Original Paper

Detecting Recovery Problems Just in Time: Application of Automated Linguistic Analysis and Supervised Machine Learning to an Online Substance Abuse Forum

Rachel Kornfield[1*], MA; Prathusha K Sarma[2*], MS; Dhavan V Shah[1], PhD; Fiona McTavish[3], MA; Gina Landucci[3], BS; Klaren Pe-Romashko[3], MS; David H Gustafson[3], PhD

[1]School of Journalism and Mass Communication, University of Wisconsin-Madison, Madison, WI, United States

[2]Department of Electrical & Computer Engineering, University of Wisconsin-Madison, Madison, WI, United States

[3]Center for Health Enhancement System Studies, University of Wisconsin-Madison, Madison, WI, United States

[*]these authors contributed equally

Corresponding Author:
Rachel Kornfield, MA
School of Journalism and Mass Communication
University of Wisconsin-Madison
5115 Vilas Hall
821 University Avenue
Madison, WI, 53706
United States
Phone: 1 4153355356
Fax: 1 608 890 1438
Email: rkornfield@gmail.com

Abstract

Background: Online discussion forums allow those in addiction recovery to seek help through text-based messages, including when facing triggers to drink or use drugs. Trained staff (or "moderators") may participate within these forums to offer guidance and support when participants are struggling but must expend considerable effort to continually review new content. Demands on moderators limit the scalability of evidence-based digital health interventions.

Objective: Automated identification of recovery problems could allow moderators to engage in more timely and efficient ways with participants who are struggling. This paper aimed to investigate whether computational linguistics and supervised machine learning can be applied to successfully flag, in real time, those discussion forum messages that moderators find most concerning.

Methods: Training data came from a trial of a mobile phone-based health intervention for individuals in recovery from alcohol use disorder, with human coders labeling discussion forum messages according to whether or not authors mentioned problems in their recovery process. Linguistic features of these messages were extracted via several computational techniques: (1) a Bag-of-Words approach, (2) the dictionary-based Linguistic Inquiry and Word Count program, and (3) a hybrid approach combining the most important features from both Bag-of-Words and Linguistic Inquiry and Word Count. These features were applied within binary classifiers leveraging several methods of supervised machine learning: support vector machines, decision trees, and boosted decision trees. Classifiers were evaluated in data from a later deployment of the recovery support intervention.

Results: To distinguish recovery problem disclosures, the Bag-of-Words approach relied on domain-specific language, including words explicitly linked to substance use and mental health ("drink," "relapse," "depression," and so on), whereas the Linguistic Inquiry and Word Count approach relied on language characteristics such as tone, affect, insight, and presence of quantifiers and time references, as well as pronouns. A boosted decision tree classifier, utilizing features from both Bag-of-Words and Linguistic Inquiry and Word Count performed best in identifying problems disclosed within the discussion forum, achieving 88% sensitivity and 82% specificity in a separate cohort of patients in recovery.

Conclusions: Differences in language use can distinguish messages disclosing recovery problems from other message types. Incorporating machine learning models based on language use allows real-time flagging of concerning content such that trained staff may engage more efficiently and focus their attention on time-sensitive issues.

(J Med Internet Res 2018;20(6):e10136) doi: 10.2196/10136

KEYWORDS

self-help groups; substance-related disorders; supervised machine learning; social support; health communication

Introduction

Background

Digital health interventions have proliferated in recent years [1], and evidence suggests they can improve management of mental health issues, including substance use disorders (SUDs) [2,3]. Once their design is established, digital interventions can be disseminated with less expense and effort than face-to-face ones [4]. Such scalability is crucial for addressing SUDs, as demand for treatment dramatically outstrips available services [5]. In addition, although SUDs are chronic and relapsing [6,7], the help conveyed through technologies is ongoing and accessible. One recent clinical trial demonstrated that, relative to a control group, individuals who accessed a mobile phone–based recovery system reported reduction in risky drinking days by more than half over a year [8].

Substantial human labor also supports many effective digital health interventions. Some evidence suggests that, relative to interventions that lack human guidance, those that combine computerized tools with human support and coaching can enhance engagement and improve effectiveness of interventions [9]. The need for human expertize extends to interventions featuring peer-to-peer communication. Digital peer-to-peer interventions have involved "moderators" in various ways, including spurring and guiding discussion; monitoring forums for problematic content; and, crucially, providing just-in-time support to patients who are struggling, including through escalating contact or recommending treatment [10,11]. Through just-in-time support, moderators contribute to efficiency of health services at a systems level, making additional attention and resources available to those who most need them, while maintaining less intensive support for those at a lower risk level. Yet, attending to changing needs of an online health community poses a considerable challenge as participants can produce a massive volume of text exchanges [12]. Demands on staff represent a key hurdle in scaling up digital health interventions [13]. In this paper, we describe how automated linguistic analysis of text-based exchanges, and supervised machine learning, may play a role in managing moderator workflow in a technology-based recovery support system.

Our approach builds on the power of language as a signal of mental health risk, with linguistic cues being increasingly discernable through computational methods. Over the past several decades, researchers have amassed an extensive body of literature showing the promise of language to reveal individuals' psychological traits, thoughts, feelings, and likely behaviors [14], including in social media contexts [15]. As similar ideas can be conveyed in different ways, individuals' risk profiles emerge not only from the explicit content of their communication (ie, what topics authors are talking or writing about) but also from the style of their language (ie, *how* authors say what they say). In this study, we investigate how recovery challenges may emerge both through the individual words that authors use within a discussion forum as well as through general psycholinguistic dimensions of their messages (eg, affect, cognitive mechanisms), as captured through a dictionary-based approach. Leveraging these linguistic features, our goal is to find classifiers that can accurately label messages as conveying or not conveying recovery problems, allowing us to prioritize this content for review and intervention.

We consider several computational linguistic and machine learning approaches. First, we extract linguistic features of messages using 3 techniques: (1) a Bag-of-Words (BoW) approach representing each message in terms of word occurrences, (2) the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker Conglomerates, Austin, TX) [16], which computes rates of language use within validated dictionaries corresponding to psychological and linguistic concepts, and (3) a novel hybrid approach combining important features from BoW and LIWC. We propose that BoW and LIWC have complementary strengths, with BoW attending to important words specific to the dataset (eg, those related to substance use), whereas LIWC attends to relevant psychological states (eg, anxiety, self-focus). We expect that a hybrid approach, capitalizing on the strengths of each, should outperform either LIWC or BoW. We test these techniques in the context of supervised machine learning models that have been utilized in social media contexts: support vector machines (SVM), decision trees, and boosted decision trees.

In interpreting performance of our computational linguistic and machine learning approaches, we consider some particularities of the domain of addiction recovery support, namely: (1) a low tolerance for false negatives, (2) a preference for understandability of the method to stakeholders, and (3) efficiency in processing language and classifying messages in real-time. In other words, in addition to considering overall accuracy of each classifier, we ask: Does it miss too many worrisome messages to be useful to forum moderators? Does it have face validity to a team of health professionals? And can we successfully implement it in real-time? To establish the utility and robustness of our approach, we test our classifiers in a separate iteration of our mobile intervention involving a cohort of primary care patients with SUDs. We finally discuss implications of our findings for future research and system design, including how to improve model performance, and how classification can serve as the basis for directing attention and resources to those who need them.

Online Support Forums

SUDs are among the most common mental health disorders in the United States, with over 20 million adults affected in 2013 [5]. SUDs precipitate distress for sufferers and their communities, as well as serious health consequences [17,18]. Although many individuals with SUDs make attempts to stop using substances, resumption of risky substance use, or relapse, is extremely common [19]. With intensive SUDs treatments

being time-limited, it is crucial to find ways to extend recovery support to prevent relapse in the long term [20].

