CHI: G: Towards Developing a Global Self-Updating Meeting List for Alcoholics Anonymous

Sabirat Rubya, University of Minnesota (rubya001@umn.edu)

ABSTRACT

Twelve-step fellowships like Alcoholics Anonymous is an effective maintenance program for millions of people trying to recover from substance use disorders. Attending twelve-step meetings is considered to play a vital role for exchanging support and for encouraging continued abstinence from addiction. However, when technology interacts with these recovery programs, issues like anonymity, identity, etc., arise and it is vital to understand the challenges faced by the member of these groups to inform design of technology for them. Through formative investigations we identify a critical problem faced by members of Alcoholics Anonymous and propose a solution to the problem using machine learning, data mining, and human computation.

INTRODUCTION

Substance use disorders are prevalent and high-impact health conditions among US population. It is estimated to cost the United States $374 billion per year, with 11.9% of Americans developing a dependence on the substance [2]. Treatments for recovery from SUDs are effective when clinical intervention is coupled with long-term maintenance programs. One of the most effective and popular maintenance program is twelve-step fellowships (e.g., AA and NA). AA and NA focus on providing face-to-face support through sharing problems and experiences in meetings.

Previous studies show that members of these programs worry about anonymity, privacy, identity, etc., as technology interacts with the program [13,16]. However, there are opportunities for designing technology to help the peer support process in these special groups. This work attempts to expand our understanding of the needs of these groups in order to inform the design of technologies for them.

In this paper we discuss our formative research with people in recovery, identify a critical problem they face, begin addressing the problem with a technical solution that leverages techniques from data mining, machine learning, and human computation. We describe our research and design in the following three phases.

PHASE-1: FORMATIVE INVESTIGATIONS

Related Work

Although there have been numerous studies on developing and evaluating technology-based intervention techniques for substance addiction [7,15], only one study has conducted formal investigations of how technology fits (or does not fit) into the culture of twelve-step recovery programs [16]. Yarosh described in that work how technology may introduce issues of achieving anonymity, consensus, autonomy, etc. as it interacts with recovery. Therefore, it is important to apply formative research methods to understand needs and challenges faced by these groups.

Methods

In order to understand the technology needs of recovery communities, we conducted an online questionnaire study and an in-depth interview study.

Questionnaire

A link to a questionnaire was distributed as a paid banner advertisement on the homepage of an online recovery community (InTheRooms[1]) for one month, as well as advertised in the weekly newsletter of the community. 285 participants completed the questionnaire. The average age of these participants was 49 years (SD=11.5) with 61% of them being female. Participants' geographic location covered 16 different countries, with majority of them (72%) being from US. The demographic information of the participants was representative of active users of the community. Survey participants had an average of 8.5 years of recovery time (SD=9.5 years). Participation in different twelve-step meetings was self-reported as an important part of recovery.

Semi-structured Interviews

We recruited 14 AA/NA members from InTheRooms. The participants had an average of 8.6 years of recovery time (SD=9.75 years). They represented four different countries and nine of them were female. These participants had frequent face-to-face and online meeting (hosted on ITR) attendances in the past three months. We conducted a 90-minute semi-structured interview with each of them. The questions asked focus on use of technology in finding or attending meetings, reasons of attending (or not attending) face-to-face meetings, and challenges or problems faced in finding or attending meetings. The transcripts were analyzed using a data-driven approach. Statements of interest (open codes) were extracted and grouped by different themes.

Results

Previous research on recovery from substance abuse has emphasized the importance of attending twelve-step meetings for abstinence [16]. The questionnaire and the interview studies resonated with this point, with many participants mentioning the meetings as sources of encouragement and support in their recovery processes. However, the participants described many challenges to find meetings.

Importance of Meeting Attendance

67% of the survey responders reported to attend 2-3 meetings per week. The self-reported average involvement score

[1] http://www.intherooms.com

for commitment to recovery was 32.5 (SD=4.3) out of 40 for those who attend at least one meeting per week vs 30.2 (SD=5.6) for those who do not attend meetings (higher score means greater involvement). The measurement was based on a modified 7-item AAI (Alcoholics Anonymous Involvement) scale [9]. The average years of recovery was also higher for participants who reported to attend only face-to-face (M = 12.2, SD = 12.3) or both face-to-face and online (M = 8.3, SD = 9.5) meetings than who attended only online meetings.

Although online meetings have been receiving attention recently, most of the participants thought that the online meetings were useful only if they were attended as a supplement of the face-to-face meetings, not as a replacement.

