Association for Information Systems
AIS Electronic Library (AISeL)

ICEB 2018 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-6-2018

Design of Back-End of Recommendation Systems Using Collective Intelligence Social Tagging

Ravi S. Sharma

Rati Kale

Design of Back-End of Recommendation Systems Using Collective Intelligence Social Tagging

(Full Paper)

Ravi S. Sharma, University of Canterbury, Christchurch, rs.sharma@canterbury.ac.nz
Rati Kale, University of Canterbury, Christchurch, rok13@uclive.ac.nz

ABSTRACT

Recommendation systems are the tools whose purpose is to suggest relevant products or services to the customers. In a movie business website, the recommendation system provides users with more options, classify movies under different types to assist in arriving at a decision. Although, with current e-commerce giants focusing on hybrid filtering approach, we have decided to explore the functionality of Content-based recommendation system. This research paper aims to delve deeper into the content-based recommendation system and adding tags to enhance its functionality. The content-based approach is more fit to the movie recommendation as it overcomes the 'cold start' issue faced by the collaborative filtering approach, meaning, even with no ratings for a movie, it can still be recommended. The proposed method is to solve the less 'data categorization' issue in content-based filtering. Collective Intelligence Social Tagging System (CIST) aims at making a significant difference in content-based recommendation system to enrich the item profile and provide more accurate suggestions. The main gist of CIST is to involve the users to contribute in tagging to build a more robust system in online movie businesses. Tags in the millennial world are the 'go to' words that everyone looks up to in an online world of E-commerce. It's the easiest way of telling a story without actual long sentences. We recommended three main solutions for the concerns of CIST, (a) clustering of tags to avoid synonymous tag confusion and create a metadata for movies under same tags, (b) 5 criteria model to motivate and give the most amount of genuine information for end users to trust and eventually contribute in tagging, and (c) clear way of distinguishing and displaying tags to separate primary tags and secondary tags and give a chance to the users to assess whether the given tags reflect the relevant theme of the film.

Keywords: digital commerce; suggestion systems; social networks

*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 (see https://www.businessinsider.com.au/) reports that 80% of Netflix' business comes from its recommendation system and less than 20% from searches. 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." 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. 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. 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.

In the era of digital lifestyles, there are numerous choices available to make a single decision like to select which movie or which song to watch and play. Sometimes, this information overload makes it difficult to arrive at decision. One such personalization technique that assesses search and online behaviour and helps in arriving at a decision is the recommendation system (RS). Most of the major online retailers with strong online presence like Amazon Prime, Netflix, and Hulu use RS (Advomavicius & Kwon, 2007).

Before the emergence of electronic commerce, the traditional of recommending products or services was 'Word of Mouth'. This approach was social and had an individual experience attached to it which swayed societies (Chartron, Kembellec, & Saleh, 2014). In this fast-paced, ever-evolving world of product-related Internet offers, information has exploded exponentially. Too much of information can deteriorate the quality of messages reaching out to the customers. There has to be a set of rules and ways to filter information from this Big Data. For an end user, it is necessary to filter and pre-processes the available data prior to display to match his/her expectations (Chartron, Kembellec, & Saleh, 2014).

Social tagging has emerged as a valuable technique to tag with tools like Flickr, Furl, Del.icio.us and more. In this time of Web 2.0, users have become more active and a major source of generating new content online (foo et al., 2015). The end users are allowed to assign a personal label or tag on a site for purposes like sharing, discovering, resourcing and more. E-tailers, on the other hand, benefit a lot from this tagging behaviour as they are able to discover user behaviour at different levels to tailor services for them. Social tagging has become popular on the Internet as it provides an efficient way to organize and manage resources. The tags themselves carry useful information regarding the product: its characteristics, expression and more (Zheng & Quidan, 2011).

Boosting sales via online purchases is supported by better and more accurate recommendations. A good RS can be a reason to upscale or downscale the cart size. The recommendation system is categorized into three types – collaborative, content-based and hybrid approaches. The online players not having a huge database or warehouse depends on the content-based approach. The content-based approach sometimes leads to less than accurate recommendation results (Ziad & Najafi, 2016). Our research paper aims to address this issue and propose a revolutionary method to mitigate the weaknesses of content-based RS by combining it with social tagging based on crowdsourcing.

Studies suggest that 'crowdsourcing' has evolved as a methodology to gather data. The word combines 'crowd' and 'outsourcing' which indicates "more heads are better than one". Crowdsourcing is a way to receive feedback from the community for whom the products and services are being designed (Haren, 2017). In our research, we want to gather the crowd to aid our combination approach which will secure better results. Little has been researched on how to use the collective intelligence and leverage it to gain competitive advantage. In our research, we are targeting the online movie sales platform, where not necessarily all are well established (Sharma, Soe, & Balasubramaniam, 2014). Our experiments of a focus group with end user will help users formulate a model to evaluate and examine the criteria for social tagging and hence used by content-based RS.

LITERATURE REVIEW

This section focuses on the current best practices with respect to content-based recommendation system and tagging behaviour of users.

Recommendation Systems

With the tremendous amount of choices available in almost all types of products and services online, it is a tough decision for the users to select. Tailor-made solutions are the need of the time in this world of opportunities and personalizes recommendations are a win-win for both customers and the businesses. Once such personalized tool is 'recommendation systems'. Recommendation systems (RS) are the information tools that aim to predict suggestions, relevant searches or a filtering tool to the customers these days. It directly or indirectly hints the users of a web service to find content, products or services relevant to their search. The essence of this recommendation service is underpinned by user behaviour. Gathering and analysing this data helps in a more accurate recommendation which will lead for the user to purchase the product or service (Park, Kim, Choy, & Kim, 2012). RS helps the online retailers (e-tailers) especially in recommending right product or service to the right consumer at right time to ultimately increase the business.

Online-media services are not behind in adapting to this era of personalization. These media website caters to a diaspora of customers. The content they deliver is related to the most popular entertainment source i.e. 'movies'. One type of personalized service the movie websites are using is the recommender system. This RS is responsible to predict ratings, and suggest similar choices depending on user browsing history.

Amongst the popular types of RS, content-based recommendation system (CBRS) approach is one of the most popular RS as it uses user preference and descriptions of the items to match with the items of the users. This approach is accurate as it considers active users and can recommend accurately even if the data from other user ratings are not large. There are drawbacks to this approach like CBRS cannot recommend items if it does not have enough information on the metadata of the item profile. There is also over specialization problem which arises due to recommendations based on selective attributes only (Son & Kim, 2017). For example, while choosing a fiction movie the banner name like Marvel or DC is more important than the director. But if more appropriate information on categorization is not available then the choices recommended are biased. There have to be more categories like actors, characters present and more. In a similar way, CF has few drawbacks like sparsity and cold start problem. Also, CF is vulnerable to frauds or fake profile injections. If the number of users' increases the CF approach degrades as it cannot recommend accurately, hence it has to be continuously upgraded (Ziad & Najafi, 2016; Son & Kim, 2017).

Keeping in mind the drawbacks of both collaborative filtering and content-based approaches in RS, content-based approach is more effective and useful in our research as we are introducing tag-based system aided from crowdsourcing platform. This type of system will aim to mitigate the drawback of CBRS. Currently, many leading movie sales platforms like Netflix, Rotten Tomatoes, IMDB and more are using content-based approach effectively (Bergamaschi & Po, 2015).

Figure 1 shows the content-based approach explained by a diagram for movie recommendation system. The diagrams explain how tags and similar ratings help in recommending a movie to a user. It shows that if more information is collected under user profile and item profile, they can be matched efficiently. Research has shown that explanations help increases user satisfaction and acceptance of recommendations (Vig, Sen, & Riedl, 2009).