Mobile phones and internet use are now ubiquitous in the United States [21], with one consequence being that individuals with mental health challenges can access social support despite physical distance and at any time of day [22]. Often, this support comes from others who share the same mental health concern, as occurs via digital peer-to-peer forums where participants seek help on an "as needed" basis, and provide it to others [23,24]. Such forums typically involve anonymity or pseudo-anonymity, allowing for candid disclosure of personal and stigmatized issues and experiences [25]. Content analyses show that participants in SUDs forums disclose a variety of recovery challenges, prompting exchange of informational and emotional support [26-28].

The Role of Moderators

Although discussion forums offer a valuable arena for peer-to-peer exchange, moderators can also play a key role. For instance, those in recovery must manage their *own* health issues, limiting the time and energy that can be applied to help others [29]. In addition, although peers can offer first-hand experience related to coping and recovery, they may lack expertise necessary to guide decision making about clinical issues [30]. In contrast, moderators often have knowledge of intervention components and health behavior change processes and may recognize instances where contact or treatment is appropriate [31,32,10]. Moderators may additionally engage in pseudo-therapeutic activities such as offering emotion-focused support or assisting participants in reassessing dysfunctional perspectives [11] and may be more effective than peers in motivating individuals earlier in their behavior change process [33].

The presence of moderators in digital health forums has been associated with benefits. Notably, studies have found greater participation and expressiveness in moderated health forums relative to unmoderated ones [34,35]. In a mobile SUDs intervention for drug court participants, trained staff played a central role in discussion networks, with many participants communicating only with staff [36]. Prior work has also found that staff can enhance the success of digital mental health interventions regardless of formal clinical training [37].

Machine Learning Applications to Moderator Engagement

To "scale up" digital interventions, designers must take steps to support and streamline moderators' work. Fortunately, such efforts can make use of extensive data generated as participants engage with technologies. A rapidly growing research area centers on leveraging the digital traces of participants' activities to gain insights into the changing contexts within which participants are embedded and the psychological states they experience [38,39]. Digital trace data collected through mobile phones may include sensor data (eg, geolocation, accelerometry), as well as patterns of engagement with the intervention itself, and the content of messages exchanged.

By capturing spontaneous, first-hand accounts of authors' beliefs, feelings, and experiences, text-based messages offer particularly powerful insights into wellness, including the risk of mental health-related outcomes [15]. For instance, prior research has shown that linguistic qualities such as self-focus (as conveyed in pronoun use) can distinguish those who go on to post about suicidal ideation [40], and that negative affective language and swearing can identify individuals who go on to relapse in alcohol recovery [41]. These approaches rely on automated linguistic analysis as described in greater detail below.

Text-based features of user-submitted messages can now be efficiently extracted through a range of computational approaches. One of the most common approaches, BoW, involves representing each message in terms of occurrences of individual words, or "unigrams." After throwing out extremely common words, and grouping together words with the same stem, a message is represented as a vector formed by the occurrence rate of each stem, relative to that stem's overall occurrence in the full set of messages. In contrast, dictionary-based approaches search within a message for lists of words corresponding to relevant concepts. For instance, LIWC searches for words representing discussion topics (eg, health, family), psychological dimensions (eg, affect, cognition), and linguistic characteristics (eg, pronouns, conjunctions). LIWC then computes the percent of words in a given message that fall in each category. LIWC has been widely used in research, with studies showing that its categories predict health-related states including suicidality, depression, and dementia [42-44].

Relevant to this study, recent work also uses the above approaches to detect self-disclosure in online forums, defined as messages wherein participants convey personally relevant thoughts, feelings, and experiences [45-47]. In the context of support forums, self-disclosures offer a promising opportunity for intervention (eg, by moderators), because participants are revealing and working through personal issues, and may be actively seeking help [46]. The prior literature suggests that self-disclosure messages involve telltale linguistic cues that aid automatic detection. For example, one study identified a number of LIWC categories predictive of self-disclosure sensitivity, including third person pronouns and discussion of family, sex, death, and negative affect [48]. In another study, individual words conveying affect (eg, "happy," "love," and "hate") were characteristic of mental health–related self-disclosure [49].

Human expertise can also play an important role in guiding the development of language-based models. In supervised machine learning, an expert will designate a subset of messages as belonging to a category of interest (such as mental health risk), and the features of labeled messages are then used to predict whether an unlabeled message would fall in the same category. Labeled data can be generated in a number of ways. For instance, naturally occurring response patterns can be used, such as where Huh and colleagues [12] labeled as problematic those messages to which moderators had previously responded in a health support forum, using their linguistic features to classify new messages that moderators would likely be interested in. Alternately, human judgment can be used to generate each label in the training set, as was implemented in efforts to detect suicidality in an online discussion forum for youth [50,51]. This approach recognizes that moderators' response patterns do not

always clearly follow from the level risk a message indicates. For instance, responding to a message need not represent concern, but could reflect interest in a particular topic or investment in an ongoing relationship. Using SVM, boosted decision trees, and other models, researchers were able to achieve F-scores over 0.9 in identifying messages urgently requiring response [50,51].

This Study

In online addiction recovery forums, messages can be posted at any time of day or night, and some convey serious or time-sensitive problems. To offer timely help, moderators must continually review new content, but this task is demanding, with potentially dozens or hundreds of new posts to consider every day. Therefore, this study focuses on automatic detection of messages suggesting risk. For instance, interpersonal conflict, legal issues, personal traumas, or encounters with substance use cues could all represent threats to recovery in a substance abuse context [52,53]. Furthermore, these events could inspire psychological states associated with relapse, such as negative affect, cravings, or reduced self-efficacy [54]. Although circumstances and states can be conveyed in a variety of ways, prior literature leads us to anticipate that common language elements should emerge making recovery problems amenable to detection.

This paper contributes to the literature on digital SUDs interventions on several fronts. First, this work has practical application to efficiently capturing concerning content, so that forum moderators can respond in time. Efficient engagement by experts has been identified as a priority for extending the effectiveness of digital mental health interventions [13]. Second, the methodological contribution of our work involves comparing common computational linguistics and machine learning approaches and determining which are suited to the context of mental health risk in support forums.

As far as linguistic analysis, we compare performance of 2 techniques and their hybridization. First, BoW is driven by word-level usage in a given dataset and may therefore have an advantage for recovery-specific words (eg, “drink”). In contrast, our dictionary-based approach, LIWC, characterizes messages along general psychological and linguistic dimensions. Through building on prior knowledge about how words relate to established psychological constructs, LIWC offers potential efficiency, interpretability, and theoretical traction; however, its distinct disadvantage is that its dictionaries are not recovery specific. Thus, although LIWC contains a general category for “health,” it lacks dictionaries corresponding to concepts like “relapse” or “cravings.” Given these trade-offs, it is unclear whether BoW or LIWC will perform best.

Importantly, LIWC and BoW differ in their treatment of common words. The BoW framework retains words that are distinctive of the data at hand. Words with consistently high use across contexts, such as “I” and “we,” are considered insignificant within the BoW framework and typically discarded. In contrast, LIWC computes usage rates of these and other so-called “function” words (eg, pronouns, conjunctions, prepositions), which lack content but hold sentences together [55]. Despite their apparent banality, function words have proved powerful in predicting well-being, with pronouns receiving substantial attention in the mental health domain as a gauge of social integration [40,56,57]. Not surprisingly, personal pronouns also indicate self-disclosure, as they can show that individuals are talking about themselves [49]. In comparing BoW with LIWC and hybrid approaches, we therefore pay particular attention to performance improvements related to function words. In a more general sense, we aim to identify linguistic features most central to manifesting recovery problems, including discussion of substance use triggers, affective states, cognitive processes, and function words.

