Winter 12-6-2018

Design of Front-End for Recommendation Systems: Towards a Hybrid Architecture

Yixiao Wang

Ravi S. Sharma

Wang, Y.X. & Sharma, R.S. (2018). Design of front-end for recommendation systems: Towards a hybrid architecture. In *Proceedings of The 18th International Conference on Electronic Business* (pp. 221-230). ICEB, Guilin, China, December 2-6.

Design of Front-End for Recommendation Systems: Towards a Hybrid Architecture

(Full Paper)

Yixiao Wang, University of Canterbury, New Zealand, yixiao9004@gmail.com
Ravi S. Sharma*, University of Canterbury, New Zealand, rs.sharma@canterbury.ac.nz

ABSTRACT

To provide personalized online shopping suggestions, recommendation systems play an increasingly important role in "closing a transaction". Some leading online movie sales platforms, such as Netflix and Rotten Tomatoes, have exploited content-based recommendation approaches. However, the issue of insufficient information about features in item profiles may lead to less accurate recommendations. In this paper, we propose a recommendation method known as Collective Intelligence Social Tagging (CIST), which combines a content-based recommendation approach with a social tagging function based on crowd-sourcing. We used an online movie sales platform as a use-case of how a CIST approach could increase the accuracy of recommended results and the overall user experience. In order t0 understand the feasibility and satisfaction level for CIST, we conducted fifteen design interviews to first determine user-developer perspectives on CIST, and then collected their overall design input.

Keywords: Online suggestion systems, crowd-sourcing, collective intelligence, social tagging.

*Corresponding author

INTRODUCTION

The business value of a recommendation or suggestion is that it helps e-commerce platforms uncover associations among large amounts of transaction information for the purpose of providing personalized shopping suggestions for consumers. Recommendation and Suggestion Systems (RSS) are therefore a source of competitive advantage in the electronic marketplace. Business Insider reports that 80% of Netflix' business comes from its recommendation system and less than 20% from searches (see https://www.businessinsider.com.au/netflix-recommendation-engine-worth-1-billion-per-year-2016-6). Netflix believes it could lose $1 billion or more every year from subscribers quitting service if it not for its personalised recommendation engine. Google recently introduced an improved design and recommendation system to the Home screen of its YouTube app, in order to attract viewers into longer watch-sessions so as to "create the feeling that YouTube understands you" (see https://www.businessinsider.com.au/youtube-app-redesign-and-recommendations-2016-4). Apple's Watch List will recommend content and accompanying marketing messages across their Apple TV devices via a "universal search and suggestion mechanism" that delivers as fast as possible (see https://www.businessinsider.com.au/apple-tv-recommendation-app-2016-10). MightyTV is a meta-RSS which allows consumers to swipe through a list of movies or shows across service-providers such as Netflix, Hulu or HBO, and consequently suggests what new ones they should watch, based on profiles (see https://www.businessinsider.com.au/mightytv-2-million-investment-2016-8). In a value-added feature known as "mash-up", consumers can connect with friends for the RSS algorithm to model what their tastes have in common, from shows to specific actors. However, while high-quality, personalized recommendations benefit users, online businesses may violate laws if they collect too much user data or if they globalize its use. It is therefore timely to consider the design of a state-of-the-art approach that may be adopted in global, electronic markets.

It is taken for granted that online customers buy digital products from a host of platforms such as iTunes, Google Play, Amazon Prime Video, or Netflix on the basis of convenience. In order to enhance user experience and sales, almost all leading e-tailers have built their own recommendation systems. The notion of a recommendation or suggestion can help e-commerce platforms uncover associations among large amounts of transaction information for the purpose of providing personalized shopping suggestions for consumers (Venkateswari & Suresh, 2011). Commonly used recommendation approaches include content-based filtering, collaborative filtering, and hybrid filtering. Websites that depend on the content-based filtering approach may face insufficient item information, an issue that often results in less accurate recommendation results (Najafi & Salam, 2016). In order to close the gap between the actual and desired performance in the absence of an effective method to work around the issue, a new method called Collective Intelligence Social Tagging (CIST) recommendation approach for which depend on content-based recommendation, is proposed.

In the intensely competitive and ever-changing e-commerce world, continuous improvement of recommendation systems is pivotal. For this reason, conducting design research to recognize the effectiveness of CIST will help online sales. Mostly, these depend on content-based recommendations to improve the functionality of their current recommend systems. Resolving insufficiency identified through the research reported in this paper can further improve user experience and satisfaction. Considering that such a hybrid approach also includes the application of content-based recommendations (Najafi & Salam, 2016), the intended objective of this research is to inform more effective design of e-tailing platforms.The remainder of the paper is organized as follows. We begin with a review of state of the art of recommendation systems and crowd-sourcing based tags in Section 2. This is followed by the research design and ethical considerations in Section 3. We interpret and analyze data collected from our design interviews to come up with our key findings in Section 4, where we also refine our proposed hybrid recommendation approach. In Section 5, we conclude with suggestions for increasing the satisfaction level of users and mitigating the key design drawbacks of CIST.

REVIEW OF CURRENT PRACTICES

2.1 Recommendation Systems

With the growth in the number of products that users can view and purchase online, it is inevitable for users to face an information overload problem as they browse. One solution to this problem is the notion of a recommendation system (Cataltepe, Uluyağmur, & Tayfur, 2016; Schafer, Konstan, & Riedl, 1999). Recommendation systems have evolved into what is known as part of predictive Business Intelligence. It directly or indirectly prescribes to users of a web service to locate content, products or services relevant to their search. The essence of a recommendation service is underpinned by user behaviour. Gathering and analyzing users' data helps in a more accurate recommendations, which will then guide users to purchase products or services that are suggested to them (Park, Kim, Choy, & Kim, 2012). Recommendation systems specially helps e-tailers to "pitch" the right product to the right consumer at the right time to ultimately increase revenue (Foo et al. 2015).

