29] conducted a user study to show that the quality of user experience and interaction with the system is crucial for recommendation performance. They provided a framework named ResQue (Recommender systems' Quality of user experience) to measure a recommender system's overall perceptive qualities and effectiveness in influencing users' behavioral intentions.

Kulesza et al. [20] showed how a user's mental model of the recommender system can be used to debug and personalize an intelligent agent. In a similar study, the authors explained that the soundness and completeness of explanations when presenting recommendations can impact the end user's mental model of the recommender system [21].

Bonhard and Sasse [8] used a grounded theory methodology to demonstrate the impact of similarity and familiarity visualizations. People prefer recommendations from people they know. Presenting the familiarity and similarity between the user and the people who have rated the items aids them in their decision-making process.

Arazy et al. [5] argue for a system that is grounded in theories of human behavior. They proposed a methodology to apply theory-driven design to social recommender systems with the aim of improving prediction accuracy. They suggest using behavioral theories to guide information system design.

Our work relies on these previous efforts in several ways. First, we must understand the mechanics of recommender systems and how users might interpret them. Second, we must use a well-established tool, in this case, qualitative study, to model the user's understanding of recommender systems. In this pilot study, we do not go as far as building a grounded theory; but, present evidence that such a model is feasible and take the first steps toward developing it. Third, we take inspiration from the Theory of Mind and previous efforts in social psychology to explore the question of whether or not users possess a theory of the recommender.

3 METHODOLOGY

In this work, we seek to determine if today's internet users understand how recommender systems function. To that end, we present our methodology including the design of our survey instrument, interpretation of the results, and preparation for a larger study.

The initial survey was designed after conducting an extensive literature review to identify fundamental aspects of recommender systems [12]. The questions were based on common scenarios users often experience when interacting with recommender systems. In particular, domains for survey questions were inspired from research by Adomavicius and Tuzhilin on the survey of state of the art and possible extensions of recommender systems [3] and the recommender systems handbook [31] by Ricci et al. Adomavicius and Tuzhilin in their survey describes the three popular types of recommender systems namely collaborative filtering [18], content-based filtering [11], model-based methods [2] and explains the new user problem [33]. Similarly, Ricci et al. delineates the basic recommender system ideas and concepts such as user representation [11], item representation [11] and goals of the recommender system [31]. Consequently, questions were designed to probe the participant's understanding of similar recommender system concepts namely user models, item models, similarity measures, the cold start problem, collaborative filtering, and content-based models.

We relied on open-ended questions to allow participants to provide in-depth accounts of their experience and understanding of the system based on their interactions with the system. To ensure the relevance of participant's answers, we provide them with information about the context of the study (i.e., recommendation systems), while being cautious not to bias their responses.

The survey instrument was evaluated by a panel of three domain experts in recommender systems to establish content validity [9, 23]. Similarly, discussions with students helped establish face validity of the survey [23]. In this instance, content validity describes the extent to which a survey instrument fairly represents the domain the instrument seeks to measure, in this case, the domain of online recommender systems. Face validity, on the other hand, assesses whether the survey is comprehensible to the participants or other technically untrained observers.

After receiving feedback the survey instrument was revised. The feedback resulted in the removal of some questions, the addition of new questions, and the rewording of others. The instrument was then returned to the panel for more feedback. This iteration continued until a consensus was achieved.

After the survey was approved, it was administered to a pool of participants. Participants were first informed of the purpose of the survey, any risks, the expected time commitment, and the payment information. They were then asked if they wish to participate. Those that wished to continue were asked basic demographic information such as age and gender. At this stage, the system asked their familiarity with recommender systems to ensure they have sufficient exposure to complete the survey. The participants were then presented with the open-ended questions based on common recommender system scenarios.

Once the surveys were completed, we evaluated the responses. Incomplete responses were discarded. The remaining responses were read by three domain experts with a deep understanding of recommender systems. Answers to the open-ended questions were coded and organized based on their association with known recommender system concepts such as similarity functions and user modeling. Disagreements among the coders were resolved by consensus decisions. Quotes from these responses were then organized for presentation.

In this pilot study, we concluded our analysis at this stage and sought to ascertain if there was sufficient evidence to warrant the time and cost of a larger study. In a larger online study, we would continue surveying participants until we observed a saturation of concepts. Coders could then develop a grounded theory representing the user's theory of the recommender.