We also attempt to identify well-performing machine learning approaches. We focus on decision trees with and without boosting, as well as SVM, approaches with good performance in prior social media data [58,59].

Finally, we consider our results in relation to several key features of the domain of recovery support. First, recovery support is an arena where false negatives may be problematic, as missing an opportunity to intervene could allow a problem to escalate, even precipitating relapse. Therefore, in generating gold standard data, we emphasize the importance of establishing a reliable definition of “recovery problems” that is broad enough to capture potentially concerning content. We also reflect our concern about false negatives by prioritizing sensitivity in weighing classifier performance. Second, we seek machine learning methods that can offer insights into the particular language patterns associated with recovery. Decision trees may have an advantage in this regard, as they provide a visualization of the mechanisms of classification that may be helpful to establish face validity among stakeholders [59]. Finally, computational linguistics approaches have different implications for implementing classification in real-time, which we discuss.

Methods

Intervention

Data for this study came from a mobile phone–based intervention that provides on-demand services for recovery maintenance and relapse prevention. These services include informational pages, self-management tools (eg, self-help meeting directories, surveys), and peer-to-peer discussion forums. The intervention has been described in detail elsewhere, and it demonstrated efficacy in reducing risky drinking days by more than half relative to a control group [8]. We used data from 2 studies of the system: (1) a clinical trial involving individuals discharged from alcohol treatment (study 1) [8] and (2) an implementation study in primary care, involving individuals who used either alcohol or illicit drugs (study 2) [60]. The institutional review board at the University of Wisconsin-Madison approved both studies. Study participants provided informed consent for collection and use of their data for research (not shared beyond the team). These data included a log of all uses of the intervention and the content of communications exchanged within the intervention.

Study participants were provided with a mobile phone loaded with the intervention: either the Palm Pre with the Palm OS (Palm, Inc, Sunnyvale, CA) or an HTC Evo running Android

4.4 (HTC Corporation, Taiwan). In study 1, 130 participants posted on the forum. They were 56.2% (73/130) male and had a mean age of 38 years (SD 9.7). Participants wrote approximately 20 messages each (average length: 31 words). In study 2227 participants posted on the forum, and they were 53.3% (121/227) male, with a mean age of 42 years (SD 10.7). Participants wrote approximately 69 messages each (average length: 29 words).

This study focuses on text-based messages that were exchanged in the system's discussion forum, where participants could either start new threads on a topic of their choosing or respond to existing threads. All forum messages were visible to those on study, but study 1 forums were gender segregated. Moreover, 3 members of the research team also monitored the forums (authors GL, FM, and KP). Although the moderators lack clinical background, they are experts in digital health support for self-management of chronic conditions, including addiction recovery.

As mentioned earlier, gold standard data in this study substantially differ from those used in some prior work using moderators' natural response patterns [12]. We instead developed and applied a standardized, reliable codebook for capturing recovery risk. The first author first conducted an initial interview with the 3 moderators to understand their role in the forum and which messages would be considered worthy of intervention, and then consulted with them throughout the hand-labeling process to ensure our process captured messages of concern.

Computational Linguistics

We represented discussion forum messages using a BoW model, the LIWC program, and a hybrid approach.

The BoW approach represents each message in a feature space characterized by word counts. Common words were discarded, and remaining words were reduced to their stems using the Lancaster Stemmer from the NLTK stem package in Python and the NLTK word_punct tokenizer. For example, the stem "drink" would capture "drinking," "drinkin," "drinker," "drinks," and so on. We also wrote an additional filter to remove emoticons and other nonstandard characters. After grouping words according to their stems, Term Frequency-Inverse Document Frequency (TF-IDF) weighting was applied to calculate the occurrence rate of each specific stem in a message, offset by the importance of the stem in the entire corpus. Specifically, TF-IDF for a term is expressed as the term frequency (the number of times a word appears in a document divided by the total number of words in that document) multiplied by inverse document frequency (log of: total number of messages in a corpus divided by the number containing the term), thus adjusting for the fact that some words appear more frequently than others in general [61]. Once computed, the TF-IDF weights are used to form a vector representation of each message. After discarding common words, our BoW representation utilized 4247 unique unigrams as features.

The LIWC 2015 program computes rates of using words that fall within approximately 90 categories representing linguistic characteristics (eg, personal pronouns), topics of discussion (eg, family), affect (eg, anger), and cognitive processes (eg, insight) [16]. Each category corresponds to a predetermined dictionary of related words and word stems. Therefore, each message is represented as a 90-dimensional vector, with each dimension corresponding to a category such as "pronouns" and "positive affect." The value in each dimension is computed as the number of words from the message belonging to that category divided by the total number of words in the message. For example, "personal pronoun" is one of the features scored by LIWC. In the message "I am doing well," 1 out of the 4 words are personal pronouns, and so the LIWC score would be 1 out of 4 words or 25%.

In a hybrid approach, we exploit linguistic features from both BoW and LIWC. In other words, for a given message, word frequencies of the most important features from the TF-IDF matrix and the percentages falling in the most important linguistic categories from LIWC are stacked together to form a single feature vector. Given that combining too many features can inhibit performance by introducing noise [62], we utilized a subset of features from each representation. After ranking features according to their importance for a random forest model [63], we picked up to 10% of the most relevant features from BoW and LIWC to form a new feature set. Feature importance is calculated using the Gini Impurity measure, defined as the sum across the number of splits over all trees containing a feature, divided by number of samples in each split [64]. The hybrid approach included 310 features.

Machine Learning Techniques

With numeric representations of each message in our training set, and a corresponding label (recovery problem or no recovery problem), we trained 3 candidate binary classifiers for our task: SVM, decision trees, and boosted decision trees. SVM is a widely used technique and involves defining an optimal hyperplane to distinguish between items falling in classes of interest [65]. Decision trees involve segmenting the feature space into a number of simple regions [66]. In a series of decision steps, represented as branches, observations are made about an item (eg, the frequency at which a particular word is used within the message), leading to corresponding conclusions about the appropriate class (represented in the leaves). Finally, a related approach, boosted decision trees, involves an ensemble of decision trees where each tree learns by fitting the residual of the trees before it, allowing iterative improvement in performance. Python scikit-learn was used for machine learning [67].

As our datasets feature unbalanced classes (ie, messages including "recovery problems" are outnumbered by messages without them), we compensated for this imbalance by oversampling from the minority class. Specifically, we used the Synthetic Minority Oversampling Technique to generate synthetic samples from the minority class [68]. Rather than creating exact copies, the algorithm samples 2 or more similar instances, with similarity being calculated by a distance measure, (eg, Euclidean, Cosine), and then slightly perturbs these instances to create synthetic samples.

Once our classes were balanced, we trained our classifiers using labeled training data from study 1, the clinical trial for those

completing alcohol treatment (n=2581), and calculated parameters for each machine learning model using k-fold cross validation. Next, we tested the best performing models in labeled messages sampled from study 2 (n=800) with its primary care population. We report F-scores, as well as sensitivity (the proportion of correctly identified true positives), specificity (the proportion of correctly identified true negatives), and area under curve (AUC). We also describe example decision trees that illustrate classification logic.