However, such a collaborative approach has two major issues: the sparsity of the user-item matrix and the cold-start problem. The latter occurs if a product is not rated by any users at all, then it cannot be recommended. In such a case, a content-based approach is more suitable for recommending movies (Uluyagmur, Cataltepe, & Tayfur, 2012; Bergamaschi & Po, 2014). Some leading movie sales platforms such as Netflix, Rotten Tomatoes and Internet Movie Database, prefer to utilize the content-based approach (Bergamaschi & Po, 2014). For example, Netflix explains movie recommendations by showing users similar movies they have highly rated before (Vig, Sen, & Riedl, 2009). If a Netflix user has watched many "detective" movies, the system will recommend movies having such tags (Rajaraman & Ullman, 2014).

A content-based approach deals with item profile and user profile (Najafi & Salam, 2016; Uluyagmur, Cataltepe, & Tayfur, 2012; Alsalama, 2015). An item profile is a collection of records representing critical characteristics of the item. The features of a movie, such as actors, years, plot keywords, and genres, which can be linked to a recommendation system, can be an example (Rajaraman & Ullman, 2014). A user profile is built for each user containing their preferences, which are captured by analyzing the similarities between the items that the user has rated. A new item can then be recommended by content-based recommendation system through matching the existing user preference in the user profile with the features of the new item not yet rated by the user (Najafi & Salam, 2016).

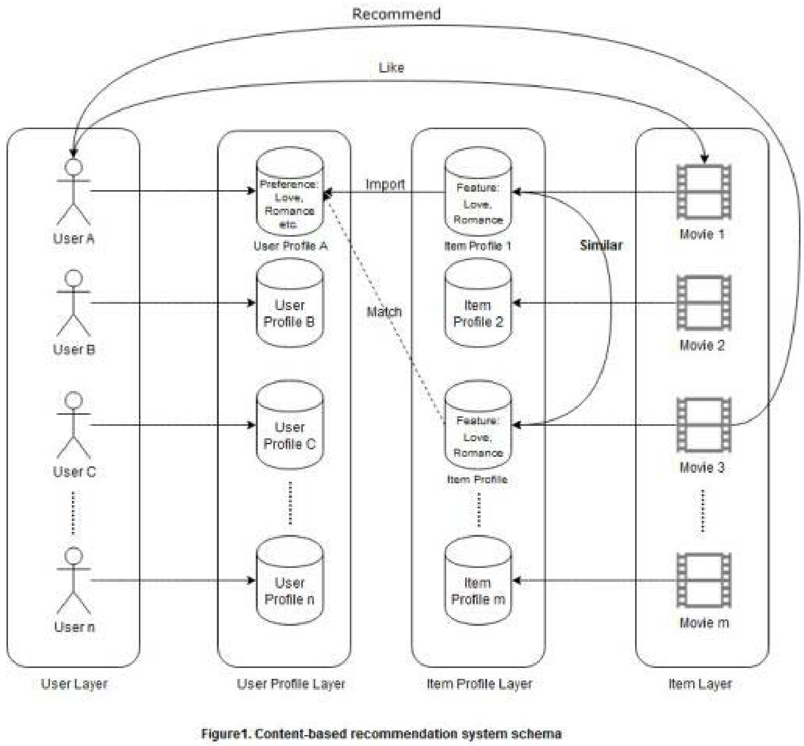

Figure 1: Content based recommendation system schema

Some researchers have indicated that tags, which are index terms, can be used to describe item features, so users can obtain information about items through knowing their tags (Rajaraman & Ullman, 2014; Vig, Sen, & Riedl, 2009). Therefore, we review the use of social tagging.

2.2 Social Tagging based on Crowd-Sourcing

In the era of web 2.0 and beyond, users have become sophisticated and in most cases, they are the content generators. Crowd-sourcing is referred to as collective intelligence in the literature. When more than one person contributes towards achieving an objective it creates more value, and collaborative efforts take less time (Sharma, Soe, & Balasubramaniam, 2014). Tags have become more popular with the likes of Flickr, Tumblr, Amazon, Facebook, and Instagram and so on. Users like tags as they are self-explanatory, and their search purpose can be solved within a shorter period of time.

It is worth mentioning that film features, such as genres and plot keywords, can be described by tags (Rajaraman & Ullman, 2014). However, if websites assign tagging tasks to internal staff, it will undoubtedly increase the workload of internal staff. Thus, some authors suggested adopting crowd-sourcing to perform tagging (Nguyen-Dinh, Waldburger, Roggen & Tröster, 2013; Hsueh, Melville, & Sindhwani, 2009). Smaller e-tailers could recruit external annotators through Internet services (e.g., Amazon Mechanical Turk (AMT)) to allow multiple tagging tasks to be done in bulk with fast completion rates and low overall costs (Hsueh, Melville, & Sindhwani, 2009).

It is not necessary for users to be an expert to create tags. Hence, crowd-sourcing can also be a good way to increase user-system trust in recommendation systems and understand users' perspectives on presentation, explanation, and priority online (Berkovsky, Taib, & Conway, 2017; Vig, Sen, & Riedl, 2009).

2.3 Content-Based Recommendation Systems

The main view currently is that recommendation systems can facilitate customers' purchasing processes and make positive impact on the user experience of online purchasing (Venkateswari & Suresh, 2011; Swathi & Reddy, 2014). In the online movie sales industry, unlike the collaborative approach, content-based approach does not have the problem of the sparsity of the user-item matrix and cold-start (Uluyagmur, Cataltepe, & Tayfur, 2012). The content-based approach is more inclined towards user information, like user preference, and item information, like ratings. It recommends items in-sync with the interest demonstrated by users in the past. Even if a movie was not rated by any users before, so long as the movie has sufficient movie information, it can be recommended to users who have similar user preference. The extent of sales diversity is then increased by allowing consumers to access to more new and niche products among seemingly endless alternatives (Fleder & Hosanagar, 2009).