4 SURVEY

Here we present the survey instrument. First, we presented the user with key information. Second, we ask the subject for basic demographic information such as gender, age, and profession. We then present the subject with five scenarios. These scenarios are meant to capture everyday situations an internet user might encounter such as signing up for a new service, visiting a familiar online retailer, or marking a recommended item as irrelevant.

4.1 Key Information

In order to conform to the standards and practices described by the Institutional Review Board, we presented the participants with several pieces of key information. We provided our contact information. We then described the purpose of this research, before discussing the risk, effort, and remuneration associated with completing the survey. Privacy issues were discussed and we explicitly stated that no personal information that could be used to identify participants would be collected. Participants were then asked if they wished to continue before being administered the survey.

4.2 Demographic Information

The users were asked to provide demographic information such as age, gender, education level and to list services they have used in the past to gauge whether a user is qualified to take the survey. The questions were:

(1) Please provide your gender.
(2) Please select your age group.
(3) Please tell us your profession.
(4) Please provide your highest education level.
(5) List up to three applications, websites or services with a recommender system that you used in the past 1-6 months.
(6) List up to three recommendations that you received from the above stated applications, websites or services.

4.3 Recommender System Scenarios

The open ended questions first described a common scenario a user might encounter when interacting with a recommender system and then asked the user to explain some aspect of the scenario. We selected the questions after a literature review identifying common scenarios users often experience when interacting with recommender systems. These themes were preference elicitation (Q1), goals (Q2), user modeling (Q3a), familiarity with recommenders (Q3b), Content-based filtering (Q3c), demographic information of a new user (Q4), implicit behavior, and (Q5a) historic profiling (Q5b). The questions were:

(1) Robin likes watching movies, TV shows, and occasionally documentaries. His friend recommended him a subscription-based video streaming website. Robin registers for the service and the website asks Robin to select from a list a few movies, TV shows, and documentaries the ones that he likes. Explain with some detail (400 characters) why you think the system asked Robin to select those few items.
(2) Recommender systems provide personalized recommendations to users on a wide variety of platforms such as movies, music, travel, news, and products. Based on your experience interacting with the system on these platforms, list and explain with some detail (400 characters) four goals and intentions of the system.
(3) Sarah is a regular customer at a popular retail website. The website sells many types of items but Sarah usually buys books, electronics and occasionally clothing. She rates her purchases and leaves feedback on items she bought. Answer the following questions.
    (a) Whenever Sarah logs on to the website. She finds a section on the web page with a list of recommended items. Explain with some detail (400 characters) what information the recommender system uses to make those recommendations.
    (b) List a few items the recommender system might recommend Sarah in this section.
    (c) Visiting the site today, Sarah is recommended a book by an author she is familiar with. Explain with some detail (400 characters) how you think the system made that recommendation.
(4) Joe likes listening to music and connects to a music streaming service as a new user by answering several demographic questions including age and gender along with his preference for music (e.g. preference for rock, jazz, and blues). Upon completing the registration, the site recommends several tracks for Joe to play. Explain with some detail (400 characters) how the recommender system made those recommendations for Joe.
(5) Emily is a user of a popular video streaming website that allows users to create profiles (channels) and upload videos on various topics including sports, music, news, and entertainment. Similarly, users who register on the site can subscribe to these channels, search, watch, like, comment, and share other videos.
    (a) Sometimes, Emily dislikes her recommended videos and finds them irrelevant. Explain with some detail (400 characters) why the recommender system might have recommended those items to Emily.
    (b) What difference in the nature of recommendations would have Emily noticed if she used the website as a registered user as compared to using the site as a guest? Explain with some detail (400 characters).

5 SURVEY RESULTS

The survey was administered to 25 participants on Amazon Mechanical Turk [26] using Qualtrics XM Platform for surveys [27]. Mechanical Turk is a crowdsourcing service that connects ‘workers’ and ‘requesters’. Requesters publish ‘human intelligence tasks’, or HITs, and workers complete these tasks online, usually for a small sum. For this HIT, workers were selected based on four criteria. First, their location was limited to the United States in order to avoid language concerns. Second, we limited the participants to those that had completed at least 50 hits. Third, we limited the participants to those that had achieved a 90% acceptance rate on their previous HITs. Forth, we limited the participants to those that had been awarded ‘Master’ status, workers identified by Amazon as maintaining a high standard across a wide range of tasks. These last three criteria were enforced to ensure the quality of the responses. We paid each participant $2.00.