Results

Identifying Recovery Problems

Our conversations with moderators first revealed that they recognized a wide range of issues and circumstances as warranting a response (relationship troubles, cravings, etc). Moderators expressed a fear of missing an important message, reporting a preference to have unconcerning messages flagged (false positives) than to miss actual problems (false negatives). Supporting our strategy of hand-labeling problem messages versus using prior response patterns as gold standard data, moderators also reported that contextual considerations influence their likelihood of responding on the forum. For instance, they might be unlikely to respond if participants had already received competent help from peers, or if they had personally had recent contact with participants outside the forum (eg, by phone call or private message). Moderators also stressed that they sometimes miss concerning messages inadvertently.

Guided by this feedback, 3 coders independently reviewed a preliminary set of 200 messages to identify ones they thought disclosed recovery challenges, broadly construed, and then mutually discussed their decisions. Coders arrived at consensus around a rule for coding the entire dataset, when "the writer describes a potential threat to well-being or recovery efforts." We further specified that the message may express either feeling vulnerable (eg, "I've been clean for about 7 months but even now I still feel like maybe I won't make it") or may outline a specific incident (eg, "it's not looking good, they are talking 0 to 5, and that's not days [in jail]. It's got my head all f.... up."). The coding rule also specified that the code should be applied even if the writer conveys that he or she has skills or abilities to handle a given problem (ie, a message may convey both a threat and mastery of that threat at the same time). Thus, by making the coding rule quite general, we avoided some subjectivity involved in making determinations about problems' seriousness. The first author next overlapped with each other coder on a set of 100 messages, allowing computation of interrater reliability, with average Cohen kappa of .77 for the 2 overlap sets deemed acceptably high [69].

Thus, our codebook captured recovery problems broadly construed. Results of hand-labeling revealed that of the 2581 messages posted to the forum over the course of the study 1, 388 (15%) disclosed some recovery problem. Review of these messages revealed themes including negative affect, cravings, and discouragement. Some described sleep problems, legal issues, medical concerns, unemployment, interpersonal conflict, financial worries, or housing. In a few cases, the writer simply shared that he or she was "struggling" or having a "hard time." Some messages relayed relapse. In contrast, messages not relaying recovery problems included small talk, affirmations, bonding, reports of doing well or feeling good, or giving support to others.

Supervised Machine Learning

To choose an optimal classifier and its parameters, we performed 10-fold cross validation on labeled data from study 1, partitioned into 70% training and 30% test datasets. Error metrics used were the average F-scores and AUC scores. Moreover, a total of 3 basic classifiers were considered (1) SVMs with linear and Gaussian kernels, (2) decision trees, and (3) boosted decision trees. Our results indicated that SVM performed worst with improvements in decision trees and best performance in boosted decision trees where we achieved F-scores of 0.88, 0.89, and 0.94 for the BoW, LIWC, and hybrid approaches, respectively. For the decision tree classifiers, we used tree depth of 3 and a minimum of 10 samples per leaf at termination when using the BoW feature space. When using the LIWC feature space, we used the same tree depth but a minimum of 8 samples per leaf at termination. For the hybrid feature space, we used a slightly deeper tree (depth=4) with a minimum of 11 samples per leaf at termination. Boosting utilized an average of 175 estimators across the 3 feature spaces.

Having set parameters, we trained on all data from study 1 and applied all 3 classifiers to test data in study 2. Recall that study 2 contained messages posted by a separate cohort of individuals with substance use disorders (in contrast to study 1 in which all individuals had alcohol abuse issues). F-scores for SVMs, decision trees, and boosted decision trees in test data are provided in Table 1.

Figures 1 and 2 show the top features extracted from the BoW and LIWC representations, respectively. For BoW, top features included words with the stems: drink, som (eg, some), because, hard, depress, feel, and hav (eg, have). For LIWC, top features are tone, clout, time, authenticity, analytic words, and insight words. Moreover, 3 top categories include pronoun forms.

Table 1. F-scores reported by 3 classifiers on the test data from study 2.

Classifier	BoW[a]	LIWC[b]	Hybrid
SVM[c]	0.76	0.71	0.76
Decision tree	0.8	0.75	0.77
Boosted decision tree	0.8	0.83	0.85

[a]BoW: Bag-of-Words.

[b]LIWC: Linguistic Inquiry and Word Count.

[c]SVM: support vector machines.

Figure 1. Fifteen most important feature words in the Bag-of-Words (BoW) framework.

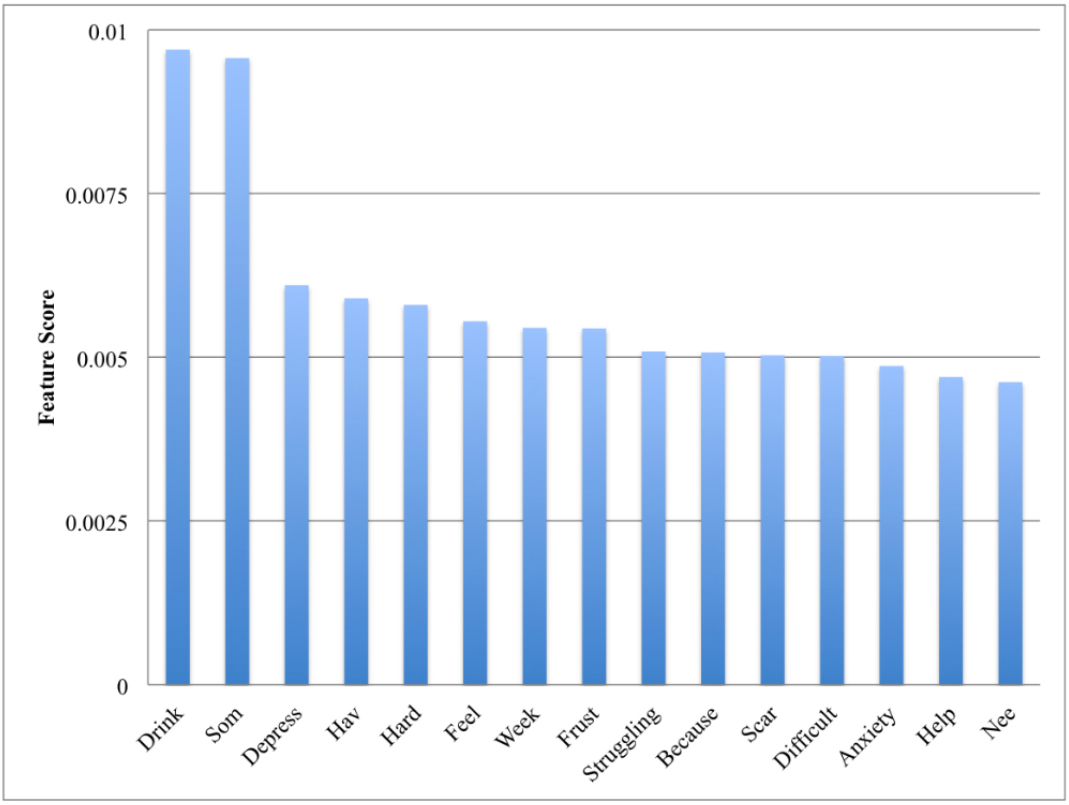
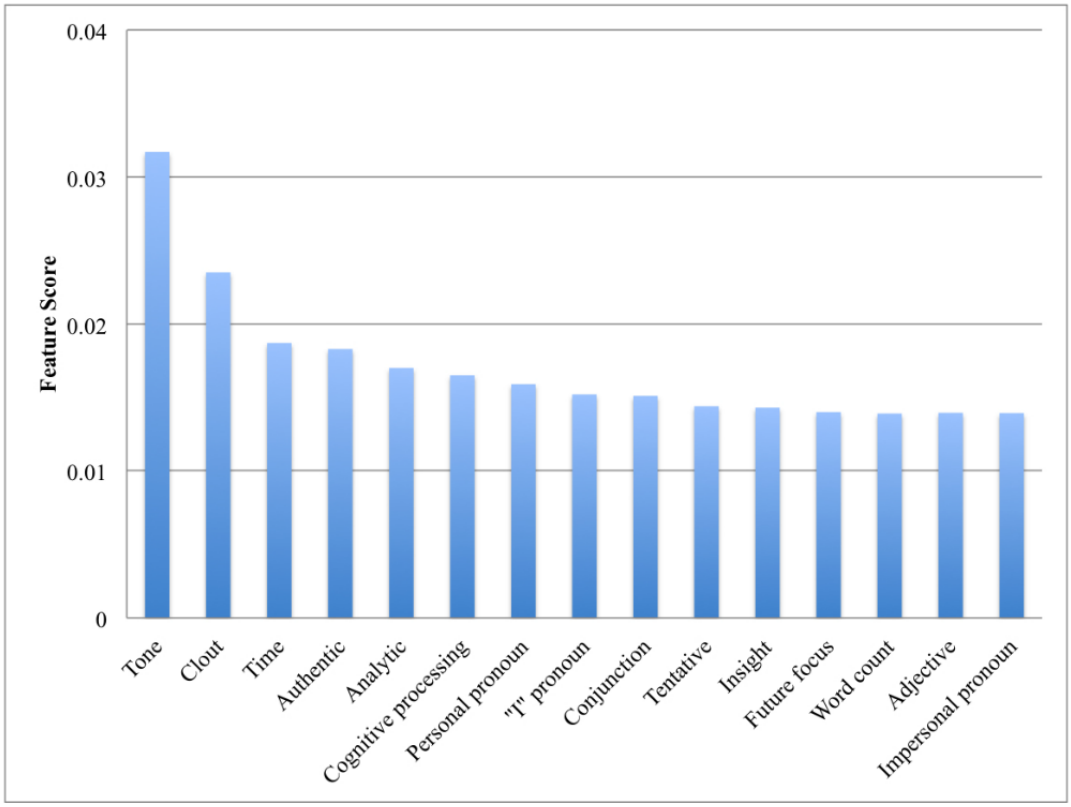