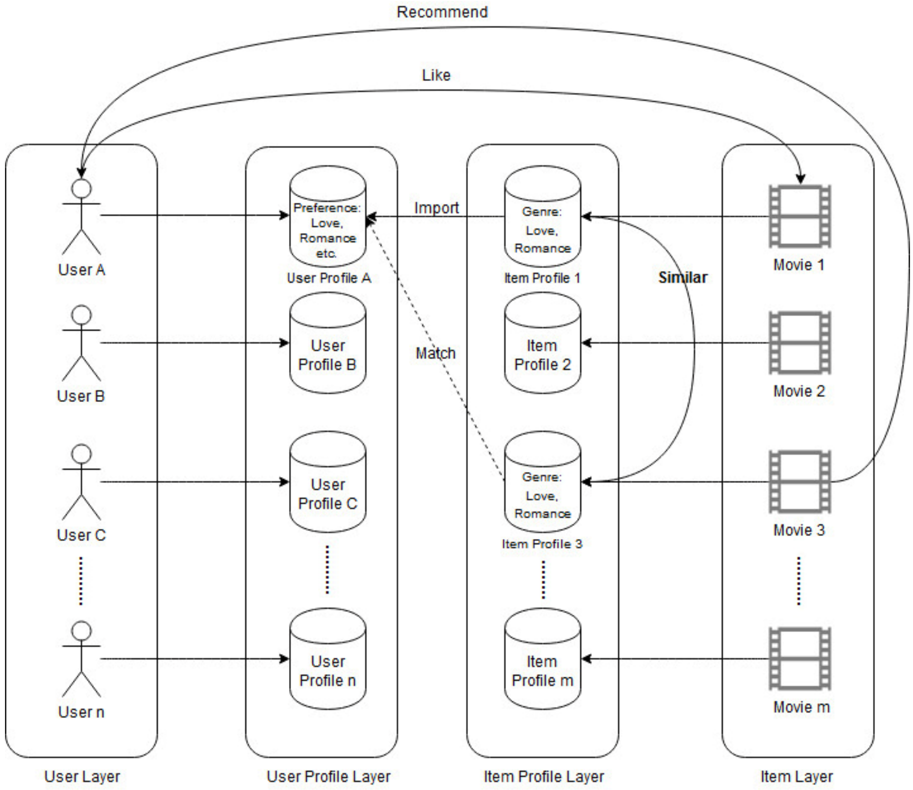


Figure1. Content-based recommendation system schema

Tagging and Crowdsourcing

In the ever-evolving era the users have become smart and, in most cases, they are the content generators. Crowdsourcing is nothing but filtered user generated content (UGC). When an appropriate content of high quality is filtered it is called as 'collective intelligence'. 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 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 earliest. Tags have various characteristics attached to it like in Figure 1, the tags explain the 'film features' of the movie like *love or romance*. This helps the user to figure out themselves about what they need (Vig, Sen, & Riedl, 2009). It will be tedious on the e-tailers to have a separate army to generate tags and keep it up to date. The reason for calling them army is the online content changes daily, the trends, the types of movies, genres are new and changing rapidly. It will be a tough grind to keep updating the metadata too. Hence, crowdsourcing the tags and giving the baton in end users fingertips to generate tags will be efficient (Zheng & Quidan, 2011; Vig, Sen, & Riedl, 2009). Hence, many researchers suggest on using collective intelligence as a tool to annotate and then use systems like Amazon Mechanic Turk for multiple tagging in less time and low cost (Nguyen-Dinh, Waldburger, Roggen, & Tröster, 2013; Hsueh, Melville, & Sindhwani, 2009).

One does not have to be an expert to create tags. Hence, crowdsourcing can be perfect to also increase user-system trust in recommender systems and understand user's perspective on presentation, explanation, and priority on the website (Berkovsky, Taib, & Conway, 2017; Vig, Sen, & Riedl, 2009).

Content based recommendation system

Content-based RS are more inclined towards user preferences and ratings. It recommends items in-sync with the interest demonstrated by user in the past. CB approach analyses the underlying information, for example in Figure 1, if User A has searched and surfed movies on 'Love' or 'Romance' then similar movies will be recommended.

A significant advantage of CB recommendation system over CF recommendation system is that it does not match one user's preferences with another, each user is unique and the user profiles are matched with their own past browsing and behaviour history online (Uluyagmur, Cataltepe, & Tayfur, 2012). In a story-telling business, each user is unique and to suggest choices depending on others' preferences can be termed as 'none of the above category' and can lead on losing customers (Gomez-Uribe & Hunt, 2015). To earn loyalty from the users, trust and accuracy are the building blocks.

As mentioned earlier social tagging has become more of a need than choice. Amazon Mechanic Turk is a crowdsourcing tool to generate tags and research suggests that the accuracy rate as high as 76% (Nguyen-Dinh, Waldburger, Roggen, & Tröster, 2013). The reason for us to believe in tags is that the content on the internet is dynamic and changes rapidly. With the likes of Flickr, Amazon, and more tagging helps a user in understanding the exact items or needs of their search (Vig, Sen, & Riedl, 2009; Vig, Sen, & Riedl, 2009). Tags are those self-explanatory metadata words that will enrich the film features section. CBRS relies heavily on the data categorization part and hence, enriched item-profile is useful.

For this social tagging, the user is generating the tags and not the IT developers of the website. This is user generated content and then we intent to generate 'criteria-based model' to support this idea. This will positively impact the user-system trust and user experiences. Success of the tagging will highly depend on the user-system trust as this will trigger the decision-making process for users (Berkovsky, Taib, & Conway, 2017).

Design Challenges

Content based recommendation systems

The effectiveness of content-based recommendation system depends on the metadata. The data categorization is crucial for the CBRS algorithm to suggest accurate results. Currently, insufficient information of item profile is a major concern (Ziad & Najafi, 2016). This leads to recommending biased choices to the user (Son & Kim, 2017).

Crowdsourcing and tagging

Willingness of users of the movie website to contribute in tagging is the most critical element of crowdsourcing. The encounter of users with systems has three dimensions of trust, dispositional trust dealing with cultural situational trust is more about the work and demographic factors, situational trust dealing with intricate tasks involving systems and learned trust that develops when user interacts with the systems and builds on the experiences (Berkovsky, Taib, & Conway, 2017). The workers are untrained that will also affect with the accuracy of the tags created, the explanation and presentation of the data. Hence, to establish criteria to maintain the user-system trust for item tagging and combining those elements with the CBRS is critical.

Crowdsourcing is excellent to collect the data real time, but it also has the issue of an incompatible system available. This means to combine the real-time data and collaborating it at the same time is not easy (Gao, Barbier, & Goolsby, 2011). Identifying a criterion for the acceptable tags is a persistent issue, which requires developing a framework which will explain specific criteria and rules. Eliminating the noise level and ambiguous tags are essential (Nguyen-Dinh, Waldburger, Roggen, & Tröster, 2013; Milicevic, Nanopoulos, & Ivanovic, 2010).

Synthesis: The CIST approach

Crowdsourcing can typically establish a task in no time with the huge pool of taggers. In perspective, this data is created by low-commitment workers and hence cannot be trusted fully. AMT (Amazon Mechanical Turk) has emerged as a low-cost option and has been effective accuracy of 76% to 92%. AMT is ranked as the most popular crowdsourcing platform (Nguyen-Dinh, Waldburger, Roggen, & Tröster, 2013). Hence, it will be an achievable option to use crowdsourced tagging to enrich the item profile and ultimately to increase the accuracy of the content-based recommendation system.

Berkovsky et.al in 2009 came up with criteria on basis of crowdsourcing viz. presentation, explanation and priority. These criteria are utmost important to establish the user-system trust and positive user experience. Netflix sited the issue by proving that the consumer loses interest after 60-90 seconds of browsing and reviewing from 10-20 movie titles (Gomez-Uribe & Hunt, 2015). From a back-end perspective and pondering on the soft system methodology root definitions, adding trust and accuracy as a part of this criteria model will enhance the stickiness of an end user on the website. Hence, to accomplish accuracy and solve the data categorization problem, my Criteria model has five main criteria for the recommender algorithms to run efficiently, viz. Accuracy, Trust, Presentation, Explanation and Priority.

It is conjectured that the Criteria model along with the Accepted Tag model can resolve the issue of item profile attributes and accuracy of content-based recommendation system.

