Independent Study Final Report

Background

The American Cancer Society (ACS), founded in 1913, is a nationwide, community-based voluntary health organization dedicated to eliminating cancer as a major health problem. In an attempt to provide support for people suffering from cancer, ACS has established an online forum, Cancer Survivors Network (CSN), where patients can receive considerable help by participating in various support groups. The CSN's effort is to encourage survivors to participate actively in the cancer recovery process by sharing their experiences with others who are going through similar things and to educate and encourage them to implement a comprehensive, integrated cancer recovery plan.

Research Goal

The research team at Carnegie Mellon University has previously built a deployed recommender system for CSN. Newcomers of the website are given the choice of whether to use the modified version of CSN to participate in the community. If users opt in to the new interface, they will receive recommendations for content and other members as well as surveys asking for their feedbacks regarding the new interface. The recommender system has four different settings, each associated with a specific algorithm (See Appendix A). This study aims to understand the effectiveness and underlying issues of the recommender system so that the team can improve the interface and thus encourage more forum members to participate in the discussions.

Method

I gathered information about the performance of the recommender system by examining the behavior of CSN users under the new interface. I used Navicat for data analysis and the

system was evaluated from two different perspectives – overall evaluation and evaluation by recommender setting.

Overall Evaluation

Overall evaluation aims to provide information about the general performance of the system. Below is a list of evaluations that I've completed.

- Overall Opt-in Situation: See how many users (both newly-registered ones and old ones who have updated their profiles) choose to use the new interface.
- Clicking Activity: Check the number of clicks each user makes in a single session.
- Web Form Submission: See how many users have submitted the survey and how valuable/relevant these threads are to them.
- Rating Feedback: Check to see how many users have provided positive feedbacks to certain threads.

Evaluation by Recommender Setting

The performances of recommender setting #3 and #4 are compared to the control group – recommender setting #2. Below is a list of evaluations done under each recommender setting.

- Hitting Evaluation: See the overall number of recommendations displayed under each setting as well as calculate and compare the hit ratio.
- Recommendation by Type: Calculate the hit ratio under different types of recommendations, including login, browsing and similar-users.
- Thread Creating/Replying: Calculate the average number of threads each user has created/replied to under each recommender setting.
- User Interaction: Calculate the average number of private messages each user has sent to other users under each recommender setting.

Results

Overall Evaluation

Overall Opt-in Situation. Since the CMU recommender system went live on Aug. 7th, 2017, it has maintained a steady opt-in rate at approximately 75%. This means that most users are willing to try out the new interface. However, according to the bar chart (See Appendix B), the number of new users registered each month has decreased a little.

Clicking Activity. The average number of clicks a user makes in a single session is 26.57 (See Appendix C), demonstrating that CSN is a content-based website where users have strong motivation to click into a variety of threads to obtain information. This evaluation is an indicator of whether users are interested in the content displayed to them, and thus can be used for examining the effectiveness of each recommender setting.

Web Form Submission. So far, CSN has displayed over 1600 pop-up surveys asking users whether a recommended thread is relevant/valuable to them and has received approximately 150 submissions (See Appendix D). The submission ratio is relatively low, indicating that people rarely reply to those surveys. Among those who replied, about 85% of the threads were rated positively mainly because they provided emotional support, or the stories described in the threads were similar to users' own experiences or their family members'. However, many users also pointed out that some threads contained outdated information but were shown repeatedly, which was quite annoying.

Rating Feedback. CSN has received 700 submissions on the ratings for some specific threads. The survey is displayed below the content during users' reading process (See Appendix E). In about 92% of the cases, the threads were rated 4 or 5 on a 5-point scale, indicating that users found the content to be interesting/relevant in general.

Evaluation by Recommender Setting

Hitting Evaluation. The table (See Appendix F) records the hit ratio under each algorithm. In general, users rarely click into recommendations displayed for them (with the maximum hit ratio being 2.82%). However, setting #3 performs significantly better than the control group in encouraging users to click into recommendations. The reason might be that, by displaying recommendations in any forum, setting #3 exposes users to a wider range of topics than the three and thus is more likely to include threads that attract users. Besides, setting #4 performs a little worse mainly because it filters threads based on users' browsing history and thus puts a limit on the content of recommendations.