2.4 Social Tagging based on Crowd-Sourcing

Using crowd-sourcing tools (e.g., AMT) to generate tags allows multiple tagging tasks to be done in bulk with fast completion rates (Hsueh, Melville, & Sindhwani, 2009). For start-ups and de-novo firms entering a digital marketplace, it provides several benefits, such as broader access to specialized skills, shorter product development cycles, more flexible and faster-hiring processes as well as relatively low costs (Knop, Durward, & Blohm, 2017). Moreover, the empirical results of Nguyen-Dinh et al. (2013) showed that the final annotation set generated by AMT is of comparable quality of the annotation of experts with high accuracy (76% to 92%). Generating tags using crowd-sourcing can increase the level of active engagement of users, and it allows users to reflect their real views through input of tags to further increase user experience and satisfaction level to some extent (Carletti, Giannachi, Price, McAULEY, & Benford, 2013).

Overall, the more tags an item has, the more features of the item that can be captured and displayed to users. Accordingly, the more user preferences that can be uncovered, expressed and matched correspondingly, the more accurate the recommended results may be expected to be. Thus, in order to generate more tags faster and at lower cost, crowd-sourcing is utilised.

2.5 Shortcomings of Content-Based Recommendation Systems

Insufficient information in item features will lead to less accurate results (Najafi & Salam, 2016). For many movie sales platforms, especially small websites, they do not have enough capacity to generate many tags to cover all features in movies. In most cases, they only create tags for main themes of movies, and many sub-themes will be ignored. For example, in Figure 1, the film genres of Movie 1 are "Love" and "Romance", and Movie 3 contains the same genres. The relevance of the items are then established based on the similarity, and Movie 3 is then recommended to User A. However, case in point is that Movie 1 has traces of "Limerence" which is a form of love, while Movie 3 contains "Cheating" plot, even if their plots in actuality account for only a few minutes of footage. But since both these features are not generated as sub-themes and enlisted there, if User A likes the former more than the later, a recommendation of Movie 3 to User A may not be a successful exercise itself because of insufficient film features related data.

One of the main issues that online recruitment of taggers brings up is that these workers are not specifically trained for tagging, and we cannot control the quality. Therefore, the quality of their output may not be as high as we have expected (Nguyen-Dinh et al., 2013; Hsueh, Melville, & Sindhwani, 2009).

2.6 Research Gap in Architecture

AMT is a useful tool for tagging items based on a crowd-sourcing technique. Nguyen-Dinh et al. (2013) conducted experiments of tagging human activities in videos by using AMT with high accuracy of tags (76% to 92%). It showed the effectiveness and feasibility of generating tags in videos by crowd-sourcing. Thus, it is possible to use crowd-sourcing to tag movies and generate more social tags to enrich movie features (e.g., film genres and plot keywords) to increase the accuracy of recommended results. Hsueh, Melville and Sindhwani (2009) introduced three selection criteria for determining high-quality annotations: 1) the noise level of a group of annotators, 2) the inherent ambiguity of an example's class label, and 3) the informativeness of an example to the current classification model. They found that the quality of tags can be improved through eliminating ambiguous examples and noisy annotators. Nguyen-Dinh et al. (2013) proposed two strategies – 1) individual filtering and 2) collaborative filtering - to detect and remove non-serious taggers to increase the accuracy of activity annotation in videos. However, few articles discussed criteria models for examining and evaluating the quality of those selected tags, which passed the selection step. This is a research gap, and in this paper we develop a criteria model for this purpose.

DESIGN RESEARCH METHOD

The research method adopted in this paper follows the five iterative stages of design thinking to generate solutions for improving the user experience (see Figure 2).Generating solutions and model validation are significant steps in design research (Peffers et al., 2008). It checks if the design artefact is acceptable and useful for addressing the problems being addressed (Macal 2005). In design research, relevance is a key aspect of solution validation (Checkland, 1995). It can be learned through observing and analysing stakeholders' views and actions in the context of a task and system. Such content analysis of discussions can be assumed as a reflection of issues in the real world and the values of stakeholders (Hanafizadeh & Mehrabioun, 2018). By drawing from the guidelines of design research, where experts and novices are sought for their comments and insights, we conducted such design interviews with users and developers in order to iteratively improve our understanding of effective recommendation systems. For the purpose of this research study, online movie platforms were chosen as a convenient yet rich context for the design interviews.