14 of the participants identified as male, 10 as female, and 1 preferred not to answer. The ages of the participants were relatively uniformly distributed from 18 to 60+, with slightly fewer participants in the 60+ range. The participants came from a wide range of professions including a baker, teacher, editor, engineer and business analyst. Nearly all of the participants indicated that they held a bachelor’s degree or higher. When asked what websites or services with a recommender system they had recently used, common answers included Amazon, eBay, YouTube, Quora, Facebook and Netflix.

In general, analysis of the survey shows that the users possess a relatively sophistical understanding of how recommender systems operate. In the remainder of this section, we discuss the recurring concepts expressed in the survey responses. We then present several important limitations of this initial study and how we might overcome these limitations in future work. We conclude this section with a discussion of the extent to which modern internet users possess a theory of the recommender.

5.1 Identification of Concepts

Recommender systems are a multifaceted application. In simplest terms, recommenders attempt to improve the user experience by steering the users toward items they would prefer and away from items they would not. However, there are many strategies employed to accomplish this goal. Moreover, each strategy might have many sub-strategies. There may also be several external goals for which the recommender is tuned. Here we attempt to dissect recommender systems into their fundamental parts and relate them to the recurring concepts identified in the analysis of the responses to the survey instrument.

5.1.1 User Representation. A fundamental concern when designing a recommender system is how to represent the users. In some systems users are represented as a vector space representation over the item space; that is to say they are represented by the collection of items they have consumed, purchased or rated in the past. In our survey, we found several examples of participants describing this type of user representation.

One user described, “user information” explaining that the system uses ratings and reviews provided by the user. Another user stated that the application uses “history” to build a list based on the items browsed or consumed by the user in the past. Perhaps the best example was,

> The information that is used to make recommendations are primarily things that the users have input into the site themselves. Their page views, prior purchases, reviews, likes and dislikes are all taken into account. The system will mainly show the user things that it thinks they will like or want to get more information about - usually similar items to what they have bought in the past.
> - Male, 41-50, Economist

In other systems users are represented by their demographic information such as sex, age, or geographic location. Such an approach works based on the notion that users with similar demographics have similar preferences. The system then uses this information to classify users into pre-existing groups.

Similar to the above concept, participants recognized that the system uses demographic data to recommend items. For example, one user described, “The system looks at user’s age and gender and then compares it to what other users in those demographics tend to prefer”. Another user answered,

> Based on ... [the] user’s identified demographics, the system can build an interest profile. The demographics will help to narrow the scope, as a particular decade may be more likely to be of interest to the user based on their age.
> - Male, 31-40, Business Analyst

5.1.2 Item Representation. Like user representation, item representation is a critical aspect of many recommender systems. For instance, in a movie application, a movie might be represented by a set of features including the genre, actors, and directors.

The survey participants appear to understand the importance of item representation citing specific attributes such as “genre”, “actors” and “authors”. Users describe two item being similar if the same users consumed them. One user described the importance of item representation,

> The system looks at user’s reading habits and trends. It sees that ... the user has read books by this author more frequently in the past. Therefore, it makes sense that the system will recommend a book by this author. If ... the user tends to read a lot of fiction, and this author has written fiction as well as non-fiction, the system probably recommended a work of fiction.
> - Male, 61+, Retired

5.1.3 Collaborative Filtering. Collaborative Filtering models work on the assumption that users who had comparable preferences in the past are likely to have comparable preferences in the future. Based on this idea, collaborative filtering identifies ‘similar’ users to a given target user based on similarity of ratings or items consumed. The system then recommends items rated by those similar users [16].

The survey responses demonstrated users having similar perceptions of the topic. One user described “... Recommendations are based on what the user ... has purchased before, how he rated it, what he browses and what other people who ... bought the same things as ... the user tend to like.” Similarly, another said,

> The system will use customer’s preferences ... to match with those of other users of that streaming service. If other users have the same or similar preferences ..., there’s a good chance that the user will like the same content as those other users. This allows the streaming service to provide relevant recommendations. If the system has enough data from a large number of users, these predictions can be fairly accurate.
> - Male, 21-30

5.1.4 Content-Based Filtering. Content-based recommender system builds a profile for each user based on past preferences and interactions. In general, the system makes recommendations by comparing features of items to the user profile.