Figure 2. Fifteen most important features in the Linguistic Inquiry and Word Count (LIWC) framework.



To understand whether demographic characteristics (gender, age) would influence how recovery problems were expressed in language, we conducted additional experiments in the 2581 messages from study 1. In these experiments, we left 1 gender or age out of the training set, reserving this gender or age for a testing set. We used decision tree classifiers with feature representation from LIWC to test this question, finding an F-score=0.76 when training on the 1618 messages posted by women and testing on the messages posted by men, which is identical to cross-validation results achieved with the gender-mixed study 1 sample (F-score=0.76). We used the same approach for age, first leaving out 486 messages from those under 30 years, then 758 messages by those in their 30s, then 881 messages by those posted in their 40s, and finally 309 messages posted by those 50 years or older. The F-scores achieved were 0.77, 0.78, 0.77, and 0.73, respectively. Thus, they were roughly consistent with full study 1 cross-validation, although slightly lower for the 50 years or older group despite the training data being largest.

We produced decision trees for each approach to represent the relationship between language features in predicting recovery problems. For our models that involved boosting, multiple trees impact each classification decision, so any individual tree will provide only a small window into the logic of classification. Figures 3 and 4 depict truncated exemplar decision trees for the BoW and LIWC approaches. Text in speech bubbles represents messages that would be correctly classified as recovery problem (red) or not a recovery problem (green) by following the associated path. In Figure 3, we can see that the BoW decision tree begins with the stem "lot," with messages having an absence of the word "lot" (0.0 rate of "lot") following the "true" branch, and messages with presence of the stem "lot" following the "false" branch and being labeled as "recovery problem" (eg, "I've been drinking a lot lately"). For those messages not mentioning "lot," we next look for the stem "thank," the presence of which leads to a "no recovery problem" label. For those without "lot" or "thank," we look for "where," the presence of which would lead to a "recovery problem" label (eg, "Fighting with my bf again and I don't know where to go").

Figure 4 shows the exemplar LIWC tree, which begins with the category of feeling words, producing a categorization of "recovery problem" when paired with time words (eg, "I've been feeling not myself for the past week") but a "no recovery problem" label when mentions of time are below a minimum threshold (eg, "I'm feeling ok."). For messages without feeling words, the "recovery problem" label would be applied where anger words appear with quantity words (eg, "I'm so pissed!").

As boosted decision trees performed better than other classifiers, error analysis was summarized in detail for this classifier, with Table 2 providing specificity, sensitivity, and AUC achieved in the test data for each language processing approach. Results reveal that performance was somewhat improved for hybrid over LIWC and for LIWC over BoW (Table 2). More specifically, the hybrid outperforms the LIWC approach in terms of the F-score and the specificity, but not sensitivity, a point we return to below. The hybrid approach makes for an especially robust classifier as seen from the receiver operating characteristics (ROC) curves in Figure 5. BoW had the lowest sensitivity. For example, the following message was correctly identified by the hybrid and LIWC approaches and missed by BoW: "This is the hardest thing I have ever done. I just wish I felt better bout recovery. I'm nervous I'm gonna go back to my old ways."

Figure 3. Example decision tree using features from the Bag-of-Word (BoW) approach. Feature importance was calculated using the Gini Impurity measure.