DESIGN RESEARCH APPROACH

The Soft System Methodology (SSM) is a mature approach for action research that help clarify ill-defined system requirements (Checkland 1995, 2000; Staker, 1999). It has been applied to as diverse a design challenge such as the Concorde and the National

Health Service. Specifically, SSM originated from a base in systems engineering and is a strand of research which "concentrates on situations in which people are trying to take action" (Checkland 2000, pS41). Since the research being reported in this paper proposes action that may be taken to increase opting-in, SSM was deemed appropriate for investigating design rules back-end design for recommendation systems. A key feature of the SSM is the expression of key requirements through the drawing of rich pictures, which follow the stages of the model's root definition and model's framework design (Checkland & Scholes 1999). The resulting artefact is then validated by testing in a production environment. The model validation in this case means conducting stakeholder interviews that examine whether the proposed blockchain-based system can improve the service levels.
Overall, the SSM consists of four stages (Checkland 2000) and seven steps (Checkland 1995), descriptions of which are beyond the scope of this paper. However, the key milestones of: i) finding out about a "problem situation" (using "rich pictures"); ii) formulating some relevant "purposeful activity" models; iii) using the models for debate and discussion (including considering changes that would be desirable and feasible as well as accommodations which will allow such changes to be acted upon); iv) taking action to improve the situation. It should be noted that the milestones of SSM are wholly consistent with processes of Design Science Research: problem identification and motivation, definition of the objectives for a solution, design and development of artefacts, demonstrations of prototypes or proofs-of-concept, evaluation against key success criteria, and communication of the design specifications to stakeholders (Peffers et al., 2008). Checkland (2000) in a retrospective journey acknowledges the holistic consistencies between SSM and other offshoots of Design Science Research such as Systems Theory and Design Thinking.

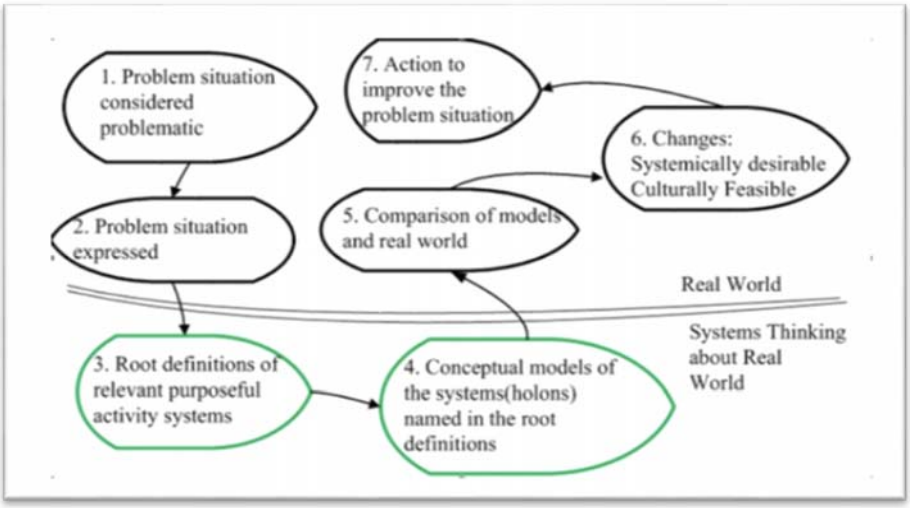


Figure 2. Soft system Methodology Overview (Source: Checkland and Scholes, 1990, p.191-242)

Model validation is a significant step 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 SSM, relevance is the key aspect model validation (Checkland, 1995). It can be learned through listening and analysing the related stakeholders' views and perceptions about the situation and the model. 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). Our design input was collected from actors in different roles in for a comprehensive insight. By drawing from the guidelines of design research, where experts and novices are sought for their comments and insights (cf. Conrath & Sharma, 1993), we conducted such design interviews to iteratively improve our design artefacts. The design interview template and guidelines may be found in the Annex to this paper.

Interview Protocol

The participants for the research are divided into three groups viz. IT developers – participants having prior knowledge of programming and working experience in recommendation systems, lead users – participants with experience in recommendation engines when using E-commerce website, and Novice users – participants fairly new to using E-commerce websites and having less knowledge on RS.

IT Developers Interview

In the pursuance of CIST and Accepted Tags Model, we will be conducting interview to get comments from IT developers. As these participants are highly knowledgeable in the area of recommendation engines, we will be interviewing five IT developers. The interview will be about 40 minutes in total.

Table 1. CATWOE for CIST system

Element	Definition
Customer (C)	Online movie sales platform
Actors (A)	Online movies sales platforms, end users, analysts
Transformation process (T)	In crowdsourcing the workers are untrained and hence a model based on important criteria will be developed for ‘Accepted Tags’
Weltanschauung (W)	This system will help the CBRS to show more accurate recommendations and use collective intelligence to positively influence user satisfaction
Owner (O)	All the users who will be one of the data source to use social tagging
Environmental constraint (E)	To develop a model to accept the tags and set rules to continuously update as trends and technology changes daily

The interview will be conducted with the SSM philosophy to delve deep into systems thinking and the real world. We will be conducting semi structured open-ended interviews. The questions are structured to guide CIST and test its feasibility on back end i.e. programming side. The rich picture and CATWOE in Table 1 will guide our research in developing a new approach in RS involving crowdsourcing and tagging to increase the accuracy of content-based recommendation systems as research showed shortcomings to the current approach like less data categorization and less features.

Our proposed solution is for those online movies selling platforms using content-based recommendation systems. Amount of data available is the primary concern for them. The interview will reflect on our empirical research. The interview will start with explaining the Rich picture depicting the process of how standard RS works. The questions focus on three categories of evaluation to support the criteria model that are recommendation system accuracy, user satisfaction and user-system trust.

Lead Users and Novice Users Interview

In order to collect feedback from the actual end users we will conduct semi-structured interview. For the convenience and detail understanding the users are divided into two categories viz. Lead users and Novice users. The purpose of dividing users is to treat lead users as the ‘producers’ of tags and novice users as ‘consumers’ of the tags. The interview will be conducted for 25 minutes.

The interview will be conducted with the SSM philosophy to delve deep into systems thinking and the real world. The interview will involve 2 tasks before beginning of the questions and answers. Task 1 is to browse Tumblr and click on 2-3 tags relevant to the shown image. Task 2 is to browse Amazon or Netflix. This will give the participant a chance to understand what recommendations mean. The questions asked are to understand whether users like tagging, their thinking about presentation of the suggestions, trust on user-generated material online and accuracy.

The sequence of the interview questions is based on the technical process of Browsing-Ordering-Confirming. This is a typical process a website desires each user to go through. The feedback will help our proposed solution to be refined and practical to implement.

Design Research Steps

Step 1: Preamble

This is the first step to introduce one another and providing information about the research. This step includes a brief description of the purpose to conduct this design experiment

Step 2. Rich Picture

Let me show you a diagram elaborating about the topic of interest today. This picture will give you a general idea on today’s discussion. we are really glad to talk to you today about recommendation systems. (If they say, ‘I don’t know how much we can help?’ – showing a bit doubt and unclear on the topic)

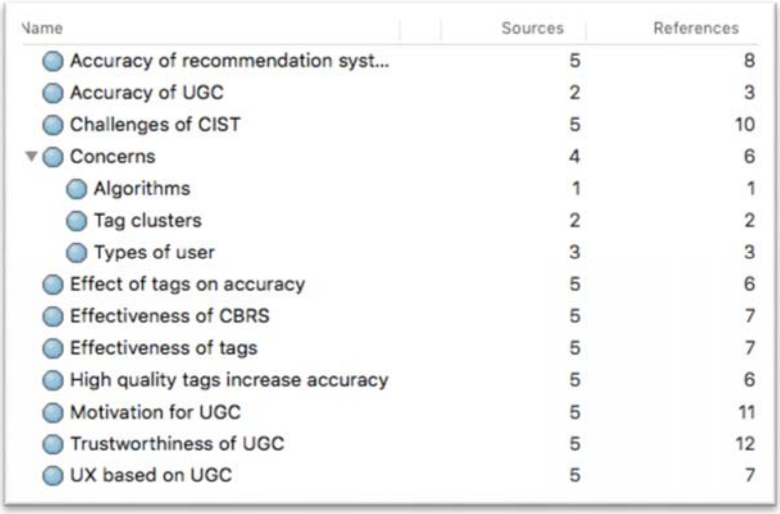
Step 3. Conducting the Interviews

The participants for the research are divided into three groups viz. IT developers – participants having prior knowledge of programming and working experience in recommendation systems, lead users – participants with experience in recommendation engines when using E-commerce website, and Novice users – participants fairly new to using E-commerce websites and having less knowledge on RS. (Please see Annex)

