A recommender system for informal bibliotherapy

Jatin Aswal
Nisheeth Srivastava
jaswal@cse.iitk.ac.in
nsrivast@cse.iitk.ac.in
Dept of Computer Science, IIT Kanpur
Kanpur, UP, India

ABSTRACT

We present an online system that recommends web-based reading passages to users based on free long-form text-based elicitations of how they're feeling right now. The system combines natural language processing techniques used to extract users' intent with an information retrieval system to yield relevant and useful narratives for users. An eight week long user study found that most people who used the system reported better mood at the end of their interaction with the system. Interestingly, our study also discovered greater user engagement with randomly recommended narratives than with narratives selected for users based on their written descriptions of their own mental states. These observations could constrain the future design of self-help recommender systems.

CCS CONCEPTS

• Human-centered computing; **• Information systems → Content ranking**; **Personalization**; *Query intent*; **Recommender systems**; **• Applied computing** → Psychology;

KEYWORDS

bibliotherapy; personalization; human factors; recommender systems

ACM Reference Format:
Jatin Aswal and Nisheeth Srivastava. 2020. A recommender system for informal bibliotherapy. In *Proceedings of the 5th International Workshop on Health Recommender Systems co-located with 14th ACM Conference on Recommender Systems (HealthRecSys'20), Online, Worldwide.* , 5 pages.

1 INTRODUCTION

Bibliotherapy is the use of books as therapy for mental distress and depression [9]. While the therapeutic value of books has been acknowledged anecdotally in literature for centuries, it has recently been empirically demonstrated that careful bibliotherapy leads to significant and long-lasting alleviation of symptoms of depression [13]. These findings have been, by and large, supported by a large body of subsequent observations in clinical settings [6, 7]. Consequently, it has found a place as a popular therapy option in several mental health programs [3, 10].

The key psychological premise of bibliotherapy is that the reader begins to identify with a particular character in the book, and thus is able to observe a situation related to their own predicament from sufficient psychological distance. This allows them to think of possibly solutions, which they eventually realize might also apply in their own situation [9].

Thus, effective bibliotherapy requires insight into the patient's condition, and the ability to recommend readings that the patient is likely to find relatable in their current condition. Since this is a complex and sensitive task, the assistance of a trained psychotherapist is generally advised for formal bibliotherapy.

It is unlikely that the role of a human therapist could be replaced via algorithmic recommendations in formal bibliotherapy for formal mental health treatments. However, bibliotherapy is not always formal. The friendly neighborhood librarian's suggestions for a book to read after hearing a teenager's anxiety about a recent distressing event is an equally valid, albeit, informal type of bibliotherapy [3].

Given the plenitude of digital reading resources available today, it is surprising to find that the prospect of conducting informal bibliotherapy using algorithmic recommendations has not been well-studied. We note an online bibliography system described in the literature using collaborative filtering based approach based on system-defined interest category tags [15]. However, this system was evaluated very weakly in a user study using 10 motivated volunteers, who filled out questionnaires before and after using the system for 6 weeks that asked them if they found the system useful. The non-blind nature of the evaluation, and the use of motivated volunteers makes it difficult to clearly evaluate the value of the system, and the system itself is not publicly accessible to permit an external evaluation.

In this paper, we propose a recommender system for informal bibliotherapy that identifies relevant recommendations for users based on long-form written elicitations from them. We also report preliminary results from a user study that assesses whether such systems are useful for participants, and whether delivering personalized recommendations for bibliotherapy is a good idea. Given the ongoing debate about the nature of 'filter bubbles' generated by recommender systems [8], and the delicate nature of the recommendation needed for effective bibliotherapy [9], it is not at all clear that personalized recommendations using standard RS methods would be beneficial, thus stimulating our research towards answering these questions.

2 AN ONLINE BIBLIOTHERAPY RECOMMENDER SYSTEM

Our system is essentially a personalized search engine for a set of online narratives (primarily blogs) related to personal resilience and growth. The system's key novelty is the use of text responses elicited from users to suggest readings to them. We use classic

HealthRecSys '20, September 26, 2020, Online, Worldwide.

information retrieval methods to index this corpus of narratives, natural language processing methods to transform long-form user text inputs into search queries, and query expansion and ranking techniques to extract relevant articles given users' searches. We describe the system in greater detail below. To correlate all details described here, our system itself is accessible at this URL for reference.

2.1 Application walk-through

Users interact with our system using anonymous IDs, preserving privacy at source. We collect no identifying information by design. However, for users participating in user studies for compensation (discussed below), there is an option to enter an email address for future identification.