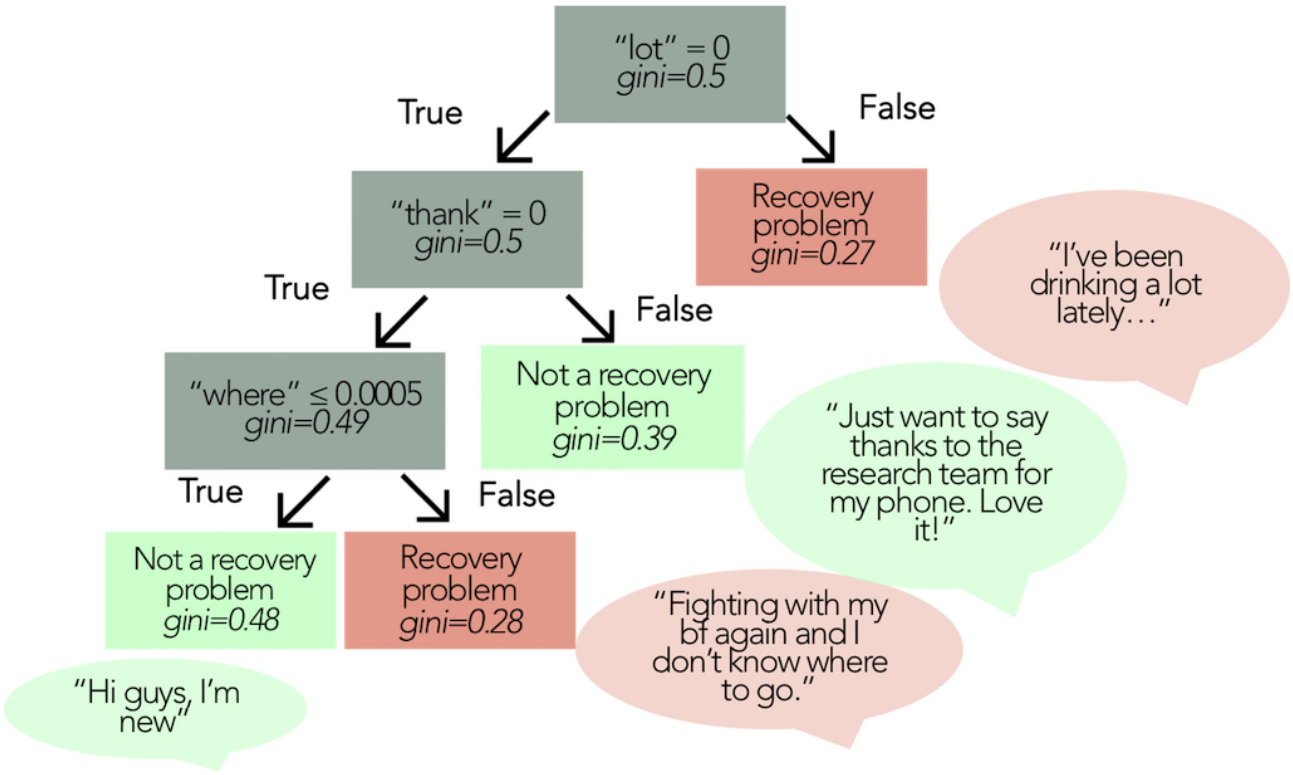


Figure 4. Example decision tree using features from the Linguistic Inquiry and Word Count (LIWC) approach. Feature importance was calculated using the Gini Impurity measure.

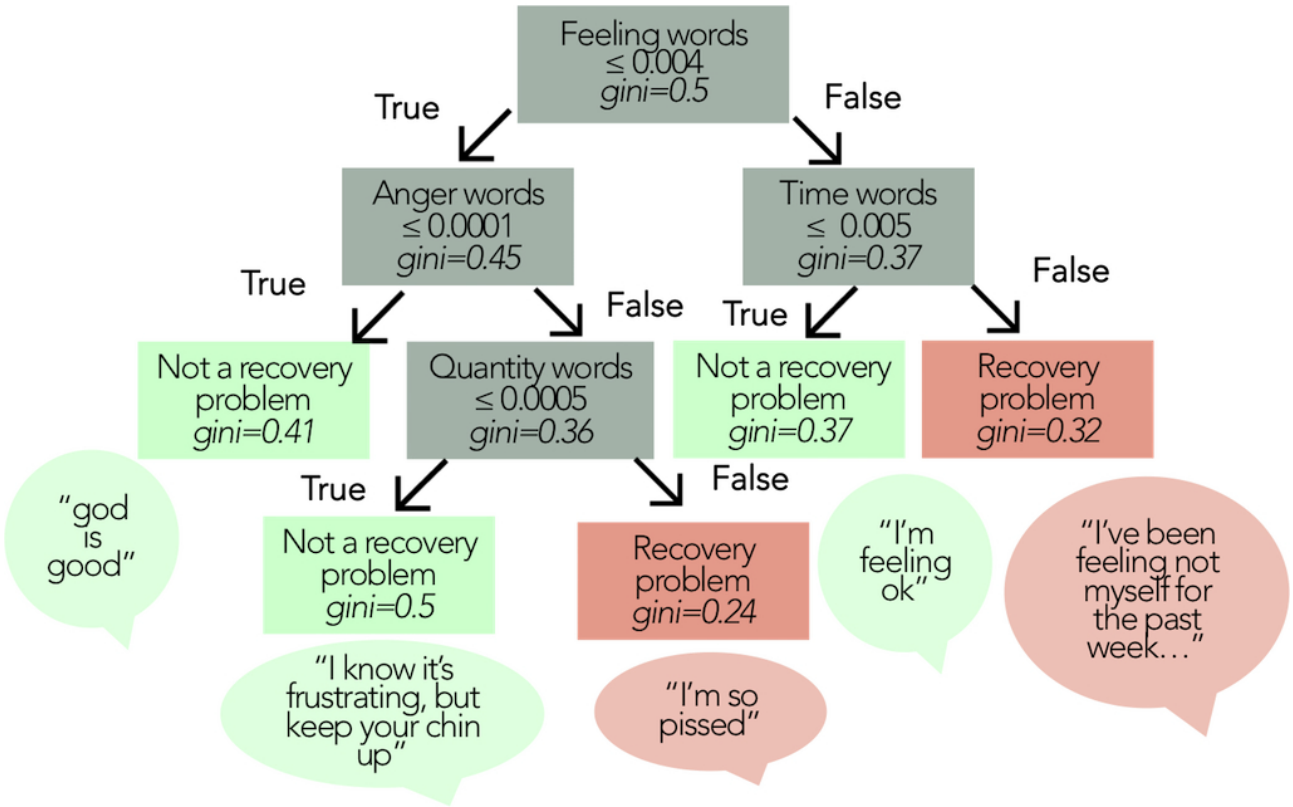


Table 2. Error analysis in study 2 for boosted decision trees using 3 language processing approaches.

Language processing approach	Sensitivity	Specificity	AUC[a]
BoW[b]	0.87	0.78	0.85
LIWC[c]	0.91	0.78	0.88
Hybrid	0.88	0.82	0.92

[a]AUC: area under curve.

[b]BoW: Bag-of-Words.

[c]LIWC: Linguistic Inquiry and Word Count.

Figure 5. Receiver operating characteristic (ROC) curves for boosted decision tree classifiers on the Bag-of-Words (BoW; left), Linguistic Inquiry and Word Count (LIWC; middle), and hybrid (right) feature spaces.

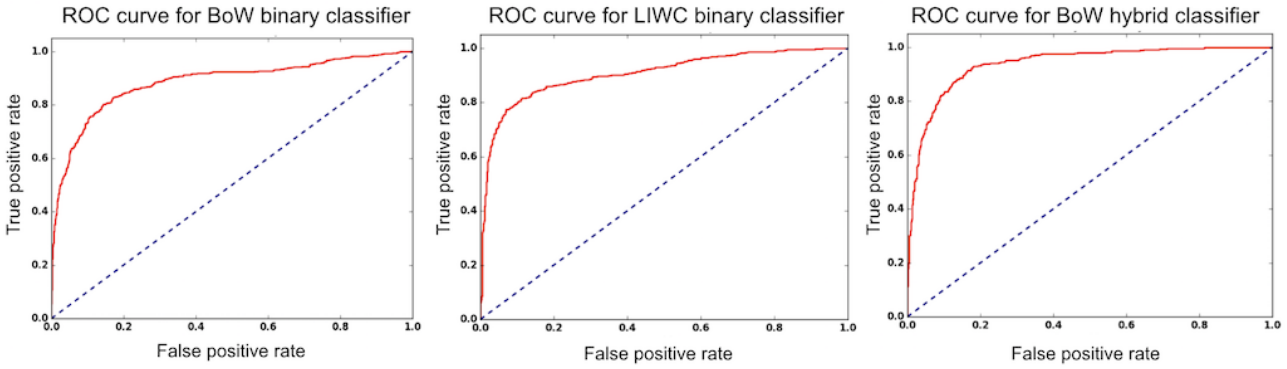
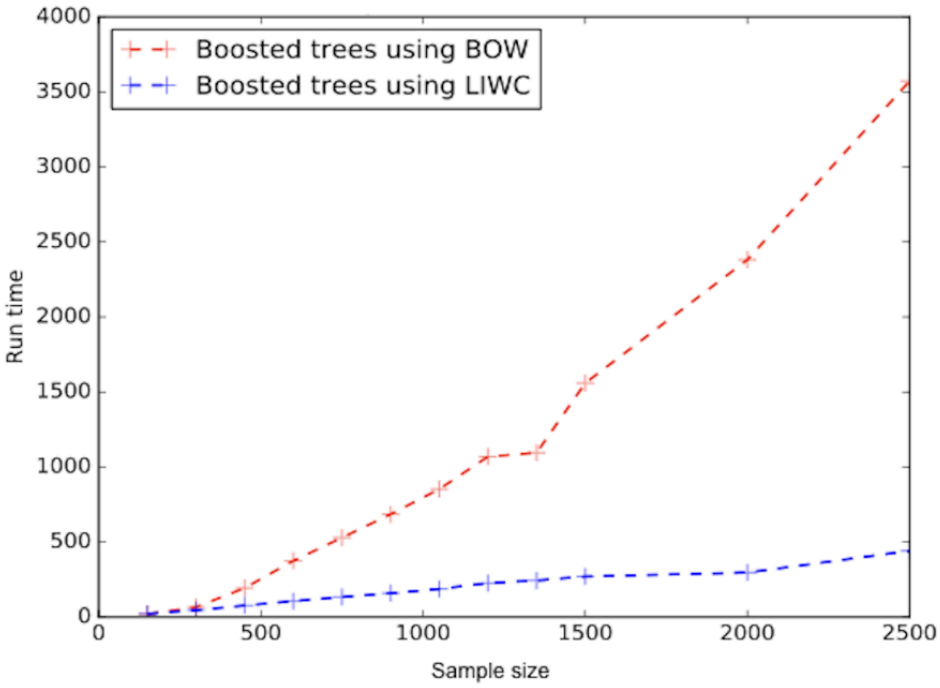