Recommendation by Type. I've calculated the hit ratio for different types of recommendations, including login, browsing, and similar-users (See Appendix G). Among the three, login recommendation has the highest hit ratio. I believe the reason is that, when users have just logged in to the website, they are looking for content to read and have no preferences on where to start. As a result, if they find something interesting under login recommendation board, they are very likely to click into that thread. On the other hand, if users have already chosen some threads to read, they tend not to click into/pay attention to threads under browsing recommendation board because that will direct them to a different sub-forum. Moreover, evaluations for similar-users recommendation have shown that users rarely click into profiles of other users. This can be attributed to the nature of the website and the way similar users are displayed. CSN is more of a content-based community than a network-based one, meaning that users are interested in the information/experiences provided by others rather than the people themselves. Also, similar-users recommendation only shows a list of names without any description or even profile image, which doesn't attract users' attention at all.

Thread Creating/Replying. As shown in the tables (See Appendix G), the average number of threads each user has replied to is approximately five times the average number of threads each one has created, which implies that users are more comfortable commenting on others' experiences than sharing theirs. According to both thread-creating and thread-replying evaluation, setting #3 has outperformed the other three. The reason behind might be that, setting #3 displays recommendations from a variety of topics, not limited to a user's specific situation, and thus reminds users of things they didn't thought of as well as inspires them to talk more about their own stories. However, since the improvement is not significant, we cannot state for sure that setting #3 encourages users to participate more in the discussion.

User Interaction. The overall average number of private messages sent out by each user under each recommender setting is shown in the table (See Appendix H). Users under setting #3, compared to those under other settings, have more interaction with other users through sending out private messages. Again, since the improvement is not significant, we cannot conclude that setting #3 encourages user interaction.

Discussion

By combining the results of different evaluations, I've reached the conclusion that, in general, recommender setting #3 has the best performance among the four settings. However, it's hard to measure how much better setting #3 is compared to the other three. This is because CSN users generally don't click into recommendations very often, and all the significance tests are done on relatively small numbers.

Despite the fundamental drawback of the evaluations, I've found several issues that should be addressed during my study.

First, in 45% of the cases, a session ID is associated with more than one algorithm. That adds difficulty to the evaluation process because it's hard to decide which setting a user's action, such as clicking, should be accredited to. Therefore, only when the database becomes more organized and consistent can we accurately measure the performance of each setting.

Second, in the web form submission, comments and ratings for "how relevant" and "how valuable" are highly correlated – most people just gave the same answer to both questions. In order to avoid correlation and redundancy, only one question should be kept in the survey. I personally prefer to keep the question of "how valuable".

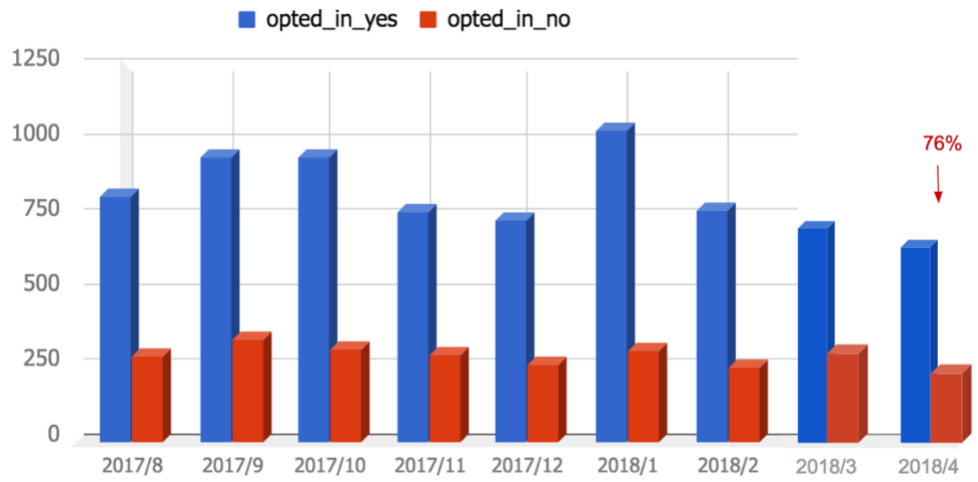
Finally, one major issue users mentioned is that same threads without updates are recommended repeatedly, causing lots of frustration. One solution could be that the system only recommends a previously shown thread to users when there is new update. In addition, an algorithm that better matches a user's situation with the content of the threads recommended could be useful for resolving issues of irrelevant or unhelpful information.