New registrants to the system are encouraged to watch a video, prominently displayed on the landing page, describing the possible modes of engagement with the system and asked to select a unique ID from a set of randomly generated ID strings. Returning users sign into the system using the IDs they selected at the time of registering.

We elicit user mood responses on a 5 point scale, with adjectival descriptions of a rating of 1 being *terrible* and a rating of 5 being *amazing*. A corresponding smiley emoticon reacts to the user's input. Following this numeric elicitation, the system asks the user whether they want to type in more about their mood (see Figure 1A). If they click 'yes', they see a 'type-it-out' text box in which they enter long-form text describing how they're feeling and why they're feeling that way (see Figure 1B). If they click 'no', they directly see reading recommendations (see Figure 1D).

Users see recommended readings sequentially in a carousel view. They may choose to rate these narratives by personal relevance on a scale of 1 to 5, where 1 represents highly irrelevant and 5 represents highly relevant. We explain in the introduction to the system that relevance, in the context of our system, means how close a particular narrative is to what users are experiencing in their life and whether they can learn something from it, and realise what must be done to get things back on track in their own lives.

The web service itself follows a Model-View-Controller (MVC) architecture. We use Jinja2 to create the application's views. All the data-related logic is handled by an SQLite database, which forms the Model component of the architecture. Python (with Flask) acts as the Controller and acts as an interface between Model and View.

2.2 Indexing narratives

Narratives are selected based on whether they are prominently related to first person accounts of mental resilience and personal growth through periods of sadness and distress. While the system is designed to handle a large number of narratives, we conducted the user study we describe below using a restricted set of 30 total narratives, which are treated as individual documents by our information retrieval system.

We apply a standard natural language pipeline for converting each document into a bag-of-words representation. Each one is tokenized, stemmed, and lemmatized and stop words are removed. The bag of words associated with each document is stored in an inverted index.

2.3 Query processing

Our system treats the text entered in the 'type-it-out' box as an implicit query, and uses it to identify relevant blogs for the user. We first extract keywords from the text using the RAKE algorithm [12]. We then expand the queries by adding the five most similar words to each keyword into the set of keywords. Word similarities are calculated using a word2vec model pre-trained on a large corpus of internet documents [11]. Finally, we use tf-idf ranking to retrieve the most relevant documents corresponding to the final query.

3 USER STUDY

Our basic expectation in designing this system is that it will help people cope with stress and unhappiness by connecting them with first-person narratives of surviving difficult situations. Such a proposition can be operationalized in many different ways, making it hard to test. However, we conducted a user study with a fixed operationalization to see if we could characterize the value of the system in some quantitative form. In our study, we focused on answering two questions:

- Does greater engagement with the system predict improvements in mood?
- Does personalization of recommendations based on users' descriptions of how they're feeling actually produce greater user engagement?

Mood ratings are already collected in the system as described above on each user visit. We operationalize engagement using relevance ratings assigned by users to the readings our system offers them, a common practice in recommender systems research [16].

3.1 Participants

We recruited participants soliciting interest from people who wanted to develop a skill set for dealing with stressful situations in life or those who wanted to offer their contribution to improving mental health care. Our primary source of participants was word of mouth mixed with chain referral sampling. In total, 190 people signed up for the study, but only 36 people completed all planned sessions, yielding an attrition rate of 81%. While high, this is not unexpected, since self-help studies, focusing as they do on people experiencing or prone to experience mental distress frequently exhibit similar or even higher attrition rates [5]. While we had originally conceived of participation as paid, this was eventually not the case when we ran the study. Thus, participants received no compensation for participation, a detail that was clearly indicated to them at the time they registered for participation, as well as in our IRB proposal.

3.2 Protocol

In the interest of maintaining ecological validity, we asked participants to simply use the system at least once in a span of 3-7 days over a period of eight weeks, and exactly seven times in total. A reminder email was sent to each participant four days from the last time they had participated. The IRB Board at IIT Kanpur reviewed and approved this protocol.