Figure 6. Training times of boosted decision tree classifiers on Bag-of-Words (BoW) and Linguistic Inquiry and Word Count (LIWC) feature spaces.



To understand this classification error, it is helpful to note that the message makes heavy use of personal pronouns such as "I," which appear in the top 10% of the LIWC features of importance in our datasets but contains just 1 highly ranked unigram from BoW ("hard"). By most metrics, the LIWC and hybrid approach outperform a BoW approach using a boosted decision tree model.

Finally, from Figure 6 we can see the training time for boosted decision trees on LIWC and BoW feature spaces, showing a huge speed advantage for LIWC, a result that was consistent across all classifiers.

Discussion

Principal Findings

The burden on mental health services has fueled recent growth of digital interventions, many of which involve text-based forums connecting a network of peers. Forums often operate with assistance from a moderator who steps in as needed, such as when participants' problems demand more formal intervention or could overwhelm peers' abilities to help. Although moderators play an important role, their time-consuming work limits the scalability of digital interventions. This study demonstrates a solution that facilitates moderation efficiency while reducing the possibility of overlooking messages of concern: a machine learning-based model to automatically flag messages that disclose recovery problems. For this work, we used several machine learning approaches, with boosted decision trees performing best, while also offering a view into the logic of classification that may be helpful in establishing face validity.

We also represented our data through a number of computational linguistics techniques. Although the BoW approach captured domain-specific language, it performed somewhat worse than LIWC, a dictionary-based approach capturing psycholinguistic features. LIWC may do well in this context because recovery problems have important affective dimensions; prior literature shows that LIWC may perform well in cases where affect is a dominant theme [42]. We further found that a hybrid approach, leveraging a combination of features from the dictionary-based LIWC program and BoW, performed best for classifying our test data with regard to AUC and F-score. However, these improvements were only marginally improved over LIWC alone. LIWC achieved a similar F-score of 0.83 (compared with 0.85 for the hybrid) and actually had a higher sensitivity of 0.91.

Ideally, analysts often seek solutions that maximize performance as measured by the F-score, which in this case points toward the hybrid approach. However, there are times when an analyst might prefer greater sensitivity (avoiding false negatives) over improved specificity (avoiding false positives), including perhaps the context of addiction treatment and other health contexts where missing problematic messages could be costly. Given our desire for high sensitivity, LIWC may even be a preferable option over the hybrid. To put this in practical terms, LIWC correctly classified 116 out of 127 true positives in our study 2 test data, compared with 112 classified by the hybrid approach. These additional 4 messages came at the expense of an additional 29 false negatives. However, given the potential consequences of a missed true positive, the additional review time may be seen as worthwhile, especially in early stages of implementing this sort of classifier when concerns about missing actionable messages may limit adoption. Digital mental health interventions seek to ease burden on providers while delivering care to patients, but adoption requires faith in system performance among those on the "front lines."

LIWC may also be preferable given its easier real-time implementation. Our experiments showed that LIWC features enabled faster training than BoW. LIWC may also have an implementation advantage as the BoW approach involves calculating TF-IDF scores that reflect the occurrence rate of the word in a single document as well as that word's occurrence across all documents in a sample. This suggests that BoW may present computational challenges when applied in a "live" forum, as the overall occurrence for unigrams may change as new messages are posted and as new unigrams may emerge over time. On the other hand, as the LIWC dictionaries are broad and fixed, classifiers may work well in a system where messages are continually added.

We found that our classifiers were flexible enough to capture numerous circumstances that present problems in recovery, including interpersonal conflicts, job and housing instability, feelings of hopelessness, and encountering triggers. Despite the variety of problems described, classifiers relied heavily on particular ways of talking about drinking, affect, and context, as evident from the important features extracted for each method. Decision rules using the BoW approach were sometimes based on weights of words explicitly linked to drinking ("drink," "relapse," "sobriety," and so on), but decision trees also revealed the use of certain location and context-related terms ("stay" and "where") in decision rules. Decision rules from the LIWC approach were rarely based on explicit topics of discussion, but instead reflected characteristics such as tone, affect, insight, and presence of quantifiers and time references, as well as pronouns. In a tree-based approach, it is not simply using words within these categories that matters but couse with words in other categories.

Comparison With Prior Work

Notably, in all cases, we achieved good performance relative to Huh and colleagues [12], who also attempted to detect appropriate messages for moderator intervention, and who achieved F-scores up to 0.54. This may in part reflect the difference in machine learning approach, as we used boosted decision trees rather than the Naive Bayes technique they report. The improvement may also reflect the labeling process for training data. Specifically, their training approach labeled messages according to whether they actually received a moderator response, presuming these to be messages of greatest concern, but we implemented a reliable human coding process that we thought would minimize error, as moderators' responses are actually driven by a number of factors beyond the level of concern a message produces.

Indeed, our results are more closely in line with studies that have used hand-labeled data for training. F-scores for our hybrid model are comparable with the best results achieved in a shared task challenge to flag messages for elevated suicide risk in a forum for Australian youth [50] and slightly lower than a follow-up study from the same forum that utilized an ensemble of feature extraction approaches (LIWC, topic modeling, meta-data, etc) [51]. However, it is important to note our more conservative approach of testing our model in a separate iteration of the forum with a separate patient population. Like Conan et al [51], we also obtained better results for boosted decision trees relative to SVM.

Implications for System Design

Moderators can play a pivotal role in digital forums for at-risk populations but face difficulties keeping up with new content. Recently, scholars have called for improving digital health interventions by emphasizing *efficiency* of human support: the level of increased engagement and intervention effectiveness relative to the effort expended by staff [13]. Our findings demonstrate an opportunity to improve efficiency through automatically identifying, in real time, when participants disclose pressing concerns. Resulting classifications could be easily used to populate an interface to display high-priority content to moderators (see Multimedia Appendix 1 for an example of how our classifier has been applied in our live mobile-based recovery support platform). The interface may also provide moderators with an opportunity to dispute message classifications they view as erroneous, generating data to refine classifiers in the future (See Multimedia Appendix 2). Upon review of flagged messages, moderators might choose to intervene in a number of ways, such as through providing emotional support, directing participants to intervention elements that might suit their needs, or connecting participants with mentors or services.