We have noticed participants describing the same idea in several different ways. One user mentioned, “similar content” stating that the system recommends similar items to those recently purchased or clicked. Another user inferred that the service likely uses features of items bought in the past explaining “... if a user is looking at books, the system will note the genre and recommend based on that.” One of the best examples was,

> I think the system is looking for a pattern in ... [the] user’s selection of movies. For example, if the user selects some movies with a particular actor, the system is likely to recommend other movies to the user with that actor. For example, let’s say a person listed a movie with Jack Nicholson in it and it was a thriller and of his earlier movies. This alerts the system that the user probably is more interested in Nicholson’s earlier movies than more recent movies, and may very well not be a fan of his comedies.
> - Male, 61+, Retired

5.1.5 Model-Based Approach. Some of the most powerful recommender systems are model-based approaches such as those based on singular value decomposition. These techniques are often more complex than collaborative filtering or content-based approaches. As such, we did not find that the users we able to describe model-based approaches in as much detail as the simpler ones. There were, however, a few examples.

One user stated, "The goal is to build as robust of a profile as possible for users so that it may inspire further purchases." Another user expressed, "... The system is ... gathering information about me in order to create a profile and refine its recommendations that can be used to determine the types of ads and marketing that I would be attracted to, and perhaps respond to."

> The algorithm can see what ... users like and determine, from there, what else ... they might be interested in. I do not know the specifics of how they work, though, so I can't really be that detailed.
> \- Female, 31-40, Editor

It seems this participant understands the goals of the system and knows that recommendations are based on historical preferences. However, she admits to not knowing the details of how these systems make recommendations.

5.1.6 *Association Rule Mining.*

Association rule mining is another useful machine learning technique that helps gain insights into the user's buying habits. Association rules capture relationships among items based on patterns of co-occurrence across transactions [4]. Amazon, for example, uses association rule when it presents a list of items to the user and states,"Customers who viewed this item also viewed these items."

While most users described the idea of recommended items being similar, one user hinted towards the idea of associated products. The user expressed that she is usually recommended gadgets that work with the electronics she previously purchased. Another user described,

> The system may recommend books in similar genres ... the user reads. [The system may recommend] electronics that are accessories or work with [items] ... the user has already bought.
> \- Female, 21-30, Graphic designer

5.1.7 *Goals.*

The fundamental goal of a recommender system is to provide users with relevant information. While identifying relevant information is integral to these systems, recommenders also possess several other motivations that may differ from the viewpoint of consumers and providers. Goals of the recommender system from a consumer's viewpoint include helping users explore the product space, actively notifying users of relevant content, and providing a satisfying experience. At the same time, from the provider's viewpoint, the recommender system's goals may be to steer user behavior in the desired direction, increase revenue, learn consumer habits, and maintain customer loyalty [17].

Participants in our survey reflected a similar understanding of the goals and motivations of the recommender system. One participant expressed, "The goal of the system is to ensure the user has a positive and useful experience with the service." Another user wrote,

> Recommender systems provide recommendations with a goal to satisfy the customer. It helps make decisions for the customer. I think another goal is for the company to make money. Buying recommendations equals profit. Another goal would be to increase user loyalty. And last, the goal of simply bettering the system overall.
> \- Male, 21-30, Sales Representative

5.1.8 *Attitude.*

Attitude is a user's overall feeling towards a recommender system. Users possess different impressions about the system based on their interactions. Examples of positive attitudes include satisfaction, confidence, and trust. Similarly, a few users find the system invading their privacy or too onerous.

When asked about this aspect, one participant answered, "The system ensures the user has a positive and useful experience with the service." On the other hand, another participant revealed,

> The ... user might find that she has generally better recommendations but is losing more of her privacy as a result. They ... would likely find their preferences appear in the form of ads not just on this particular site, but other sites related to or sites that are part of the same conglomerate.
> \- Male, 21-30, Civil Engineer

5.1.9 *Cold Start Problem.*

A crucial challenge for any recommender system is how to recommend items to new users. For a new user, the system lacks the valuable history of the user's interaction with the system on which to base the recommendations [33]. Similarly, new items with few ratings becomes difficult to recommend.