We made a slight modification in the design of the system used for the user study by inserting a short questionnaire in the system at the point where participants are asked whether they wanted to type

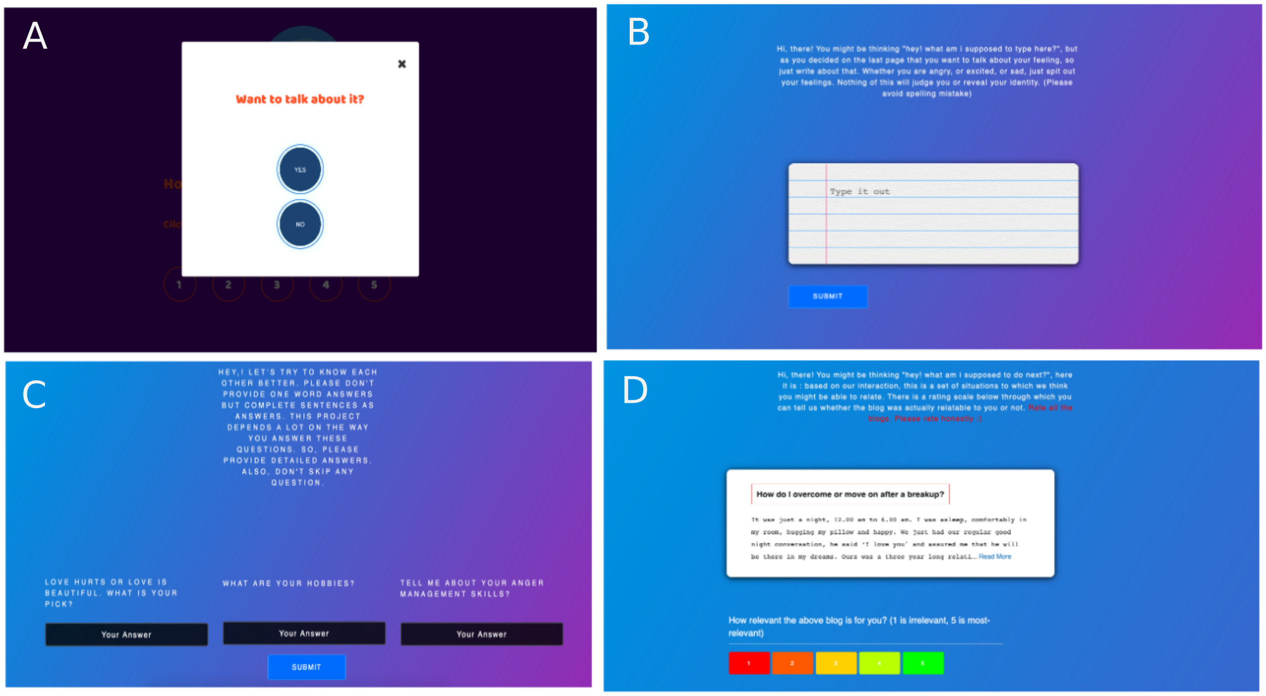


Figure 1: Screenshots showing the basic flow of the online bibliotherapy system. Long form text elicited conversationally by the system is used to recommend specific readings to users. (A) asking the user if they want to describe why they're feeling the way they're feeling currently, (B) a text box where the user can enter their thoughts in long-form text (C) a questionnaire introduced into the system as part of the user study, and (D) the main interface where users see blogs recommended by our RS based on their earlier inputs and rate them for relevance.

in something about their mood (see Figure 1C). The questionnaire consisted of 3 short answer questions selected from a bank of 21 questions, rotated such that participants didn't have to respond to the same question twice across 7 sessions.

This addition was necessary because we did not want to force participants to write about their mood, but at the same time wanted text samples to guide the recommender system. By adding 3 short questions, we obtained sufficient text content to guide recommendations even for participants who did not want to respond in long-form to the mood description probe. Thus, for participants who responded in long-form, we used their long-form responses to generate recommendations. For participants who chose not to respond in long-form, we generated recommendations using answers to the questionnaire questions. The questionnaire is available on the version of the system linked to from this paper, but is not expected to be a part of the actual system in deployment.

Also, we fixed the number of readings each person would read and rate in a particular session at 10, and changed our recommendation algorithm to ensure that 5 of these 10 readings would be generated based on the users' text inputs, treated as queries by our system, and the rest would be randomly selected from the remaining narratives.

3.3 Results

3.3.1 Engagement with the system is weakly correlated with improvement in mood. Our first research question in this project was to identify whether engagement with our system resulted in improvement in participants' mood ratings. To characterize this relationship, we estimated the mood trend for each participant across the seven sessions for which they gave us mood ratings. For each user, we fit an ARIMA(1,0,0) model to tease apart the underlying trend line from transient fluctuations. We correlated this estimated trend in mood with our measure of user engagement - the average relevance scores assigned to the readings each participant rated across all seven sessions. Greater engagement was expected to yield higher average relevance scores, viz. the participant felt that they could identify with the themes or protagonists strongly in several narratives. As is visually evident in Figure 2, we found a moderate positive correlation between these two quantities ($\rho = 0.25$), although this relationship did not reach statistical significance because of the small number of participants who completed the study ($p = 0.14$).