DISCUSSION AND ANALYSIS RESULTS

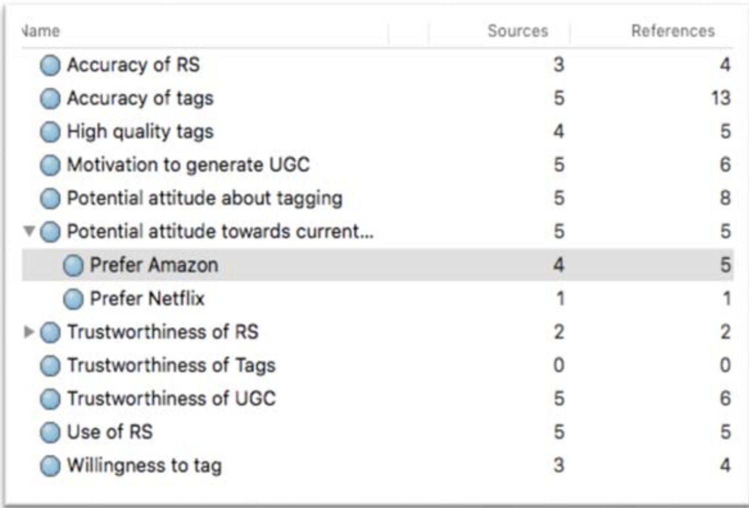
High-Level Consolidation of Design Feedback

After finished the interviews with IT developers, lead and novice users, we did transcription for all the participants' audio recordings. As mentioned earlier, the questions in the interview belong to particular design statement and, any design statement has more than one questions. Hence, we used NVivo for Mac version 11.4.3, a qualitative data analysis software to sort out the themes from all the transcriptions. Figures 3 and Figure 4 shows all the themes for IT developers and user groups respectively created in NVivo software.



Name	Sources	References
Accuracy of recommendation syst...	5	8
Accuracy of UGC	2	3
Challenges of CIST	5	10
Concerns	4	6
Algorithms	1	1
Tag clusters	2	2
Types of user	3	3
Effect of tags on accuracy	5	6
Effectiveness of CBRS	5	7
Effectiveness of tags	5	7
High quality tags increase accuracy	5	6
Motivation for UGC	5	11
Trustworthiness of UGC	5	12
UX based on UGC	5	7

Figure 3. Themes for IT developers



Name	Sources	References
Accuracy of RS	3	4
Accuracy of tags	5	13
High quality tags	4	5
Motivation to generate UGC	5	6
Potential attitude about tagging	5	8
Potential attitude towards current...	5	5
Prefer Amazon	4	5
Prefer Netflix	1	1
Trustworthiness of RS	2	2
Trustworthiness of Tags	0	0
Trustworthiness of UGC	5	6
Use of RS	5	5
Willingness to tag	3	4

Figure 4. Themes for Lead/Novice users

After the themes were sorted, the questions and design statement were allocated with the themes affiliated to them (Figure 5 and 6 below).

IT Developers			
Design statement	Questions	Themes	Main theme
1	1,2,3	T1,T2,T5	Effectiveness of Content-based recommendation system
2	4	T4,T6,T7	Social tagging as a solution for recommendation system
3	5,6,7	T6, T7, T8,T10	Attitude about user generated tags for recommendation system
4	8	T7,T10,T11	User generated tags more accurate

Figure 5. Affiliation of themes for IT developers

Lead/Novice user				
Design statement	Questions	Themes	Main theme	
1	1,2,3	T1,T4,T5,T6	Acuracy level of recommendation system	
2	4,5,6	T3,T6,T8,T9,T10	Attitude about User-system trust	
3	7,8,9,	T1,T2,T3,T6,T9	Personalized recommenations preferred	
4	10,11,12	T1,T2,T9,T10	Acceptance of tagging in recommendation system	

Figure 6. Affiliation of themes for User groups

The main purpose of conducting IT developers, Lead users and Novice users differently is to validate our design statement to understand the feasibility and challenges of CIST, user-system trust parameters and attitude of user generated content. The challenges and concerns raised by the participants' will be important reference for us to refine our proposed solution.

Validation of Design Statements

The design statements consisted of particular themes derived from the NVivo data analysis tool. In order to validate them, sentiment of those transcriptions had to be performed. We used IBM Watson's Natural language Understanding online tool to perform the sentiment analysis. The design statements act as a guideline to find out the underlying concerns, challenges, new ideas, about recommendations system, tags and user generated content. Themes such as, Concerns (T4) from IT developers' feedback will not be used for sentiment analysis as it talks about the challenges of CIST, the algorithm and user types. Similarly, themes like Preference of Amazon/Netflix (T6) derived from user groups help in understanding the popularity of the recommendation systems, hence, will not be considered for sentiment analysis.

The sentiment in IBM Watson tool is a numeric representation of negative, neutral and positive attitude of the text. The maximum and minimum scores are '1' and '-1' respectively. The sentiment scores for themes of IT developers are shown in Figure 7. Similarly, Lead and User groups is shown in Figure 8.

IT developers											
Theme	T1	T2	T4	T5	T6	T7	T8	T9	T12	T13	T14
IBM SCOR	0.64	0.18	0.08	0.74	0.67	0.43	0.79	-0.05	0.38	0.04	0.59

Figure 7. IT Developers' Sentiment Scores

Themes	T1	T2	T3	T4	T5	T7	T8	T9	T10	T11	T12	T13
Lead users	0.46	0.40	0.72	0.57	0.44	0.54	0	0.55	0.78	0.77	0.69	0.72
Novice users	0.24	0.55	0.40	0.37	0.57	0.68	-0.02	0.26	0.33	0.77	0.59	-0.28

Figure 8. Lead and Novice Users' Sentiment Scores

Effectiveness of Content-based Recommendation Systems

IT Developers Design statement 1 talks about content-based recommendation system. According to Figure 5 the themes affiliated to this design statement are T1, T2, and T5. Figure below shows the scores on a pie chart to clearly show that IT developers trust the CBRS and overall the sentiment about accuracy of recommendation system is positive. But, score for T2 shows less trust on user-generated content. Thus, in general, IT developers' feedback shows trust and accuracy on the functionality of content-based recommendation system but hesitation on the accuracy of user-generated content and item profile enrichment. Hence, the current state of content-based recommendation can be improved by increasing the trustworthiness of user profile and item profile.

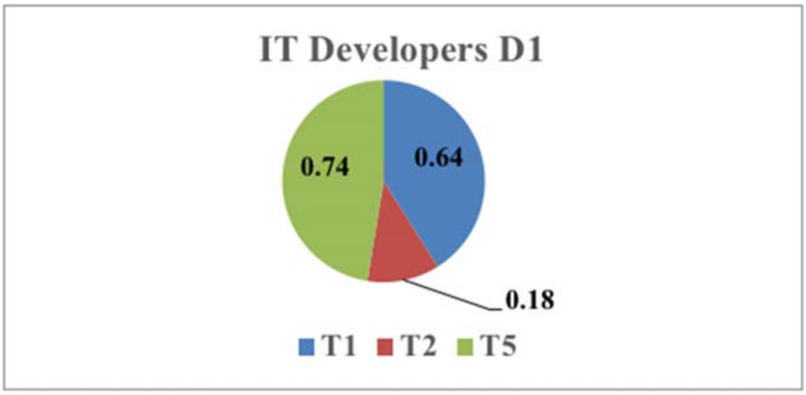


Figure 9. Sentiment Scores for D1

Social Tagging as a Solution for Recommendation System
Social tagging is common in social media. My Design statement 2 is an attempt to understand how effective it can be as a solution for Movie recommendation engines. This design statement is a part of CIST to understand its feasibility of tags. As shown in Figure 10 the scores are drastically different. T4 is almost towards negative score which depicts that all tags may or may not be the best keywords to describe the film features. However, when we look at the entire theme of D2, it shows that the feed from developers points towards 'high quality tags'. Thus, the IT developers' attitude towards tagging establishes that without any measures taken to eradicate tags, it cannot be a feasible solution. The effectiveness of tags will increase when only high-quality tags will be allowed. Also, high quality tags will ensure users will believe more and develop more trust on the website. One more point to be mentioned here is from the research by Berkovsky et.al (2017), where the Star ratings like that of IMDb ratings is a popular way to attract customers. The sense of the ratings given by 'one of them' for the users is crucial. Hence, tags can improve the functionality of the recommendation system but with the depiction of Star Ratings is compelling.