Participants who took our survey conveyed similar interpretations of the concept. One user discussed, "initial recommendations" recognizing that they might differ from those the user would receive later on. Another presumed that the system could overcome a lack of knowledge about the user by using an "enrollment survey" to get a basic idea of their preferences. A third mentioned the use of "user demographics" to recommend items preferred by those within the same age group or location. A user expressed:

> The systems use ... past purchase history to make recommendations, since it's a clear indication of what the user likes to purchase...The suggestions gain strength in accuracy the more a user searches on the site and importantly the suggestions will be very weak or non-existent with a new user. However, if the systems use their demographics (as provided when she signed up on the site), the system can provide recommendations based on her age, sex, and location.
> \- Female, 41-50, Transcriptionist

5.1.10 *Diversity.*

Diversity in a recommendation list is the variation in items being presented to the users [34]. Often there is inherent uncertainty in user's interests, therefore, recommending a variety of items may preserve user interests and avoid disappointment.

We witnessed only a few participants describe the notion of diversity with imprecise details in the survey. One user mentioned,

> The system looks at the item that the user searches and comes up with similar items. The system also

> looks at the things that she has bought in the past and decide to put other items that are close to those on a recommended list in order to give her more of a choice to choose from.
> -Male, 21-30, Engineer

5.1.11 Serendipity. Serendipity is another characteristic of items in a recommendation list that have shown to improve user's overall impression of the service. The important aspects of serendipity are for an item to be relevant, novel, and unexpected [25].

Though inaccurate, survey participants expressed the ideas of "new, interesting and relevant" frequently throughout the survey. One user stated, "The application attempts to provide value, entertainment and interesting content". Another one said, "I believe that the system has ... some form of a baseline understanding of [the user's] interests so that it can make relevant recommendations." A third stated,

> The system is trying to expand ... user engagement with the system, by suggesting new topics for ... them to watch, they may reveal a new set of videos that she will watch on the platform.
> - Male, 31-40, Business Analyst

5.1.12 Context. Context is any information that can be used to describe the situation of an entity; information such as location, time, and season [1] . Users tend to have different preferences under different circumstances. For example, in a movie recommender, a user may prefer a different genre of the movie based on his companion. Incorporating such information in the recommendation process helps personalize recommendations that are relevant to a user's specific context.

A few participants demonstrated an understanding of the use of contextual information in recommendations. A user answered,

> If the system sees that I tend to make purchases more often in the summer than other [seasons], it may recommend more items related to vacations, summer clothing, etc.
> - Male, 60+, Retired

5.2 Limitations

This study is the first in a larger research agenda. Being the first foray into the study of how users understand recommender systems, it does suffer from several limitations. We discuss some of those limitations here and plan to overcome them in the future work.

This survey has two noticeable biases. First, subjects were recruited from Amazon's Mechanical Turk. These individuals are likely more technologically savvy than the average internet user. Second, subjects self selected to complete the survey. The Mechanical Turk volunteers opted to complete this survey after viewing the title, "Tell us Your Experience with Recommender Systems." Workers without an interest in recommender systems may have opted not to complete the survey thereby skewing the sample populations. In retrospect, the strict criteria we placed on the worker's qualifications – number of completed HITs, HIT acceptance rate, etc. – may have exacerbated these biases.

Another limitation of the study is the number of participants. Only 25 Mechanical Turk workers completed the survey. While this was enough to provide subjective evidence that many users maintain a mental model of the recommender system, we cannot claim that we have collected enough responses to have identified all the ways in which users understand recommender systems.

A third limitation stems from the nature of online surveys. They are inflexible. While we were able to ask nearly any question we would like, we could not improve questions based on previous responses.

A forth limitation also stemming from the nature of online surveys is the lack of depth. The survey questions are identical for each participant regardless of their experience, background or education. We were unable to adapt questions or explore answers based on the interaction with the participants as we might have in an in-person interview.

Despite these limitations, this pilot study provides strong subjective evidence that users possess a cognitive model of how recommender systems work. Next, we discuss future work to further explore this research direction while addressing the limitations.

5.3 Future Work

This pilot study was conducted to assess the feasibility, time and cost of a larger more in-depth survey and to improve upon the study design. Here we discuss our plans for future work both in the short term and the long term.

In the short term, we plan to conduct a larger online survey. This survey will take an iterative process. Participants will be selected to complete the survey. Their responses will be coded by multiple coders, identifying key recurring concepts expressed by the participants. Another batch of participants will then be selected to complete the survey. This process will continue until we reach saturation of key concepts. Agreement between the coders would be evaluated to compute the inter-coder reliability [22].