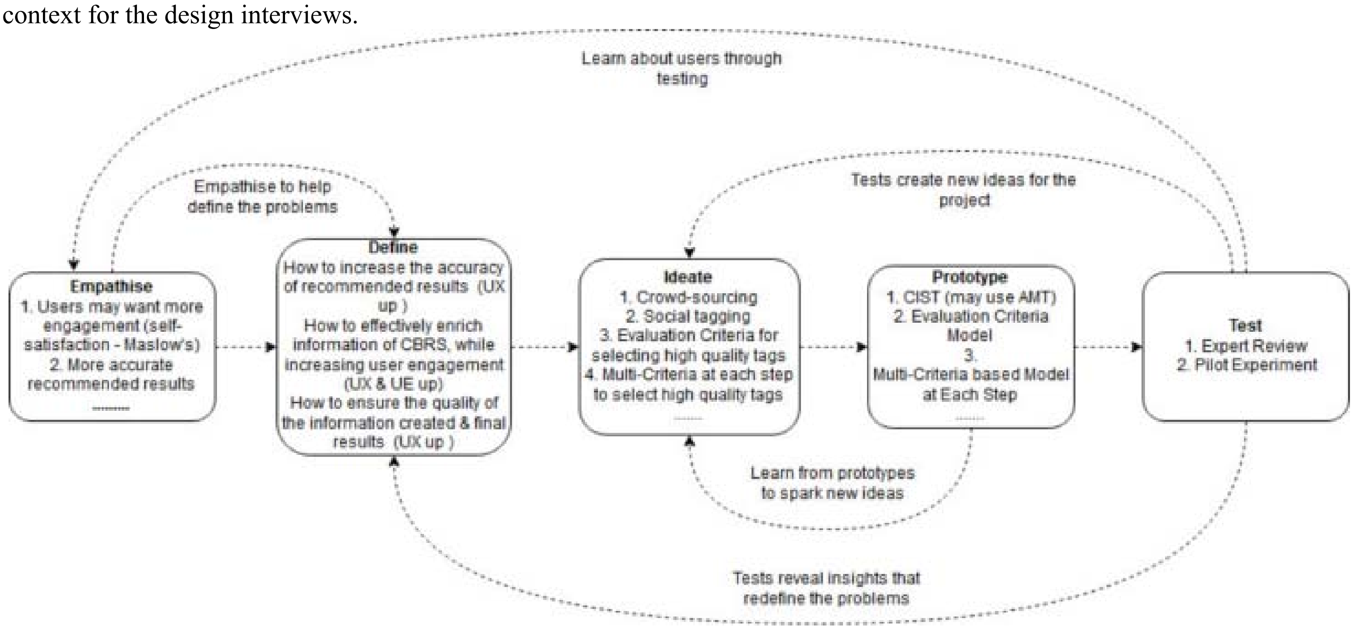

Figure 2: Stages of Design Thinking (Adopted from Dam & Siang, 2017)

3.1 Design Interviews with Developers

In order to obtain design feedback from experts to improve our CIST architecture, we conducted 60-90 minute, individual interviews with five developers having relevant knowledge and experience with in developing applications for 3 years or greater. We applied design thinking to generate a solution for examining and evaluating tags created by the crowd to improve the accuracy of the recommendation. We previously noted that few researchers discuss systematic models for evaluating the quality of tags.

Step 1 was to gain an empathic understanding of the problem that needs to be solved. We adopted a role-playing scenario where we advocated a user-perspective in our design questions. Our context was that of online platforms which depend on content-based movie recommendation systems. From a customer's point of view, we derived sixteen examination and evaluation criteria that will possibly have an influence on the quality of tags based on our experience. Developers who were interviewed helped us validate and improve our criteria, and they are asked to complete a form and tick the criteria that would influence the quality of tags. In addition, they could list any new criteria that they come up with.

Step 2 involved a definition of key vocabulary and this was corroborated with the developers.
The over-arching focus of the design interview was the question "How might we use these criteria to examine and evaluate the quality of tags for online movie sales platforms in order to get accepted tags with high quality and further increase end-users'

experience." The objective of interviews was next (as Step 3) to ideate as many as possible solutions to resolve any human-centered problems based on the comments of respondents (Conrath & Sharma, 1993).

Table 1: Assessment and evaluation criteria

Criteria of Quality for Tags	Criteria of Quality for User Accounts	Quality for Examining Mechanism
1. Relevant: Tags must be relevant to the contents of items.	1. Last Active: Last Login Time.	1. The Frequency of the Examining Mechanism Version Update: Update frequency about the rules/version of Tagging-related knowledge of Examining Mechanism. Version, which contain rules, must keep up with the development of state-of-the-art technology.
2. Searchable: Tags can be recognized by search engines and can be used to search relevant items.	2. Credibility of Accounts: The level of trustworthiness rated by other users about all activities of one account.	2. Quality of Examiners: The knowledge level of tag examiners.
3. Reusable: Tags for one item can be reused for other items.	3. Total Cumulative Posts: The total number of post generated by one account.	3. Quality of Examining Technology: The performance level of back-end examining technology for tags.
4. Valid: Tags must have a sound basis in logic or fact; reasonable or cogent.	4. Rating for Posts: The level of usefulness and helpfulness rated by other users about all review posts generated by one account.	4. Degree of Implementation of Relevant Laws and Rules: Tags must comply with local laws and regulations.
5. Meaningful: Tags must have meanings.	5. Degree of Completion for User Information: The level of completion of the account information filled in by the owner of the account.	
6. Extensible: Tags can be given or added new meanings.		
7. Stable: The meaning of tags must be consistent within a certain period of time.		

3.2 Design Interviews with Users

In order to complete the comprehensive view of recommendation systems, feedback from end-users were also obtained from design interviews. Ten users, including five lead users and five novice users participated in hour-long, individual sessions. Design thinking theorizes that users with different levels of usage experience with recommendation systems and tags may have different views on social tagging and hence we divided users into lead and novice users. We define lead users as those who have experience in using online movie sales platforms in excess of 2 years and novice users as those who do not.

The design interviews proceeded as follows. First, all participants were required to complete a task in one scenario to help them to understand the application of tags in recommendation systems and its benefits to users. Because CIST currently is only a conceptual idea and has not been implemented in any online movie platform, we allowed the participants to use Tumblr and Amazon.com to experience the idea of tagging. On Tumblr, one instance of CIST, almost every picture is tagged, and the website can recommend relevant blogs and searches based on users' history about tags that are frequently searched or viewed. In addition, if a given user follows a particular blog, then other blogs will be recommended based on the similarity of tags and the number of similar tags used among those blogs. Participants were required to browse and search any tags and pictures that they liked. They could then check if Tumblr recommended blogs and searches that they also liked. Amazon.com, as a traditional e-commerce website, does not use tags to recommend items. Participants were required to browse and search items that they liked. They could later check the recommended results provided by Amazon. After the exercise, the interviewed “users” were given time to compare the recommended results provided by two websites, in terms of accuracy, user experience and trustworthiness.