Many of the interview participants, especially newcomers, considered the meetings as a way of avoiding self-destructive and delusional thinking, loneliness, and isolation. Many newcomers see these as "*a place to hear about experience, strength, and hope from others*", whereas old-timers (members who have fairly longer recovery time) view it as a way to "*pass on the love that's once being given*" and to "*let people know that somebody cares*". Meeting different people in the meetings enabled them to "*share problems with like-minded others*," as well as to "*know about diverse perspectives*."

Face-to-face meetings typically have a consistent set of attendees which makes it easy to identify a newcomer. Newcomers are usually nervous; hence a welcoming gesture enables them to overcome their struggle. Besides, they can form close friendship with people from their local meetings and even find a sponsor.

Challenges in Finding Meetings

Despite the popularity of the 12-step programs and potential of meeting attendance for positive impact, there is no global list of AA meetings. The fourteen interview participants mentioned seven different sources for finding right meetings for themselves (Google: 6, booklet/pamphlet: 2, local AA websites: 3, calling hotline/ intergroup: 1, sponsor: 1, InTheRooms list: 3, email: 1). Regular meeting attendees in a particular region are usually aware of their groups' meeting schedule and any change in that, but it becomes difficult to find the right meeting when they are traveling.

> "*I went to visit family and I looked up for some types of meeting. One of the meetings that I was going to go to, I drove down the road wasn't there. I really really needed a meeting, I didn't have my computer. It was awful.*" [P6]

Besides, static apps or websites that are built based on information from these multiple websites become outdated once a meeting information in any of the domains changes. Websites of different AA regional intergroups contain meeting lists in different formats (e.g., webpage, PDF, scanned images, etc.) and accessing these lists from different devices become even more difficult. What is worse is a newcomer may lose faith in the program if they feel the immediacy of attending a meeting but cannot find one.

Discussion

Norms around designing technology for peer support usually include online communities and social networks. However, participants in our study utilized online support as a supplement to the offline support and this support is very important to build their recovery network. When asked about problems they faced, they mostly mentioned the challenge in finding face-to-face peer support. We believe that it's important not to limit the design of technology to only facilitate online communication. Therefore, we identify the lack of AA global meeting list as a critical problem for people in recovery and describe how technology can help.

PHASE-2: DESIGNING A GLOBAL SELF-UPDATING AA MEETING LIST USING MACHINE LEARNING AND DATA MINING

We aim to address the challenge of finding meetings by extracting meeting information from all different regional AA websites and by creating a global meeting finder that can keep track of updated meeting lists. Regional websites have many webpages other than the meeting list (e.g., office hours, AA resources, events, etc), and different websites maintain list of local meetings in different formats (e.g., HTML table, PDF, map, calendar, etc.). Therefore, differentiating a page that contains a meeting list from one that does not and extracting meaningful information related to different meetings (e.g., day, time, address) are not very straight-forward problems. We combine existing techniques from Data Mining and Information Extraction domains to automatically extract and collate meeting information.

Related Work

The process of finding meetings includes aspect of Machine Learning (for identifying meeting pages) and Data Mining (for extracting meeting information).

There have been extensive investigations in classifying webpages of different types using machine learning [14]. In service of better online search, researchers constructed features based on the text of the URLs [6], structure and content of webpages (e.g., [12]), and aspects of neighboring pages [15]. We primarily use methods from the former category to generate the features for our machine learning model. Past research showed that Random Forest is the most effective technique in these kinds of problems [1], so we applied Random Forest with Bagging to distinguish meeting pages from non-meeting pages.

Previously researchers have used natural language processing, pattern matching, rule based mining, and machine learning [3,10] for extracting data from heterogeneous semi-structured or unstructured data sources. The use of rules and pattern-matching exploits basic patterns over a variety of structures, such as text strings, part-of-speech tags, semantic pairs, and dictionary entries. Regular expressions are effective when the structure of the text and the tokens are consistent, as in our case (e.g., each meeting must have an address, a day, and a time associated with it). We have followed a modified version of the algorithm developed in [3], where the treelike structure of webpages and repetitive patterns were utilized to extract records.

Methods

(a) Extracting All Webpages from Local AA Domains

The official website of AA (www.aa.org) contains the URLs for all regional AA domains. These domains usually consist of pages describing the AA program, meetings, events, resources for newcomers, etc. We wrote a web-scraping script using Python Selenium library that extracted all the domain URLs and recursively traversed through the links referenced by each page in a depth-first search manner. We obtained over 10,000 URLs from the 385 domains. Although there are 592 total AA regions throughout the United States 207 of them listed their hotline numbers only. Besides, Multiple pages with same content were considered as duplicate pages, as it is possible for different URL strings to point to the same page, e.g. http://www.aalakesumter.com/index.html and http://www.aalakesumter.com/, which could affect the classification accuracy if counted twice. Only one instance of these pages was included After filtering and duplicate removal, we ended up with 9468 pages in total.