3.3.2 Engagement with narratives is selective and heterogeneous. While the average relevance score across narratives gives a reasonable summary of a user's engagement with the system, it also conceals important information about the specificity users' interests in self-help readings. To uncover this information, we show

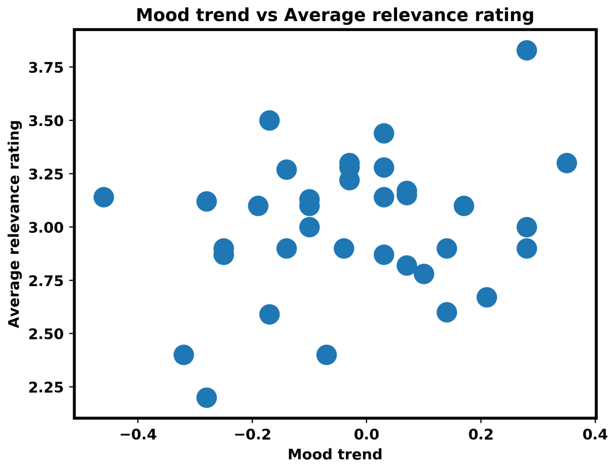
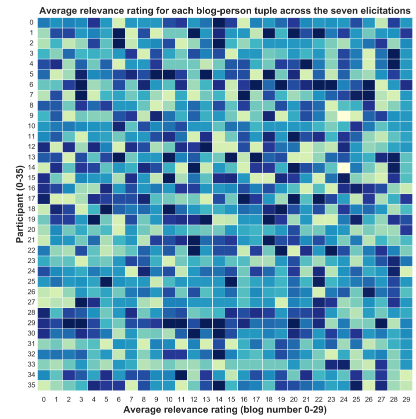
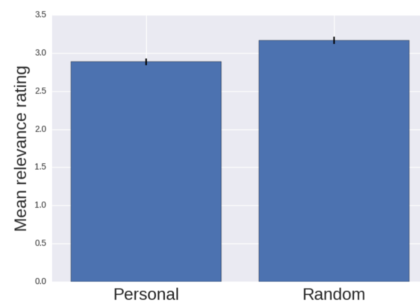


Figure 2: A scatter plot of trend in mood ratings over sessions versus average relevance scores across sessions for all participants



(a) Average relevance rating for each blog-person tuple across seven elicitations



(b) Average relevance rating across participants while rating personalized and random suggestions from the recommender system

Figure 3: Lack of evidence for efficacy of collaborative filtering or content-based recommendations in online bibliotherapy

the average relevance ratings for all user-narrative pairs in the user study in Figure 3a.

The interesting observation here is the lack of any global structure in the heatmap, which would have indicated clusters of user or narrative similarities, which in turn would have suggested value in collaborative filtering-based approaches for making recommendations in this domain. The absence of such structure recommends caution in applying such approaches: it looks like users are highly selective about which narratives they find personally relevant, and there is very little similarity in these preferences across users. Tolstoy said that every unhappy family is unhappy in its own way. It looks like individuals find resonate with self-help narratives also each in their own way.

3.3.3 Personalized recommendations create significantly lower engagement. The last important insight our data reveals answers our second research question: do participants find value in receiving recommendations matching their own personal narratives of their present sense of being. Recall that each user rates 5 personalized and 5 randomly selected narratives in each session, effectively generating a two condition within subject manipulation of personalization.

Figure 3b plots the average relevance rating for all participants in each of these two conditions - while rating narratives selected by the RS, and while rating randomly selected narratives from the same corpus. Interestingly, the randomly generated narratives are rated significantly higher than personalized narratives (two sample t-test $t = 4.4$, $p < 0.001$). Thus, it looks like personalized recommendations are actually reducing user engagement in our user study.

4 DISCUSSION

In this paper, we have presented a recommender system that attempts to mimic the role of a counsellor suggesting informal bibliotherapy. The key novelty of the system is the use of natural language processing to transform long-form narratives generated by the user as a description of their current mental condition into relevance cues used to retrieve semantically related readings.

We also report preliminary results from a user study, wherein we found that engagement with the system is weakly correlated with elevated mood among participants who completed the study. This correlation, however, should not be mistaken for causation. It is certainly possible that experimenter-demand effects induced by responding on the same system over and over again might account for much of the measured improvement [2]. Further research, potentially including retrospective self-reports and clinical interview-based debriefs of study participants, are needed to provide more substantive evidence for a causal relationship.