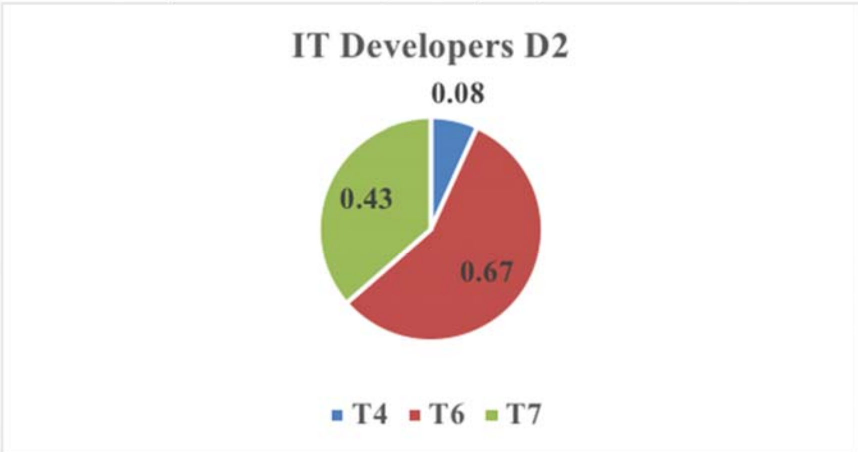


Figure 10. Sentiment scores for D2

Attitude About User Generated Tags for Recommendation System
Design statement 3 is about the potential attitude of IT developers with respect to user generated tags for recommendation system. All the four themes here are inter-related. The most shocking result was that of T8. IT developers think that users will have no motivation to generate content unless some incentive is attached. But they also think that the functionality of RS will increase if the high-quality user-generated tags are allowed as users are the best judge of the content and features of the movies. But accuracy of the RS results is something which will depend on the allowance of tags. The whole solutions of CIST become feasible if high-quality and user-generated tags are allowed. The accuracy of the engine will surely multi-fold. Thus, according to IT developers there is a positive side to CIST if quality is controlled. Essentially, CBRS can work more effectively as high-quality tags means enriched Film features, in technical terms enriching 'item profile'. (Figure 11 below)

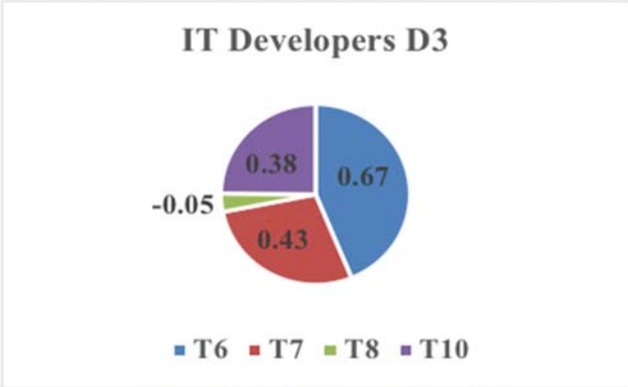


Figure 11. Sentiment scores for D3

User-Generated Tags are More Accurate

Recommendation systems depend upon user behaviour online. The more a user interacts with the website, the more patterns of his/her behaviour are saved. The more data available, the more accurate and trustable suggestions can be given. One of the important purposes of RS is to increase customer trust and loyalty. Trust and loyalty depend on numerous things like, quality of content, customer connect to the content, ratings, presentation, features, filtering options and more in case of an online Movie website (Figure 12 below)

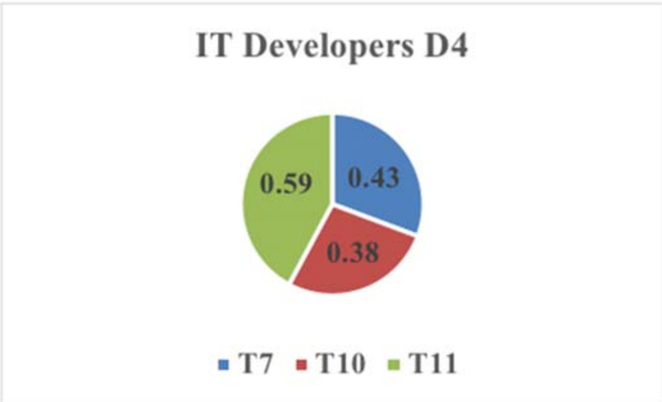


Figure 12. Sentiment scores for D4

Satisfaction Level of Recommendation System for Decision Making

Unlike IT developers' the user group has a different lens to evaluate or give feedback for recommendation system. User groups treat the recommendations system as a tool to browse more options and to narrow down choices using filter. To understand how user groups, define accuracy we had to probe them on their process of watching a movie online. The overall process for both lead and novice users of shopping online or watching movie is somewhat similar. But few differences can be pointed out like, in Figure 13, T1 scores for both lead and novice users are similar but novice users believing the RS is more accurate than lead users. For T2, the novice users believe more in RS than lead users in terms of trustworthiness. Here, the experience of the two groups in using information system as a whole comes handy. The lead users are aware about how RS works based on their browsing history and preferences, but novice users aren't aware on the technicalities of RS and they trust on the face value.

Figure 13. Sentiment scores for D1

The purpose of D1 is to understand how use of RS is more satisfying and assists in decision making. The themes erupted in the answers of questions in D1 have accuracy, trust and experience in using RS. Out of the three parameters UX is the most positive, meaning the presentation, ease of use, features on the website is appealing. The user-system trust is important when examining the effectiveness of the RS. Satisfaction level of RS depends as much on trust and accuracy as on UX. Thus, to establish user-system trust, accuracy and trust plays a key role. More of this was explored in D2.

Attitude About User-System Trust

In order to evaluate CIST's main sub-part 'user-generated tags', the questions in D2 were designed to understand the trust level to generated tags on online platforms, willingness to generate tags, and the quality of the UGC. From Figure 14, T8 and T10 are the most important themes. T8 which talks about trust on UGC is less in Novice users. The one definitive reason from the transcriptions was the inexperience with the website review platforms and generating content. Also, to note was, when addressed the concerns of UGC tags with the 'Accepted Tag Model' the trust factor increased. Hence, in order for CIST to work, building on the user-system trust is the factor of accuracy and trustworthiness of UGC in addition to UX.

The most important finding from D2 was derived from T10, which talks about 'willingness to generate tags'. The enthusiasm to contribute in tagging was blunt negative (-0.28) amongst the novice user group. The reasons for this were their inexperience with the systems, confused with respect to the concept of tags, and unclear of what is expected. On the other hand, sentiment for lead users was positive with a score of 0.72. The reason is aplenty given the experience and the zeal to contribute tags, or curiosity to know about new trends and more.

Keeping in mind the concept of CIST, it needs to be fed with tags and there have to be consumers to use those tags. CIST depends on user-generated tags, hence there is a need for users to keep generating more tags, ultimately enriching the 'film-features'. Thus, the concept of 'lead users' being 'tag generators' and 'novice users' being 'tag consumers'.

Figure 14. Sentiment scores for D2

Acceptance of Tagging in Recommendation System

Design statement 4 consisted of the questions to understand acceptance of tags as a way to understand about a movie and convenience in browsing through suggestions. From the Figure 15, the overall concern from users was the trust and accuracy of the tags. The concept was luring to the users but the uncertainty of accuracy of those generated tags and the sureness of UGC could be sensed. As mentioned above, novice users where negative about contributing to tags as well. Over all, for the Trust and Accuracy formed a big part of users' interaction with the system. Thus, it is good to establish that user-system interaction is based on 'trust' and 'accuracy' a lot.

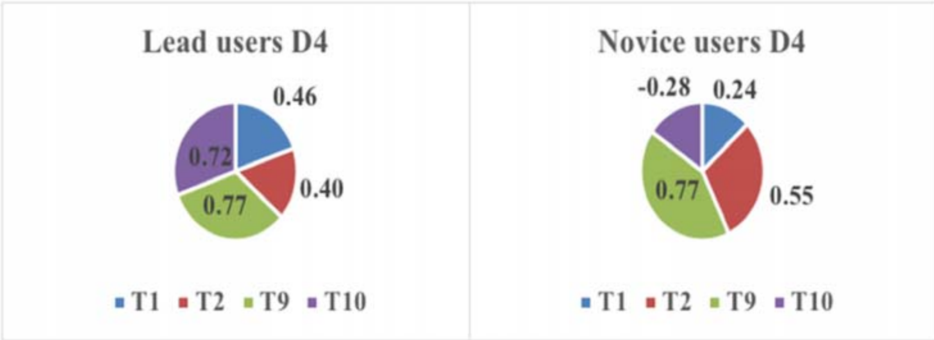


Figure 15. Sentiment score for D4

Surprisingly, security and privacy concern of the recommendation system is one of the concerns that users had. Issue like what data will be stored or shared if they contributed in online content was one main questions. As security and privacy issue was not in the scope of this research, the answer was not available.