3.3 Design Statements & Interview Questions

The box below is a shortened version of design interview template for developers (4 statements, 9 questions) and users (3 statements, 6 questions) respectively.

Design Statement 1*: IT developers think that the current functionality of recommendation systems can improve online shopping.* Question 1: Someone said that recommended results on online shopping websites can help them to find the products that they like to buy quickly, do you agree with this statement? Please provide some comments. Question 2: Are the products recommended by the recommendation system what you really want to buy most of the time? In other words, how do you think about the accuracy of recommendation systems currently?
Design Statement 2: *IT developers think that tags can improve the functionality of recommendation systems and further motivate online purchase.* Question 3: Since tags can present information for online contents, how do you think about using tags to enrich item profile to further make recommendation results more accurate?
Design Statement 3: *IT developers think that it is feasible to allow users to generate tags for items online.* Question 4: Have you contributed tags as a user? Do you agree with the statement that user-generated contents can lead to accuracy/trust for other users? Question 5: Would you like to contribute tags yourself? If yes, what motivates you to do this? If no, why? Question 6: What are the main advantage and disadvantage of user-generated tags, compared with developer-generated tags? Please list one separately.
Design Statement 4: *IT developers think that the functionality of recommendation system can be improved by applying tags generated by users and an Accepted Tag Model, which is to ensure the quality of tags.* Question 7: Do you agree that more tags generated by more users for a product can help to enrich item profile and discover latent needs of other customers? Please provide some comments. Question 8: Do you think that the quality of tags can be improved by applying an accepted tags model, which will contain criteria listed on Table 1? Question 9: Overall, do you think that the accuracy, user experience and trustworthiness for recommendation system can be improved by adding high-quality tags to describe items?
Design Statement 1: *Lead users are more satisfied with the current functionality of recommendation systems, in terms of accuracy, trustworthiness and user experience.* Question 1: Someone said that recommended results on online shopping websites can help them to find the products that they like to buy quickly, do you agree with this statement? Please provide some comments. Question 2: How does recommendation system change or affect your purchasing process? Please provide some comments in terms of trustworthiness and user experience. Question 3: Do you feel satisfied with the accuracy of recommended results based on your purchase? Please provide reasons.
Design Statement 2: *Lead users are more likely to contribute to creating tags for online products, and they are more satisfied with the impact of using more tags for describing items on the improvement of the accuracy of content-based recommended system.* Question 4: Would you like to add your own tags as feedback for online movie products that you purchased or watched? If yes, what motivates you to do this? If no, please provide reasons. Question 5: "The more tags about a product, the more you know about it" - do you agree with this statement? Please provide some comments.
Design Statement 3: *Novice users are more concerned about the use of social tagging based recommendation systems and crowd-sourcing based tags.* Question 6: Among security, ease of use, accuracy and/or trustworthiness, what is your biggest concern about recommendation systems and /or user-generated contents (e.g. tags)?

ANALYSIS OF DESIGN STATEMENTS

4.1 Consolidation of Feedback

The purpose of the interviews for IT developer group is to help to refine CIST. IT developers, who have relevant knowledge and experience, are more likely to provide reliable and professional feedback. The user group, comprising lead and novice users, is mainly used to measure the satisfaction level of CIST from a user's perspective. Any concerns raised by participants are an important reference for us to improve CIST. Seven key findings synthesized from our design interviews are summarized as follows: 1) The current functionality of recommendation systems can improve online shopping. 2) Tags can improve the functionality of recommendation systems and further motivate online purchase. 3) It is feasible to adopt user-generated tags for describing online products. 3) The current functionality of content-based recommendation system can be improved by adding high-quality tags screened out by an Accepted Tag Model. 4) Lead users are more satisfied with the current functionality of recommendation systems, in terms of accuracy, trustworthiness and user experience. 5) Both lead and novice users think that recommendation systems can positively affect their purchasing process. 6) Both lead and novice users are likely to contribute tags for online products, but lead users are more satisfied with the potential impact of using more tags for describing items on the improvement of the accuracy of content-based recommended system. 7) Lead users are more concerned about the use of social tagging based recommendation systems and crowd-sourcing based tags than novice users. The above findings guide design guidelines for CIST described in the Section 4.2.