The user study also uncovers intriguing evidence supporting the case that conventional recommendation strategies may not work very well in self-help settings. The pattern of ratings displayed by our participants suggests that this domain does not lend itself very easily to collaborative filtering based recommendations. Our controlled within subject manipulation of personalization also reveals that content-based recommendations appear to lead to lower user engagement than random recommendations from the same thematic corpus of readings. These observations substantiate recent theoretical proposals about the inappropriateness of conventional recommendation strategies in self-help settings [8, 14].

The possibility that suggesting narratives resembling the user's personal narrative of distress may backfire is not counter-intuitive. It is quite conceivable that suggesting too related a narrative might cause the user to not acquire sufficient psychological distance from their own predicament for therapeutic mental transformations to occur. This observation joins the ever-growing list of unexpected difficulties that the psychological feedback loops perpetuated by recommender systems generate [1, 4].

The research we have reported in this paper is very much a work in progress, and we are in the process of extending it in two dimensions. One, collecting data from more participants will enable us to draw conclusions with greater statistical confidence, as well as perform more interesting sub-cohort analyses using individual differences between users. Two, as stated above, interview-based debriefing of study participants may give us more direct evidence to support the inferences we are currently drawing indirectly from data analysis.

ACKNOWLEDGMENTS

We are grateful to our study participants for their support in testing the system.

REFERENCES

[1] Homanga Bharadhwaj and Nisheeth Srivastava. 2019. New Tab Page Recommendations Strongly Concentrate Web Browsing to Familiar Sources. In *Proceedings of the 10th ACM Conference on Web Science*. 7–16.
[2] Jonathan de Quidt, Lise Vesterlund, and Alistair J Wilson. 2019. Experimenter demand effects. In *Handbook of research methods and applications in experimental economics*. Edward Elgar Publishing.
[3] Deborah Fanner and Christine Urquhart. 2008. Bibliotherapy for mental health service users Part 1: a systematic review. *Health Information & Libraries Journal* 25, 4 (2008), 237–252.
[4] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.
[5] Russell E Glasgow and Gerald M Rosen. 1978. Behavioral bibliotherapy: a review of self-help behavior therapy manuals. *Psychological bulletin* 85, 1 (1978), 1.
[6] R. J. Gregory, S. Schwer Canning, T. W. Lee, and J. C. Wise. 2004. Cognitive Bibliotherapy for Depression: A Meta-Analysis. *Professional Psychology: Research and Practice* 35 (2004), 275–280. Issue 3. https://doi.org/10.1037/0735-7028.35.3.275
[7] M.R. Gualano, F. Bert, M. Martorana, G. Voglino, V. Andriolo, R. Thomas, C. Gramaglia, P. Zeppegno, and R. Siliquini. 2017. The long-term effects of bibliotherapy in depression treatment: Systematic review of randomized clinical trials. *Clinical psychology review* 58 (2017), 49–58.
[8] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender systems for self-actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 11–14.
[9] Fran Lehr. 1981. ERIC/RCS: Bibliotherapy. *Journal of Reading* 25, 1 (1981), 76–79.
[10] Grainne McKenna, David Hevey, and Elaine Martin. 2010. Patients' and providers' perspectives on bibliotherapy in primary care. *Clinical psychology & psychotherapy* 17, 6 (2010), 497–509.
[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[12] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory* 1 (2010), 1–20.
[13] Nancy M Smith, Mark R Floyd, Forrest Scogin, and Christine S Jamison. 1997. Three-year follow-up of bibliotherapy for depression. *Journal of consulting and clinical psychology* 65, 2 (1997), 324.
[14] Daricia Wilkinson. 2018. Testing a recommender system for self-actualization. In *Proceedings of the 12th acm conference on recommender systems*. 543–547.
[15] Yunxing Xin, Yongqiang Chen, Li Jin, Yici Cai, and Ling Feng. 2017. TeenRead: An Adolescents Reading Recommendation System Towards Online Bibliotherapy. In *2017 IEEE International Congress on Big Data (BigData Congress)*. IEEE, 431–434.
[16] Qian Zhao, F Maxwell Harper, Gediminas Adomavicius, and Joseph A Konstan. 2018. Explicit or implicit feedback? Engagement or satisfaction? A field experiment on machine-learning-based recommender systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. 1331–1340.