FINDINGS AND RECOMMENDATIONS

From the analysis of the design interviews described in Section 4, the following are key findings with respect to the design statements.

1. The current state of content-based recommendation can be improved by increasing the trustworthiness of user profile and item profile

2. Tags can improve the functionality of the recommendation system but with the depiction of Star Ratings

3. Tags will definitely enrich the film features but *accuracy and trust* on user generated content is doubtful

4. More '*personalized suggestions*' are feasible if tags are distinguished clearly

5. Recommendation system is a trusted system for online shopping

6. The suggestions currently are effective in presentation, explanation and prioritization but trustworthiness and accuracy are utmost

7. Tags are effective ways of depicting personalized recommendations, but '*information overload*' has to be avoided

8. Lead users are willing to generate tags, but novice user is not in complete favour

9. Security and privacy concern of the recommendation system and the CIST approach

From the above key findings, three significant issues that could be highlighted are: trust and accuracy of UGC, personalized suggestions and information overload. For CIST to avoid this and improve the CBRS functionality we suggest the following three recommendations.

Clustering of Tags to Avoid Confusion

One of the most important concern from IT developers and User groups was that of information overload. CIST relies on user-generated tags, the number of tags will also grow as the tags are personal preferred words which will differ for each user.

Since users will be generating tags, there is a high chance that there will be synonyms, vernacular words, confusing meanings and so on. For example, for a movie *The Avengers*: the film features can have many tags like *superpower, super-natural, mission, objective, pursuit, etc.* Similarly, *Justice League*: *woman-power, strong, superman, etc.* can be added. The main problem at times the recommendation engine might face is, movies related to what content can be suggested. Hence, the solution of 'clustering the tags'. As in Figure 16, the similar movies with most common tags can be clustered. Randomness and personalization of tags is common, especially in UGC. This leads to issue of information overload, crunch in data space and semantic obstacles. For this kind of solution, three main algorithms can be used viz. MMSK (min-max similarity *k*-means), tag clustering based on semantic analysis, and LMMSK (latent min-max similarity *k*-means) (Yang & Wang, 2017).

The preferred option here can be 'tag clustering based on semantic analysis as those colloquial words, synonyms can be all bundled for better results. CIST is as of now only designed keeping in mind the movie selling platforms online. Hence, Film feature is the most important element of the content-based RS. The more item profile is built on the better will be the result.

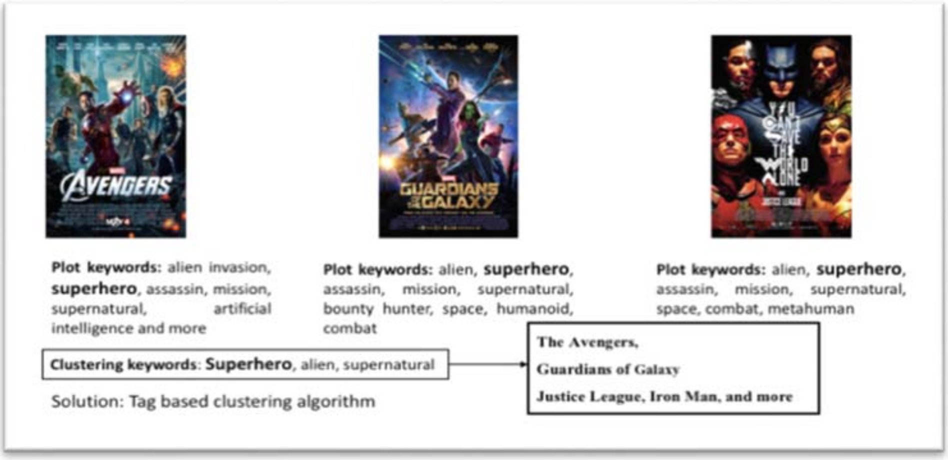


Figure 16. Clustering of tags

Five Criteria of User-System Trust

Commercial recommendation engines owe their success to the users to a great extent. It is a two-way communication, the more your exceptional data you offer on plate, and the more the customers will spend time to keep the uptake. Now, the success also depends on parameters like the freshness of content, the precise suggestions, but the most important is promising value to the customer (Berkovsky et.al, 2017). Users these days, need personalized suggestions, need more data to filter, hassle free appearance on the website. During the data analysis it was established that the correctness of the content is directly proportionate to the trustworthiness of it. Hence, we synthesized the UX factors from the previous research and added two more criteria broadly outlining the user-system trust viz. accuracy, trustworthiness, presentation, explanation and priority. Figure 17 outlines the criteria to establish user-system trust and ultimately encouraging the users to contribute in tagging and provide more functionality on the website.

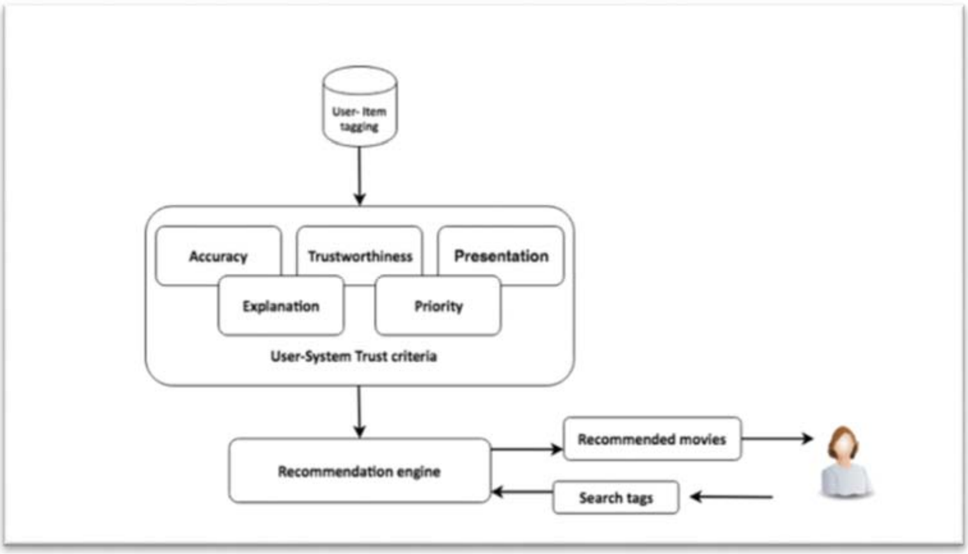


Figure 17. Proposed solution for user-system trust

As the research advocates the recommender engine in Movie website is hardly aid in taking critical decisions like, medicines, doctors, etc. The IT developers were aware of the fact that the users will have little domain expertise about some or the other movies flashing under recommended list. In this case, the RS needs to support decision making but also to keep the users engaged, use more functionality available to aid the decision. Hence, RS is more than just pushing to arrive at a decision. Keeping all this in mind accuracy and trust are of paramount importance for CIST as the main resource of tags are users' themselves. Hence, these criteria are of utmost value to keep the vigour in the users burning.

Clear Way of Distinguishing and Displaying Tags

To address the issue of personalised suggestions and target the users to engage more, the CIST approach needs to consider which tags will appear primarily. So, to avoid confusion a simple method of dividing the 'film features' into primary and secondary tags is essential. In Figure 18, the primary tags characterize the main theme of the film. The primary tag is different that the 'genre' of a movie. The 'Genre' explains the category to which the movie belongs. For example, *The Avengers*, this movie belongs to 'Science-Fiction' (referred as 'Sci-Fi' in future), but the tags like *Superhero* is its 'primary tag'. All the other plot keywords or film features describing the storyline or themes of the movie are 'secondary tags'. Also, to make it simpler, the primary tag can be highlighted font-wise to separate from the secondary tags.