4.2 Design Guidelines for Accepted Tag Model and CIST

4.2.1 Accepted Tag Model

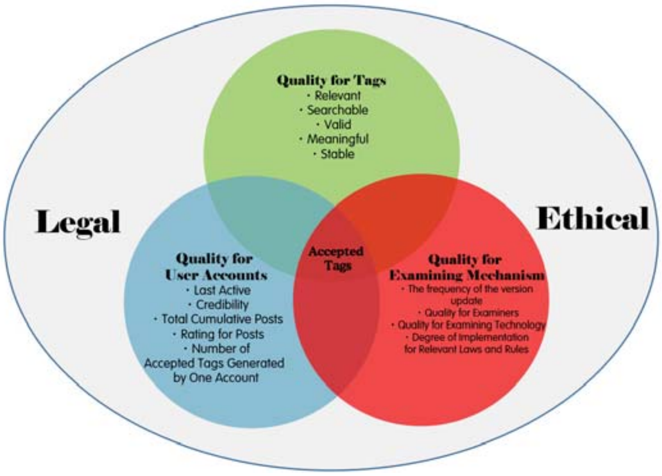

Figure 3. Accepted Tag Model

Figure 3 illustrates the key design concepts associated with Accepted Tag Models. For CIST, user-generated tag is the main source for recommending items. However, from the feedback of our participants, we determined that most are concerned about the quality of user-generated tags. Meanwhile, generating too many tags, especially irrelevant tags, under products will confuse users because of information overload. In order to improve the quality of tags as well as cut down the number of tags, especially the number of low-quality tags, we propose an accepted tag model. Only when tags meet most of the criteria outlined in Table 1 can they be considered as high-quality tags. However, since the capacity of database and software processing may be limited, and we may not accept and process all tags created by all users, we should have criteria in the blue sphere to cut down the number of tags through setting the threshold for user accounts. Therefore, only high-quality tags created by high-quality user accounts will potentially be accepted. In spite of this, whether tags are finally accepted or not depending largely on the quality of examining mechanism. Only when the version of mechanism is up-to-date, it can learn to recognize latest tags. For example, if you tag #8KUHD for a movie before 2015, it may not be accepted, because at that time there is no 8K UHD movie at all. However, if you tag #8KUHD five years later for a real 8K UHD movie, but the version of examining mechanism is not up-to-date, the tag may still not be accepted. This is not the fault of that tag, but the fault of the outdated examining mechanism. In order to make this model effective all the time and not miss any high-quality tags, the version of mechanism should be updated frequently to be up-to-date. Quality of examiners and examining technology decide whether high-quality tags can be finally accepted by websites as well. Knowledgeable examiners and advanced AI examining technology are more likely to capture any potential high-quality tags. Finally, examining mechanism should always ensure compliance with relevant laws and rules to evaluate and examine tags, in order to make sure that all high-quality tags generated by high-quality users are also legal and ethical. Developers said that it is feasible to use current technology to implement such a model for real movie sales websites. Therefore, this approach can be treated as a solution for controlling the quality of tags and avoiding information overload caused by too many tags.

4.2.2 CIST Schema

The Accepted Tag Model can be seen as one part of CIST. The main purpose of the model is to improve the quality of tags to further provide a high-quality source for accurate recommendations. With the benefit of design input obtained in the manner described in Section 3, the modified schema of CIST is shown in Figure 3.

To simplify the use-case, we label item profile to represent a movie and label user A as an example. The issue of insufficient information in item profile now can be resolved by using user-generated high-quality tags to enrich item profiles. Previously, many movie sales platforms, do not have enough capacity to generate many tags to cover all features in movies. In most cases, they only create tags for main themes of movies, and many sub-themes are ignored. In Figure 4, the item profile 1 and 3 have limited numbers of features, for example, only "Love" and "Romance" as main themes are generated there. However, if Movie 1 contains "Limerence", and Movie 3 contains "Cheating", but both features as sub-themes are not generated and listed there, it may not be a good fit to recommend Movie 3 to User A, whose preference does not contain "Cheating", according to his viewing and purchasing history.

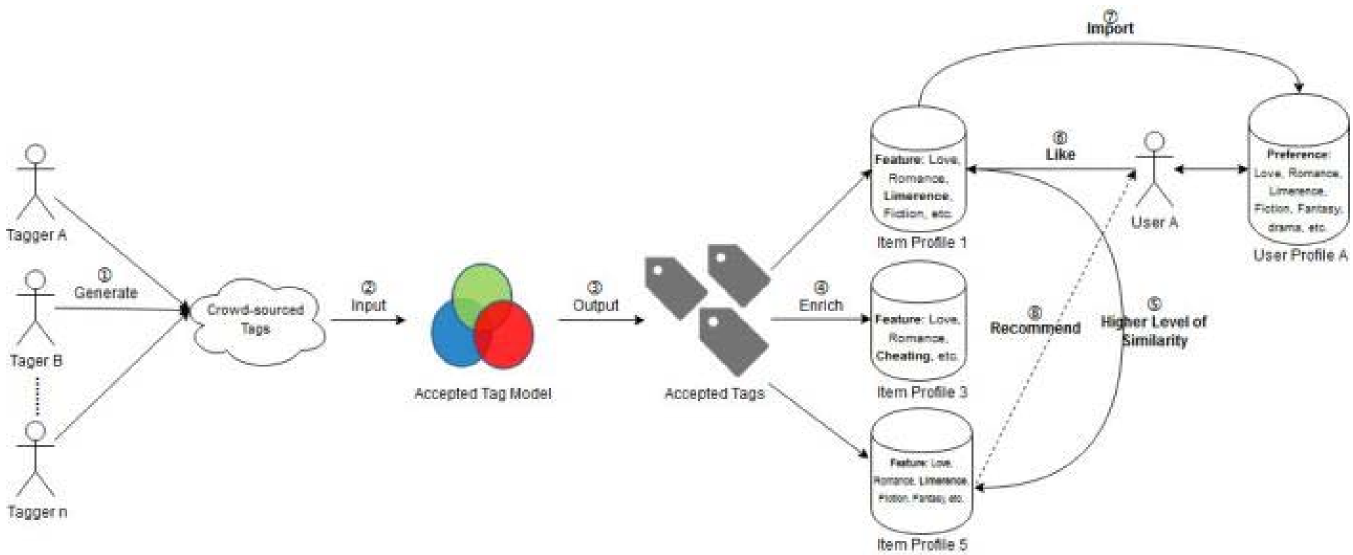

Figure 4. Modified CIST Schema

Since we use user-generated high-quality tags to enrich item profiles, item profiles can include more tags. Even sub-themes, which only account for a few minutes in one movie, can be covered and described by tags. We can then find that actually item profile 1 and 5 have higher levels of similarity, and item profile 5 is closer to User A's preference, which is accumulated from past history. Therefore, recommending Movie 5 to User A will be a better choice, after we have enriched item profiles. The number of tags expected to be generated for each movie may vary from movie to movie, but we expect that the estimated number should be three times of the current number or more.

4.2 Validating the Design Process

In order to validate our design statements, we cross checked the sentiment expressed by developers and users against each design statement. We used IBM Watson's natural language understanding tool (www.ibm.com/watson/services/natural-language-understanding/) to conduct sentiment analysis to understand participants' attitudes towards each of the design statements discussed with them. Note that the sentiment score calculated by IBM Watson only reflects the sentiment level of text (which were transcribed from voice recordings of the design interviews), and it can be positive or negative. The scores for design statements (except Design Statement 3 for users) are shown in Figure 5 for all three groups. Higher scores reflect greater concordance or positivity towards a design statement.