We seek to develop a theory of the user's knowledge and perceptions rather than test any preconceived hypothesis. To this end, we will use a grounded theory methodology to develop a rich and detailed theoretical account of the user's understanding that is purely grounded in observations of their knowledge and personal experience. Consistent with the grounded theory approach, data collection and analysis will be conducted simultaneously allowing emerging concepts to guide the process of further data collection. Several coding schemes will be used to identify emerging concepts and relationships, and unify them to formulate our 'theory of the recommender'. Finally, we will establish the trustworthiness of our finding by using an inter-rater reliability test.

To obtain a better outlook on the user's perceptual world of the recommender system, we will also use in-person interviews as our method of data collection alongside the online survey. This will allow us to ask more detailed questions and delve deeper into unique ideas that arise during the interview process.

In the long term, we are interested in how users change their behavior based upon their cognitive model of how recommender systems function. Users, sensitive to privacy concerns, may forgo the benefits of a recommender system and opt to view news stories in 'incognito' mode. Others, when presented with a political viewpoint with which they disagree, may aggressively downvote a content creator in a video sharing service in order to signal to the system that they are uninterested in those viewpoints. On the other hand, a user unhappy with the current set of recommendations may purposely search for and upvote items they have previously enjoyed in order to improve their user profile.

Anecdotal evidence suggests that, in recent years, these behaviors have become more commonplace. We must then ask: what is the impact of these behaviors on the health of the recommender system? We plan to perform experimental work on historical and synthetic data to understand the impact of these behaviors in a variety of domains.

Finally, we imagine a framework that can automatically identify and exploit the signals arising from these behaviors in order to improve the user experience. If a user assigns one star to a recommended movie, does it mean that the user truly dislikes it or simply that he does not want it taking up space in his recommendation queue? If a user uncharacteristically spends an online session rating several new songs, is this an indication that she is looking for more variety in her recommendations? In sum, the motivation of this research agenda is to understand the user's understanding of how recommender systems function, observe what behaviors that understanding manifests, and engineer recommender systems to take advantage of these behaviors to improve the system's performance.

5.4 Discussion

The participants of the online survey instrument have developed a cognitive model of how recommender systems function going far beyond a primitive understanding of inputs and outputs. They have developed a theory of the recommender. Just as the theory of mind describes an individual's ability to attribute goals, perceptions, knowledge, and beliefs to others, the theory of the recommender describes the ability of users to attribute these qualities to the recommender system. We have purposely adopted the theory of mind as an exemplar because in many ways the interaction a user has with a recommender system mimics more closely their interaction with other human beings than with other online applications.

It is common for a system to query the user's interests during a registration process. It may then suggest items to the user. The user may then consume the item, rate it, write a review of it, or even ignore it. Recommender systems exploit this feedback to make new recommendations which the user can then view. This cyclical interaction can be likened to a conversation with the recommender system. Word processors, online shopping carts, and wikis do not share this form of interaction.

Such an interaction directly impacts the relationship the user has with the recommender. The user can witness, sometimes immediately, the result of liking or disliking an item. The user can predict what would happen if they read news stories about a particular city or event. The user can reason why a recommender system is promoting a new song. The rich and ubiquitous interactions users have with recommender systems enables the users to refine over time a cognitive model of how they function.

Often the user is able to interpret the goals of the system. We observed in the survey responses several examples of perceived goals including 1) satisfying the user's interests, 2) aiding the decision-making process, 3) increasing loyalty, and 4) maximizing profit.

The user is also able to infer the perceptions of a recommender. Users clearly understand that their ratings, demographics, and click-throughs are observed by the system. Some users even understand that they can manipulate the output of the recommender by changing the information they allow the recommender to perceive.

The user may infer what knowledge the recommender captures. This knowledge can include how the system represents users or items. It can include metadata about the items. It may even capture the context of the user as the recommendations are being made.

Finally, the user can make assumptions about what the recommender believes. A fundamental belief of collaborative filtering is that users who have agreed in the past are likely to agree in the future. Content-based recommenders, on the other hand, believe that as a user consumes a product, she is also labeling herself with the characteristics of the item. Recommenders relying on association rule mining believe that consumption patterns observed in the past are relevant for the present. Users, experienced with recommender systems, have incorporated these beliefs into their mental model of recommender systems. All these beliefs were identified by the survey participants.

Implications of a user's theory of the recommender are many. Even though this work presents compelling evidence that users possess such a model, this pilot study is only the first step in formalizing it. How user behavior is informed by their mental model of the recommender and how recommender systems can adapt to these behaviors remains an important research direction.