(b) Detection of Meeting Pages using Machine Learning

Out of the 9468 URLs that we scraped, only around 30% include meeting information. To significantly automate the process of determine which of these extracted webpages contain one or more open AA meeting, we developed a binary classifier using random forest with bagging.

Ground truth: To train our classifier we selected a random subset of our data (containing 642 pages), making sure to pick at least one page from each domain. Each page was classified manually and independently by at least two researchers, with a 0 (non-meeting page) or a 1 (meeting page). Cases of disagreement were discussed among researchers until consensus was reached. 452 and 190 pages were classified as nonmeeting pages and meeting pages respectively.

Feature selection: From our experience of manually classifying hundreds of pages, we developed a list of possible page features to use in page classification. For example, basic information about a meeting usually contains a day, a time, and an address. Though the meetings are listed in different formats in different domains, meeting pages often contain a larger number of times and addresses than the non-meeting pages. Therefore, we considered number of times and number of addresses as distinguishing features between meeting and nonmeeting pages. We took into consideration whether the names of days of the week are present on the page. We also considered the presence of the word "meeting" in the link (e.g. http://aaminneapolis.org/meetings/ is a meeting page whereas http://aaminneapolis.org/a-arelated-links/ is not), and number of occurrences of the word "meeting" on the page. The set of final features included: if the word "meeting" is present in the URL, proportion of occurrences of text structured as a time on page, proportion of occurrences of text structured as an address on page, number of occurrences of weekday names in the text, number of occurrences of word "meeting". We wrote a Python script to extract these features.

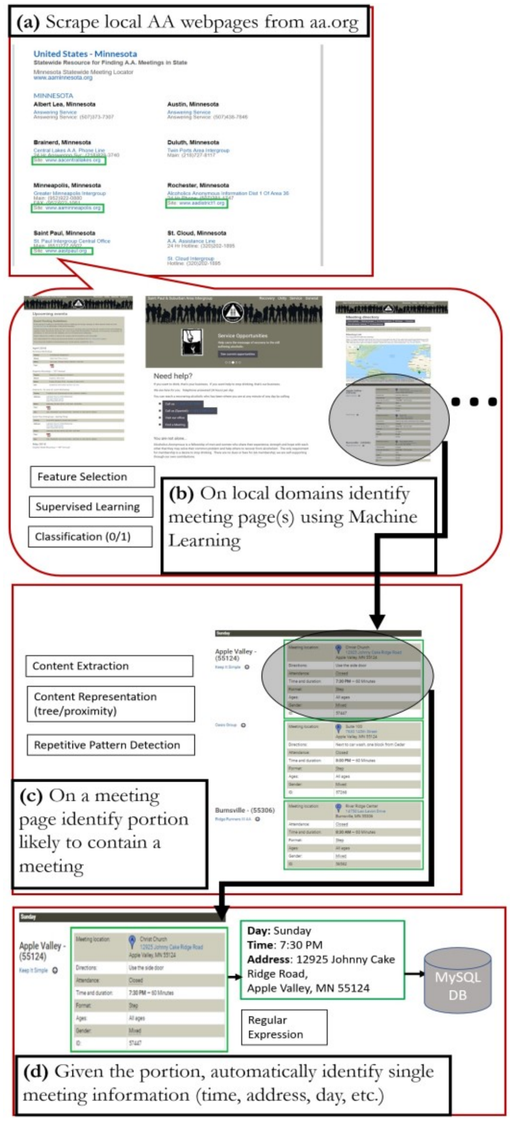

Figure 1: Steps of Extracting Meetings

Classification: We determined that it is more important that meeting pages are identified correctly than to eliminate nonmeetings classified as meetings (since we want our global meeting list to include all meetings). We tried out a number of classification algorithms [1] to determine which one performs better with respect to recall and overall F1 score, and selected random bagging with forest. The model also outputs a confidence of classification with each page.

% of meetings identified accurately	72.2%
% of meetings identified with partially correct information (e.g., wrong time, day, or address)	10.9%
% of meetings unidentified	16.9%

Table 1: Summary of Extracted Meeting Information from Saint Paul Domain

(c) Identifying Meeting Location on Pages

For the pages classified as meeting pages in previous step, we applied repetitive pattern detection techniques [3] to identify which portion of those pages contain meeting information. This step was necessary for separating individual meeting information from a list of meetings.