User Type \ Design Statement	D1	D2	D3	D4
IT Developer	1.91	0.7	1.22	3.3

User Type \ Design Statement	D1	D2
Lead User	1.93	0.68
Novice User	1.31	0.15

Figure 5. Output from IBM Watson Natural Language Sentiment Analyzer

On the one hand there was consistency in terms of positive inclination towards machine generated recommendations. Yet there were subtle differences in the perspectives of novice or lead users on the one hand and developers on the other.

PRACTICAL IMPLICATIONS AND CONCLUDING REMARKS

In Section 4, some key findings showed that different stakeholder groups have different attitudes towards recommendation systems or tags. Meanwhile, many concerns about CIST were raised by participants. In this concluding section, we will analyze possible reasons why the attitudes of some users are not as positive as the rest and come up with three main suggestions for increasing lower satisfaction level and mitigating their concerns.

5.1. Accepted Tag Model

As mentioned previously, the two major disadvantages are the lack of control over the quality of tags and information overload due to too many tags. In order to resolve them, we have introduced an Accepted Tag Model in Section 4.4.1. Each of three spheres has its own function. When user-generated tags pass all criteria in the green sphere, we can get high-quality tags. The criteria in the blue sphere limit the number of users who can generate tags and further cut down the number of tags. These two spheres ensure that all high-quality tags are generated by high-quality users. The red sphere contains some limiting factors of examining mechanism, which may or may not further reduce the number of high-quality tags. Although we do not expect any further reduction in the number of high-quality tags, we still need to consider any possible factors, which may lead to this result, for the mechanism. Finally, we can get accepted tags with high quality and not an excessive quantity.

5.1.1 Justifying CIST Results and Tagging Function

From the design research, it was found that the satisfaction level of novice users for recommendation systems is lower than that of other two groups, and we assume that the situation is due to novice user's lack of basic knowledge about how recommendation systems work. In order to allow users to understand how CIST works, we can place a conspicuous question mark icon at the top right of the recommended results area. When users click on that icon, a dialog box, which shows the reasons why these items are recommended, will pop up. A pie chart, which contains tags that users frequently search or are included in the products which users frequently buy, can also be shown in that dialog box. The more frequently a tag interacts with a user, the larger percentage the tag will take in the pie chart. All data in the pie chart are based on big data statistics of users' history. Through showing the explanation for recommended items, users will know how CIST works and understand that all recommendations are based only on their own history and are not interfered by producers of items. In addition, to make tagging function easy to use, a step-by-step instruction is employed to guide users to tag. The instruction will tell users what they should do next when users start to tag.

5.1.2 Concise and Marked Way of Displaying Tags in the Front-end Interface

User-generated tags are not only used to recommend items, but also provide simple descriptions of items. However, if we randomly put many tags on one movie's information page, users might still feel confused about which tags can best describe the movie, and it will make users' attitudes for tags more negative. In order to avoid this problem, we can allow users to "+1" or "-1" tags. When users put their mouse cursors on existing tags, a dialog box, which contains "+1" or "-1", will appear. When users think that some tags can reflect the main theme of the movie, they may "+1" those tags, otherwise, "-1". When a tag receives a certain number of "+1" (for example, 100 "+1"), the font size of that tag will become bigger (Up to three times bigger). Hence, even when users see many tags under one film, they can quickly know what the film is mainly about. Meanwhile, only six tags with the most "+1" will be shown directly in the front-end interface, while the remaining tags can only be shown when users click on "See More" at the end of all six tags (see Figure 6.).

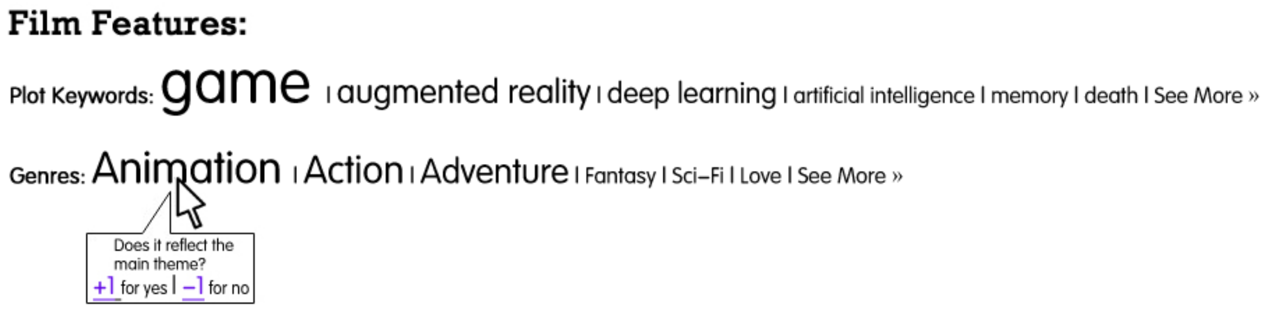

Figure 6. An Example of How to Display Tags

Taking into account information amount limit, the number of tags shown on one web page may be under fifty after users click on "See More", and other tags will be stored in the back-end database for recommendation purpose only.

In this paper, Collective Intelligence Social Tagging (CIST) is proposed as a recommendation approach for online movie sales platforms. In order to test the feasibility and user-acceptance level for CIST, design thinking was used to illicit attitudes on two major parts of CIST: the recommendation and user-generated tagging. Although developers and lead users' overall attitudes towards CIST were positive, there are still some concerns raised by novice users. The design contribution of this research are three suggestions to improve CIST: 1) Adopting Accepted Tag Model; 2) Adding necessary explanation for recommended results and tagging function; 3) Applying concise and marked ways of displaying tags in the front-end interface.