6 CONCLUSION

In this paper, we asked the question: Does the User have a Theory of the Recommender. To answer this question, we developed a survey instrument to elicit a subject's understanding of recommender systems based on several scenarios. This survey was given to 25 participants. An exhaustive analysis of their responses demonstrates that the participants possess a keen understanding of many of the recommenders' basic algorithms and design goals. For example, many users seem to understand that recommenders often keep track of past behavior, identify similar users, leverage metadata, and seek to provide relevant and diverse recommendation.

This paper is the first step of a larger research agenda. Future research milestones include conducting a larger online survey until we reach a saturation of key concepts, constructing a grounded theory from these key concepts, and conducting in-person interviews to verify and improve upon the grounded theory. Later, we plan to evaluate how users modify their behavior based on their cognitive model of the system, what impact this behavior might have on the recommender, and how a recommender system can identify and leverage that behavior.

ACKNOWLEDGMENTS

We would like to thank the participants who completed the survey. We would also like to thank members of the Institutional Review Board who assisted in the design and approval of the survey instrument administered to human subjects. This work was supported in part by two PhD Scholarship Awards granted by DePaul University.

REFERENCES

[1] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. 1999. Towards a better understanding of context and context-awareness. In *International symposium on handheld and ubiquitous computing*. Springer, 304–307.

[2] P. H. Aditya, I. Budi, and Q. Munajat. 2016. A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for E-commerce in Indonesia: A case study PT X. In *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. 303–308. https://doi.org/10.1109/ICACSIS.2016.7872755

[3] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering* 6 (2005), 734–749.

[4] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. 1993. Mining association rules between sets of items in large databases. In *Acm sigmod record*, Vol. 22. ACM, 207–216.

[5] Ofer Arazy, Nanda Kumar, and Bracha Shapira. 2010. A theory-driven design framework for social recommender systems. *Journal of the Association for Information Systems* 11, 9 (2010), 455.

[6] Marko Balabanović and Yoav Shoham. 1997. Fab: content-based, collaborative recommendation. *Commun. ACM* 40, 3 (1997), 66–72.

[7] Hennie Boeije. 2002. A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality and quantity* 36, 4 (2002), 391–409.

[8] P Bonhard and MA Sasse. 2006. 'Knowing me, knowing you'-using profiles and social networking to improve recommender systems. *BT Technology Journal* 24, 3 (2006), 84.

[9] M-C Boudreau, D Gefen, and DW Straub. 2001. Validation in information systems research. *MIS quarterly* 25, 1 (2001), 1–14.

[10] Juliet M Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* 13, 1 (1990), 3–21.

[11] Marco De Gemmis, Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Giovanni Semeraro. 2015. Semantics-aware content-based recommender systems. In *Recommender Systems Handbook*. Springer, 119–159.

[12] David De Vaus and David de Vaus. 2013. *Surveys in social research*. Routledge.

[13] Chris Frith and Uta Frith. 2005. Theory of mind. *Current Biology* 15, 17 (2005), R644–R645.

[14] Barney G Glaser and Anselm L Strauss. 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

[15] Alvin I Goldman et al. 2012. Theory of mind. *The Oxford handbook of philosophy of cognitive science* (2012), 402–424.

[16] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22, 1 (2004), 5–53.

[17] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 7–10.

[18] Yehuda Koren and Robert Bell. 2015. Advances in collaborative filtering. In *Recommender systems handbook*. Springer, 77–118.

[19] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.

[20] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell me more?: the effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1–10.

[21] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, 3–10.

[22] Karen S Kurasaki. 2000. Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field methods* 12, 3 (2000), 179–194.

[23] Mark S Litwin. 1995. *How to measure survey reliability and validity*. Vol. 7. Sage.

[24] 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.

[25] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1097–1101.

[26] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making* 5, 5 (2010), 411–419.

[27] Eyal Peer, Gabriele Paolacci, Jesse Chandler, and Pam Mueller. 2012. Selectively recruiting participants from Amazon Mechanical Turk using qualtrics. *Available at SSRN 2100631* (2012).

[28] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences* 1, 4 (1978), 515–526.

[29] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 157–164.

[30] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. ACM, 175–186.

[31] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer, 1–35.

[32] Badrul Munir Sarwar, George Karypis, Joseph A Konstan, John Riedl, et al. 2001. Item-based collaborative filtering recommendation algorithms. *Www* 1 (2001), 285–295.

[33] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. 2002. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 253–260.

[34] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. ACM, 22–32.