The HTML content of the webpages was represented as a tree of HTML tags and similar patterns were extracted. As described earlier, the meeting pages are maintained in different formats which make the extraction difficult. Other than the HTML webpages (which are about 60% of the total number of meeting pages) we considered PDFs, documents, spreadsheets, google calendar, and images. For the meeting pages which were saved as Google Calendar we used Google Calendar API. Any type of document file was converted to PDF and then we used python PDFMiner library to separate text segments and coordinates of those segments. Similarly, we used Python Tesseract library to extract text segments and their coordinates from the images.

(d) Extraction of Meeting Information

For the webpages, we checked for presence of regular expressions of time, day, or address in the repetitive patterns. If one of these three aspects was missing from a pattern (e.g., some webpages list meetings by day, so the repetitive patterns will contain time and address and the day information will appear before a list of meetings) we looked for that particular information in the siblings and parent of that pattern in the pattern tree. For Google Calendars, events were extracted using Google Calendar API. In case of PDFs and images, the text segments were checked similarly for presence of time or address. Missing information was assigned for each meeting based on proximity of the text. We saved all of these meetings in a MySQL database.

As a proof of concept, we applied this step to the Saint Paul domain (http://www.aastpaul.org/). First, we manually recorded meeting information from the meeting page of this domain. We compared these meetings with the meetings found after applying the pattern detection technique. These steps are illustrated with an example in Figure-1.

Results

Page Classification Accuracy

Out of the 8826 test instances, 4268 pages were classified as meeting pages. The overall F1 score was 73.8% (recall: 90.5%)

Extraction of Meeting Information

After applying the random bagging on pages extracted from www.aastpaul.org, we got 12 pages classified as meeting pages and 48 as nonmeeting pages. We manually checked those 12 pages to find out which were actual meeting pages. We identified that many of these pages list AA events which also have similar information (day, time, etc.) and thus have been classified as meeting pages. There were two actual meeting pages where we ran the pattern detection algorithm and the results are summarized in Table 1. From ground truth data we had identified a total of 728 meetings from this domain.

Discussion

The presence of other pages that look like and have similar information like a meeting page makes the automatic extraction difficult. Moreover, the regular expressions often capture wrong addresses of the meetings. Therefore, a lot of the extracted information are often wrong or redundant. It leads us to the idea of applying human computation techniques to get rid of this problem, since tasks like identifying if an address is correctly sought from a webpage, or differentiating a meeting and an event are easier for a human, whereas a machine may fail.

PHASE-3: INTEGRATING HUMAN COMPUTATION

We plan to apply the crowdsourcing in two steps of our pipeline. First, after (step b) the machine learning classifier outputs the probable meeting pages, we send those pages to the crowd workers for validation. Then we would run the pattern detection algorithm only on the set of pages that have been validated by crowd as a final meeting page. We run the pattern detection for only these pages and for the extracted meetings from these pages (step d), we validate the meeting information and also check if any meeting was missed. We recently completed the first stage of crowdsourcing and are currently exploring the second stage.

Related Work

Researches have been utilizing crowdsourcing primarily for obtaining ground truth labels for datasets and large-scale user studies of systems. It is not until recently that crowdsourcing has been combined with machine computations to improve accuracy of tasks that is difficult for either one to do by itself [4]. This combination has been proved efficient in solving problems like video-captioning [8], cheating detection [11], etc.

Initial Methods

Initially we have applied crowdsourcing after step (b). This has improved our results for the page classification.

Determining Which Pages to Send to Crowd Workers

To determine which instances to send for crowdsourcing, we used a prediction margin that the classifier outputs for each classified instance, which rates the confidence of the classifier associated with each prediction. We determined a cutoff value where the set of instances below the margin are sent to crowd workers and the remaining were accepted as correct classification. We sorted the pages by prediction

	Total from ML	With <90% confidence
Meeting page	4268	3569
Nonmeeting page	4558	312

Table 2: Pages sent to Crowd

margin and selected the cutoff (90%) that minimizes the size of the set below the threshold (less likely to be correct) and minimizes the number of errors above the threshold.

The Crowdsourcing Interface

We developed an interface (http://www.cs-aameeting.cs.umn.edu) using Python Flask for the crowd workers to label pages as meeting or nonmeeting pages (with an additional option of selecting "not sure," in which case they have to answer two or three additional questions to help us determine the class). On the backend we used the same MySQL database holding the information about meeting pages and individual meetings.