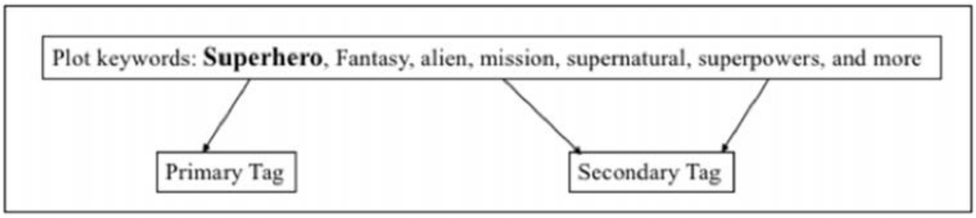


Figure 18. Primary and Secondary tags

During the interviews, one main piece of information that came forward was, to clearly display the more liked and relevant tags. The relevant tags after passing through the Accepted tag model, still needs to be eye-catching, intriguing for users to click on them. Hence, we found out a clear way to make this possible. In Figure 19, *the Avengers,* is a movie which can have 'n' number of tags, but not all tags can be eye-catching. To keep it interesting, interactive and relevant, we can have a system, where the users are allowed to click '+1' if they think the tag is relevant or '-1' if they think it's not relevant. This '+1' and '-1' system will also add to the most liked tags for the item profile and less crowded. The number of tags added to the system should also have a limit, but mostly it can be thrice as much as the current tags. This solution will also alleviate the role of CIST to be more likely the most fitting solution for CBRS.

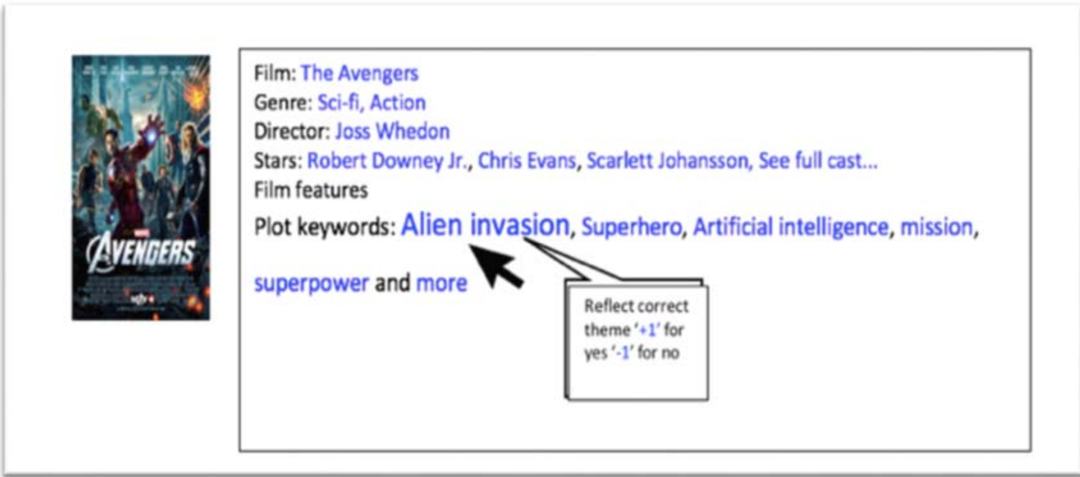


Figure 19. Displaying 'most preferred' tags

CONCLUDING REMARKS

This research paper aims to address the issue of 'less data categorization' in a content-based recommendation system by proposing a novel method known as 'Collective Intelligence Social Tagging' (CIST). This hybrid solution will increase the current functionality of the content-based approach and make a significant difference to its accuracy. The category selected for researching is the online movie sales platforms. The reason being, to recommend choices based on user's browsing history than that of his/her neighbours' search. To test the feasibility of CIST semi-structured interviews were conducted on IT developers and Users. The user group was further divided into lead and novice users to anticipate difference of opinions. The feedback received from both the group, developers and users gave tested the satisfaction level and acceptability of CIST. After careful analysis, following three recommendations were suggested in order to improve CIST to function better.

1. Clustering of tags to avoid confusion
2. Five criteria of user-system trust
3. Clear way of distinguishing and displaying tags

This research throws the light on the broader aspects like crowdsourcing, tags and recommendation system, put together the response collected was positive.

The scope of this research was to understand the acceptance and feasibility of CIST. But there are two main limitations: 1) Smaller sample size of design interviews. 2) Lack of working prototype for this design experiment of CIST. Down the line, it will be insightful and interesting to understand the experience and performance of a larger sample size. Also, to acquire more knowledge on the algorithm to code a working prototype will aid in any changes required.

REFERENCES

[1] Advomavicius, G., & Kwon, Y. (2007). New Recommendation Techniques for Multi-Criteria Rating Systems. *IEEE Intelligent Systems, 22*(3), 1-28.

[2] Bergamaschi, S., & Po, l. (2015). Comparing LDA and LSA Topic Models for Content-Based Movie Recommendation Systems. *Web Information Systems and Technologies* (pp. 247-263). Modena, Italy: Springer Link.

[3] Berkovsky, S., Taib, R., & Conway, D. (2017). How to Recommend? User Trust Factors in Movie Recommender Systems. *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 287-300). Limassol: ACM .

[4] Burge, S. (2015). An Overview of the Soft Systems Methodology. *System Thinking: Approaches and Methodologies*.

[5] Chartron, G., Kembellec, G., & Saleh, I. (2014). General Introduction to Recommender Systems. In G. Chartron, & G. Kembellec, *Recommender Systems* (pp. 1-23). London/Hoboken, UK/USA: John Wiley & Sons, Incorporated.

[6] Checkland, P., & Poulter, J. (2017). Soft Systems Methodology. In M. Reynolds, & S. Holwell, *Systems Approaches to Managing Change: A Practical Guide* (pp. 191-242). Springer International Publishing AG.

[7] Chenting, Z., & Chun'e , M. (2011, March 16). *Recommended engine to explore*. Retrieved from www.ibm.com: https://www.ibm.com/developerworks/cn/web/1103_zhaoct_recommstudy1/index.html

[8] Cooper, D., & Schindler, P. S. (2014). Ethics in Business Research. In D. Cooper, & P. S. Schindler, *Business Research Methods* (pp. 28-47). The Mcgraw-Hill/Irwin.

[9] 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.

[10] Faggella, D. (2017, October 30). *The ROI of recommendation engines for marketing*. Retrieved November 10, 2017, from www.martechtoday.com: https://martechtoday.com/roi-recommendation-engines-marketing-205787.

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

[12] Gao, H., Barbier, G., & Goolsby, R. (2011, May-June). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. *IEEE Intelligent Systems, 26*(3), 10-14.

[13] Gomez-Uribe, C. A., & Hunt, N. (2015, December). The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems, 6*(4), 13.1-13.19.

[14] Haren, R. V. (2017). Crowdsourcing feedback. *Literacy Learning: the Middle Years, 25*(1), 7-15.

[15] Hsueh, P., Melville, P., & Sindhwani, V. (2009). Data Quality from Crowdsourcing: A Study of Annotation Selection Criteria. *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing* (pp. 27-35). New York: HLT.

[16] Lakhani, K. R., Cohen, W. M., Ingram, K., & Kothalkar, T. (2014, August). *Netflix: Designing the Netflix Prize (A)*. (Harvard Business School) Retrieved November 9, 2017, from www.hbs.edu: http://www.hbs.edu/faculty/Pages/item.aspx?num=47650

[17] Milicevic, A. K., Nanopoulos, A., & Ivanovic, M. (2010, March). Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions. *Artificial Intelligence Review, 33*(3), 187-209.

[18] Nguyen-Dinh, L., Waldburger, C., Roggen, D., & Tröster, G. (2013). Tagging Human Activities in Video by Crowdsourcing. *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval* (pp. 263-270). Dallas: ICMR.

[19] 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.

[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] Soares, M., & Viana, P. (2015, September). Tuning metadata for better movie content-based recommendation systems. *Multimedia Tools and Applications, 74*(17), 7015-7036.

[22] Son, J., & Kim, S. B. (2017, August 3). Content-based filtering for recommendation systems using multiattribute networks. *Expert Systems with Applications, 89*, 404-412.