One main open issue we could not effectively resolve was regarding the security and privacy concerns of users about CIST. Although every e-commerce website has clear agreements containing privacy and security terms, it does not completely eliminate user fears of violations. For instance, it is well known that IP addresses may still leak users' private information related to their purchasing and tagging preferences. Hence, finding effective solutions and improvements for CIST to avoid invasions of privacy through IP addresses is a necessary design feature for CIST.

Our future work includes constructing a practical prototype of CIST and using larger groups of developers and users in order to test user experience and system performance. In addition, solutions for resolving security and privacy concerns about CIST may be investigated.

REFERENCES

[1] Alsalama, A. (2015). A hybrid recommendation system based on association rules. International Journal of Computer and Information Engineering, 9(1), 55-62. Retrieved (4th, Nov., 2017) from http://waset.org/publications/10000147/a-hybrid-recommendation-system-based-on-association-rules

[2] Berkovsky, S., Taib, R., & Conway, D,. (2017). How to Recommend?: User Trust Factors in Movie Recommender Systems. 287-300. doi:10.1145/3025171.3025209

[3] Bergamaschi, S., & Po, L. (2014, April). Comparing LDA and LSA Topic Models for Content-Based Movie Recommendation Systems. In *International Conference on Web Information Systems and Technologies* (pp. 247-263). Springer, Cham.

[4] Conrath, D.W. & RS Sharma, R.S. (1993). "A note on some overall evaluation measures for computer-based information systems", Computers in Industry, 21, pp 267-271.

[5] Carletti, L., Giannachi, G., Price, D., McAuley, D., & Benford, S. (2013). Digital humanities and crowdsourcing: An exploration. *Museums and the Web 2013*,in N. Proctor & R. Cherry (eds). Museums and the Web, Silver Spring, MD.

[6] Cataltepe, Z., Uluyağmur, M., & Tayfur, E. (2016). Feature selection for movie recommendation. *Turkish Journal of Electrical Engineering & Computer Sciences, 24*(3), 833-848.

[7] Corti, L., Day, A., & Backhouse, G. (2000). Confidentiality and Informed Consent: Issues for Consideration in the Preservation of and Provision of Access to Qualitative Data Archives. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1*(3). Retrieved (3rd, Nov., 2017) from http://www.qualitative-research.net/index.php/fqs/article/view/1024/2207

[8] Dam, R., & Siang, T. (2017, October 25). 5 Stages in the Design Thinking Process. Retrieved (4th, Nov., 2017) from https://www.interaction-design.org/literature/article/5-stages-in-the-design-thinking-process

[9] Eysenbach, G., & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. *British Medical Journal, 323* (7321), 1103-1105.

[10] Fleder, D., & Hosanagar, K. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science, 55*(5), 697-712.

[11] Foo, F., Sharma, R. & Chua, A. (2015). Knowledge Management Tools and Techniques, (3rd edition), Prentice-Hall, Singapore. ISBN: 9789810678463. Revised Digital Edition.

[12] Hsueh, P. Y., Melville, P., & Sindhwani, V. (2009, June). Data quality from crowdsourcing: a study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing* (pp. 27-35). Association for Computational Linguistics.

[13] Knop, N., Durward, D., & Blohm, I. (2017). How to Design an Internal Crowdsourcing System. Proceedings of the 38th International Conference on Information Systems. Seoul. Retrieved February 14, 2018.

[14] Najafi, S., & Salam Patrous, Z. (2016). Evaluating Prediction Accuracy for Collaborative Filtering Algorithms in Recommender Systems (Unpublished PhD Dissertation). Retrieved (2nd, Nov.,2017) from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-186456

[15] Nguyen-Dinh, L. V., Waldburger, C., Roggen, D., & Tröster, G. (2013, April). Tagging human activities in video by crowdsourcing. In *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval* (pp. 263-270).

[16] Park, D. H., Kim, H. K., Choy, I. Y., & Kim, J. K. (2012, September 1). A literature review and classification of recommender systems research. *Expert Systems with Applications, 39*(11), 10059-10072.

[17] Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24* (3), 45-77.

[18] Rajaraman, A., & Ullman, J. D. (2014). *Mining of Massive Datasets*. Cambridge: Cambridge University Press.

[19] Schafer, J. B., Konstan, J., & Riedl, J. (1999, November). Recommender systems in e-commerce. *Proceedings of the 1st ACM Conference on Electronic Commerce* (pp. 158-166). ACM.

[20] Sharma, R., Soe, K. M., & Balasubramaniam, D. (2014). Case studies on the exploitation of crowd-sourcing with Web 2.0 functionalities. *International Journal of Electronic Business, 11*(4), 384-408

[21] Swathi, V., & Reddy, M. S. (2014). Music Recommendation System Using Association Rules. *International Journal of Technology Enhancements and Emerging Engineering Research, 2*(7), 31-34. Retrieved (6th, Nov., 2017) from http://www.ijteee.org/final-print/july2014/Music-Recommendation-System-Using-Association-Rules.pdf

[22] Talerico, D. (2012, April). Privacy and Confidentiality: Issues in Research. Retrieved (3rd, Nov., 2017) from https://evergreen.edu/sites/default/files/humansubjectsreview/docs/PrivacyConfidentiality.NEEP.May2012.pdf

[23] Uluyagmur, M., Cataltepe, Z., & Tayfur, E. (2012). Content-based movie recommendation using different feature sets. *Proceedings of the World Congress on Engineering and Computer Science, 1*, pp. 17-24.

[24] Venkateswari, S., & Suresh, R. M. (2011). Association Rule Mining in E-commerce: A Survey. *International Journal of Engineering Science and Technology, 3*(4), 3086-3089.

[25] Vig, J., Sen, S., & Riedl, J. (2009, February). Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th ACM International Conference on Intelligent User Interfaces* (pp. 47-56). Association for Computing Machinery.