Appendix

A.

Recommender Setting
1: Cold Start: Recommend most popular threads for new users
2: Recommend recent threads in favorite subforum
3: CMU recommender in any forum
4: CMU recommender in favorite subforum

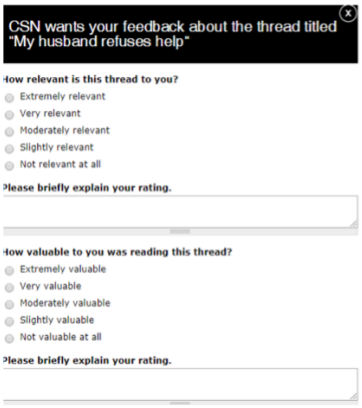
B.



C.

Number of Unique Session	Number of Clicks	Average Number of Clicks Per Session*
7797	207200	26.57

D.



1648 webforms displayed

152 submissions

9.22% submission ratio (↓)

Threads evaluated positively (↑)
(**85.9%** relevant & **86.96%** valuable)

*People rarely reply to webforms, and sample is too small to breakdown by recommender condition

E.

686 submissions
In **92.42**% cases, 4 or 5 scores on 5-point scale

F.

Recommender Setting	Displayed	Hits	Hit Ratio (%)	Improve over control (#2)
1: Cold Start: Recommend most popular threads for new users	67719	690	1.02	
2: Recommend recent threads in favorite subforum	23406	459	1.96	
3: CMU recommender in any forum	22541	636	2.82	+44% (s!)
4: CMU recommender in favorite subforum	23491	430	1.83	-7% (ns)

G.

Evaluation: Browsing Recommendation

Recommender Setting	Displayed	Hits	Hit Ratio (%)	Improve over control (#2)
Cold Start: Recommend most popular threads for new users	48710	101	0.21	
Recommend recent threads in subforums	9243	52	0.56	
CMU recommender	8932	53	0.59	+5% (ns)
CMU recommender + filtering	9623	93	0.97	+72% (s!)

Evaluation: Similar Users Recommendation

Recommender Setting	Display	Hits	Hit Ratio (%)	Improve over control (#2)
Cold Start: Recommend most popular threads for new users	17865	64	0.36	
Recommend recent threads in subforums	14036	25	0.18	
CMU recommender	13543	33	0.23	+37% (ns)
CMU recommender + filtering	13460	41	0.3	+71% (s!)

Evaluation: Login Recommendation

Recommender Setting	Displayed	Hits	Hit Ratio (%)	Improve over control (#2)
1. Cold Start: Recommend most popular threads for new users	19009	524	2.76	
2. Recommend recent threads in subforums	14163	381	2.69	
3. CMU recommender	13609	549	4.03	+50% (s!)
4. CMU recommender + filtering	13868	296	2.13	-26% (s!)

H.

Recommender Setting	*Average Number	Improve over control (#2)
1: Cold Start: Recommend most popular threads for new users	4.94	
2: Recommend recent threads in favorite subforum	6.35	
3: CMU recommender in any forum	6.51	+2.52% (ns)
4: CMU recommender in favorite subforum	6.41	+0.94% (ns)

*The overall average number of private messages sent out by each user