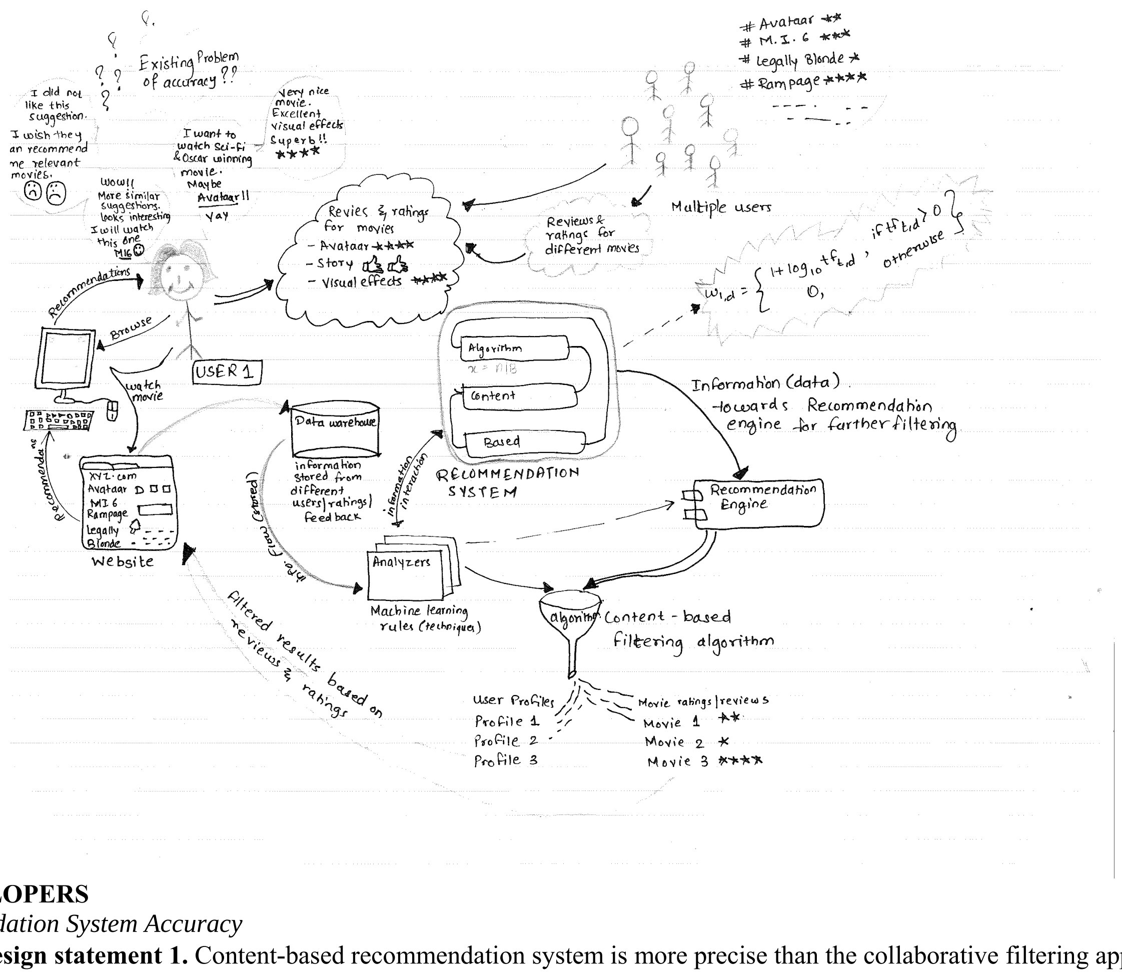
[23] Uluyagmur, M., Cataltepe, Z., & Tayfur, E. (2012). Content-Based Movie Recommendation Using Different Feature Sets. *Proceedings of the World Congress on Engineering and Computer Science* (pp. 17-21). San Francisco: World Congress on Engineering and Computer Science.

[24] University of Greenwich. (2015). *What might be the ethical issues and risks that arise in my research?* Retrieved October 30, 2017, from www2.gre.ac.uk: http://www2.gre.ac.uk/research/ethics/what-might-be-the-ethical-issues-and-risks-that-arise-in-my-research

[25] University of Southern California. (2017, October 9). *Organizing Your Social Sciences Research Paper*. Retrieved October 30, 2017, from www.libguides.usc.edu: http://libguides.usc.edu/writingguide/quantitative
[26] Vig, J., Sen, S., & Riedl, J. (2009). Tagsplanations: explaining recommendations using tags. *In Proceedings of the 14th international conference on Intelligent user interfaces* (pp. 47-56). Florida: ACM.
[27] Wang, Y. (2017). Content-based recommendation system: to tag or not to tag? *MBIS research project report*. Unpublished reseach report.
[28] Yang, J., & Wang, J. (2017, May 11). Tag clustering algorithm LMMSK: Improved K-means algorithm based on latent semantic analysis. *Journal of Systems Engineering and Electronics, 28*(2), 374 - 384.
[29] Yeoman, I., McMahon-Beattie, U., & Wheatley, C. (2016). Keeping it pure – a pedagogical case study of teaching soft systems methodology in scenario and policy analysis . *Journal of Tourism Futures, 2*(2), 175-195.
[30] Zarei, B., Azizian, S., Shishehgar, M., & Sadeghi, S. (2014). Networking in Marketing and Advertising Companies: A Soft System Methodology Approach. *The International Technology Management Review, 4*(1), 42-57.
[31] Zheng, N., & Quidan, L. (2011). A recommendation systembased on tag and time information for social tagging systems. *Experts Systems with Applications, 38*, 4575-4587.
[32] Ziad, S. P., & Najafi, S. (2016). Evaluating Prediction Accuracy for Collaborative Filtering Algorithms in Recommender Systems.

ANNEX: DESIGN INTERVIEW TEMPLATE

CONDUCTING THE DESIGN INTERVIEW



IT DEVELOPERS

Recommendation System Accuracy

- **Design statement 1.** Content-based recommendation system is more precise than the collaborative filtering approach

1. In your professional experience, do you agree with this statement? If yes, why so?
2. What about the trustworthiness? Do you think ratings and user profile are credible to increase precision in RS?
3. Can you elaborate on your experience of using recommended choices when you browse?

Explanation

- **Design statement 2.** Tags will make the RS more personalized and persuasive to the user

4. Which of the classification users prefer with persuasive, personalized or IMDb ratings?
5. What if movies were suggested with the tags alongside the IMDb ratings? How do you think they would help?
6. Tags will enrich the recommender engine to suggest more personalized suggestions. Agree? Content based RS is all about suggestions based on content. Please provide comments.

User-System Trust

- **Design statement 4.** User generated tags increases the confidence of users to review items.

7. 'The more tags about a product, the more you know about it'- Agree? Comments
8. What motivates users to engage in generating content i.e. tagging?
9. What are the implementing challenges in developing user generated based tagging system? (*quality, trustworthy, security, privacy, relevance*)

Presentation

- **Design statement 3.** Current functionality of RS is more precise in appearance of results.

10. Displaying tags like that of Tumblr can increase personalized suggestions. Do you agree? Why?
11. Generally, RS uses Genre, Human preference and Star rating to display the suggestions. Which method do you prefer and why?

LEAD USERS (PRODUCTION OF TAGS)

Task: Open Netflix and ask them to browse and see how the suggestions have come

Accuracy

- **Design statement 1.** Recommendations systems assist in decision making

1. Tell me about how you make your decision to watch a movie?
2. Do you think the recommended results influence your decision?
3. Tell me about your experience with RS in Amazon or Netflix? Why you prefer this one?

▪ User-system trust (Showing Tumblr)

- **Design statement 2.** Contributing to ratings and reviews increases the sense of involvement with after joy

4. When do you write feedback or reviews for items/products/services you use?
5. User generated content is more trustable than generated by the system. Agree? Please provide some comments
6. How about you get to generate tags for a movie? What will motivate you to do so?

▪ Explanation

- **Design statement 3.** User generated tagging is an effective way to receive personalize recommendations

7. Tags or no tags – which one will help in shortlisting the items you want to buy?
8. 'The more tags about a product, the more you know about it'. Agree? Please provide some comments.
9. What if movies were suggested with the tags alongside IMDb rating? How do you think they would help?

Presentation

- **Design statement 4.** Tags will be an effective appearance method on the website

10. Above picture is about the types of recommended results. Which one do you prefer? Why?
11. Recommendation results on search are preferred. Will tags help in narrowing the search to the relevant results?

NOVICE USERS (CONSUMERS FOR TAGS)

User-system trust (Showing Tumblr)

- **Design statement 2.** The preciseness of user generated Tags will motivate to become the part of Tagging community

1. Like the Tumblr website you just saw, would you like the same concept of tags in other ecommerce websites like Netflix or Amazon prime?
2. User generated content is more trustable than generated by system. Agree? Please provide some comments
3. Would you like to contribute to the content by creating tags movies?y