Recruitment of Workers and Labeling

We recruited workers from Amazon Mechanical Turk in a two months period. The HIT was designed as a survey with instructions for the workers on differentiating meeting pages from nonmeeting ones, provided examples, and redirected them to our interface for labeling. Each HIT would include labeling of twenty different webpages for 50 cents. Each page was classified by five different workers and the final class was determined from majority voting.

Filtering Crowdsourced Results

To filter out answers from dishonest workers we applied methods from [5] to assign penalty to workers for each incorrect answer which would affect their accuracy. We approved answers from only those workers whose accuracy was above 50%.

Initial Results

We decided based on the prediction margin which pages to send to the crowd for verification. This was necessary for the reduction of crowdsourcing cost. The details of pages sent for validation is listed in Table 2. After this step, we obtained a total of 1275 meeting pages (Recall: 93%, F1: 77.8%).

Discussion

In our initial deployment we received feedback from workers that some pages were confusing for them to classify. Even workers had confusion about the difference between a meeting page or an event page (or a service meeting page that is not open to all). We incorporated more examples and instructions with our HIT description and it helped the workers. The ultimate goal is to utilize the people in recovery as they can perform these tasks as a service to help other alcoholics. Besides, they have more experience and knowledge about the structure of the meeting pages and can provide us with more accurate result in shorter period of time.

CONCLUSION

People in recovery programs like AA exchange peer support and build their recovery network through attending face-to-face meetings. However, through our formative investigations we found that there is a lack of updated list of meetings they want to attend which make it hard for them to get to the right meetings. We address this problem by designing a global meeting finder combing machine learning, and human computation techniques.

ACKNOWLEDGMENTS

I thank my adviser, Dr. Svetlana Yarosh, for her guidance. This work was funded by the NSF grant (1464376).

REFERENCES

1. Ansam A. AbdulHussien. 2017. Comparison of Machine Learning Algorithms to Classify Web Pages. *IJACSA* 8, 11.
2. National Institute on Drug Abuse. 2015. Trends & Statistics. Retrieved January 13, 2017 from https://www.drugabuse.gov/related-topics/trends-statistics
3. Chia-Hui Chang and Shao-Chen Lui. 2001. IEPAD: Information Extraction Based on Pattern Discovery. In *Proc. of* WWW '01.
4. Ed H. Chi. 2017. Technical Perspective: Humans and Computers Working Together on Hard Tasks. *Commun. ACM* 60, 9: 92–92.
5. Nilesh Dalvi *et al*. 2013. Aggregating Crowdsourced Binary Ratings. In *Proceedings of* WWW '13, 285–294.
6. M. I. Devi, R. Rajaram, and K. Selvakuberan. 2007. Machine Learning Techniques for Automated Web Page Classification Using URL Features. ICCIMA *2007,* 116–120.
7. David H. Gustafson *et al*. 2002. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. *Intl Journal of Medical Informatics*.
8. Yun Huang *et al*. 2017. Leveraging Complementary Contributions of Different Workers for Efficient Crowdsourcing of Video Captions. In *Proc of* CHI '17.
9. Keith Humphreys *et al*. 1998. The AA Affiliation Scale: Development, Reliability, and Norms for Diverse Treated and Untreated Populations. *Alc: Clinical and Exp Research*
10. Wendy Lehnert *et al*. 1995. Inductive text classification for medical applications. *Journal of Experimental & Theoretical AI* 7, 1: 49–80.
11. Xuanchong Li *et al*. 2015. Massive Open Online Proctor: Protecting the Credibility of MOOCs Certificates. In *Proc. of* CSCW '15, 1129–1137.
12. Xiaoguang Qi and Brian D. Davison. 2009. Web Page Classification: Features and Algorithms. *ACM Comput. Surv.* 41, 2: 12:1–12:31.
13. Sabirat Rubya and Svetlana Yarosh. 2017. Interpretations of Online Anonymity in Alcoholics Anonymous and Narcotics Aonymous. *PACM on HCI* 1, 2.
14. Makoto Tsukada, Takashi Washio, and Hiroshi Motoda. 2001. Automatic Web-Page Classification by Using Machine Learning Methods. In *Proc of* WI '01, 303–313
15. Kuo-Cheng Wang *et al*. 2014. SoberDiary: A Phone-based Support System for Assisting Recovery from Alcohol Dependence. In *Proc of* UbiComp '14 Adjunct.
16. Svetlana Yarosh. 2013. Shifting Dynamics or Breaking Sacred Traditions?: The Role of Technology in Twelve-step Fellowships. In *Proceedings of* CHI '13, 3413–3422.