Although our present solution requires human review and response, it is worth noting an alternate approach of fully automating responses. For instance, flagged messages could prompt the system to provide immediate contact information for treatment providers or emergency services, thus offering support even late at night and early in the morning. Some systems have also used machine learning methods to match newly posted content to semantically similar earlier content, displaying these older messages alongside the responses they generated in case they are useful to the current poster [70]. In addition, more complex dialogue systems have been applied to further reduce the human labor behind digital health interventions, including interactive "conversational agents," software programs that mimic human conversation, and that may further display human-like cues through voice or visual representations [71,72]. Such techniques are promising but involve trade-offs relative to trained staff who develop personal relationships with participants and can exercise expert judgment [73]. For instance, unlike software programs, moderators can choose to ignore messages they believe are "false positives," not warranting their expression of concern. Of course, moderators also vary in personal and professional qualities that make them effective. For instance, staff may be particularly successful through conveying a combination of trustworthiness, benevolence, and expertise [74].

In the future, efficient just-in-time support may involve judicious use of both human support and automated messages. Short of full automation, efficiency could be enhanced through providing moderators with a drop-down list of common responses that may be appropriate after a problem is disclosed, with an editor allowing optional personalization. Information about a given participant (eg, risk score from the last completed survey) could also indicate whether a flagged message should be sent to the moderator for a personalized response or managed through automation.

Future Research Directions

Findings from this study suggest promising areas for future research. First, a number of additional optimizations of our classifiers may be possible. For instance, additional dictionaries have also been developed in the realm of electronic medical records and these could prove promising in capturing recovery-related concepts [75]. Conditional Random Fields methods also work well in classifying natural language [76]. In future, we may also improve our BoW-based model through attention to dimensionality reduction, latent semantic analysis, and potentially extracting bigrams (or trigrams, etc) in addition to unigrams. As far as our hybrid approach is concerned, we might further optimize performance by giving further consideration to the number of features pulled from each component method. Specifically, to determine the number of important features from LIWC and BoW to include in the hybrid model, we tested cut-off points at 5% intervals (10%, 15%, 20%, etc) and found the best results for 10%, but more fine-grained adjustments could be tested, including plotting F-score relative to the number of features, and perhaps allowing for different cut-offs for LIWC and BoW.

Although our models were robust regardless of type of substance of abuse (which varied across Studies 1 and 2) and by gender, our leave-one-out experiments suggest that further research may also be needed to understand if older adults use similar language to convey recovery problems. We also did not test our model across differences such as race or education, leaving it unclear whether our models would work well in populations of different compositions.

Models might also take additional data into account. This analysis was conducted at the message level, but it may be possible to improve our models by considering each individual's pattern of messaging. Those who habitually post recovery problems may require a different level and style of response than those who escalate posting of worrisome messages. Other system use or sensor data may also inform our model, such that patterns of reading messages, interacting with intervention features (eg, pressing a "panic button"), or moving to new geographic locations may be integrated into decision rules around moderator involvement [77]. Similar work in the domain of suicide risk has incorporated additional features reflecting metadata from the discussion forum (eg, How many usernames are referenced in a message? Where does a message fall in sequence within a thread?) [51].

Ultimately, the efficiency of our approach to flagging concerning messages should be addressed empirically, such as through a trial randomizing some participants to a system where moderators manually review the forum and others to a system where moderators rely on text-based classification. Outcomes may include moderators' workload as well as patients' satisfaction and health outcomes. Further research is also needed to establish how to best intervene after a recovery problem message, including through personalized responses from moderators or automated messages.

A final future direction relates to privacy. Our surveillance approach offers opportunities to intervene early to help those in need, but introduces an important trade-off as far as privacy.

Specifically, we use passively collected data to infer underlying risk levels that patients may not even be aware of, with these data being highly sensitive [50]. Future research is needed to clarify how patients understand uses of their data for surveillance, how they balance surveillance and privacy concerns, and the contexts under which they find surveillance acceptable. In this study, it is possible that we allayed some privacy concerns by recruiting patients through trusted treatment providers and clinicians and obtaining informed consent, but patients may have greater privacy concerns in the domain of commercial mental health platforms.

Limitations

This study has limitations. First, our approach would not allow us to assist participants who do not post on a discussion forum. Furthermore, we do not look at private messages, where participants potentially disclose even more sensitive information [50]. In addition, as we did not label subtypes of recovery problems, it is possible that our classifier may be biased toward recognizing certain types of common problematic messages over others. Future work should consider coding subtypes of recovery problems. For instance, relatively rare problems that are nonetheless highly concerning may include mentions of suicide risk or solicitations to buy or sell drugs. Finally, one of the core strengths of our dataset is also tied to one of our study limitations. Specifically, we have access to a dataset of anonymous messages exchanged in a system restricted to those who share a SUDs diagnosis (a condition of study eligibility). These factors mean that discussion may be particularly candid and may offer unusual insight into mental health risk. At the same time, these considerations imply that existing labeled datasets cannot easily be adapted to train classifiers within our dataset. Our model leverages a relatively small set of training messages, which has implications for the machine learning approaches available and the results obtained.

Conclusions

Digital interventions hold promise to offer cost-effective, constantly available support to those in recovery, and to reduce human workload relative to face- to-face SUDs interventions. However, human support still plays a vital role in many effective digital interventions. For interventions involving discussion forums, trained moderators can respond in real time to help participants who are facing challenges. Yet, these moderators must dedicate substantial time and effort to manually review newly posted messages to identify serious problems, and the process can be error-prone. Our results show that message content can be effectively leveraged toward facilitating just-in-time supportive intervention. Language-based classification models have potential for massive scalability as digital interventions for addiction support continue to expand.

Individuals' language use, both through its content and composition, offers a means of understanding psychological states and traits. Our work expands on the existing literature by combining and layering computational linguistics and machine learning techniques in the context of streamlining human support within digital substance abuse recovery interventions. Yet, this work also has theoretical and methodological value beyond this specific context, suggesting useful directions for applying language classification to digital mental health more broadly.

Acknowledgments

This research was funded by the National Institute of Alcohol Abuse and Alcoholism (R01 AA017192) and the National Institute on Drug Abuse (R01DA034279, R01DA040449, and DP2DA042424). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funder had no role in any aspect of the development, conduct, analysis, or reporting of the study. The authors wish to thank Adam Maus, Ryan Westergaard, Alexandria Cull, Yan Liu, and Andrew Quanbeck for their important contributions to the project.

Conflicts of Interest

DHG has a shareholder interest in CHESS Mobile Health, a public benefit corporation that develops health care technology for patients struggling with addiction. DHG and the University of Wisconsin–Madison's Conflict of Interest Committee manage this relationship. All other authors have no conflicts of interests to declare.

Multimedia Appendix 1

Interface for moderator to review messages classified as indicating recovery problems.

[PDF File (Adobe PDF File), 145KB-Multimedia Appendix 1]

Multimedia Appendix 2

Interface for moderator to provide feedback on message classification.

[PDF File (Adobe PDF File), 138KB-Multimedia Appendix 2]

References

1. Barak A, Grohol JM. Current and future trends in internet-supported mental health interventions. J Technol Hum Serv 2011 Jul;29(3):155-196. [doi: 10.1080/15228835.2011.616939]

2. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
3. Moore BA, Fazzino T, Garnet B, Cutter CJ, Barry DT. Computer-based interventions for drug use disorders: a systematic review. J Subst Abuse Treat 2011 Apr;40(3):215-223 [FREE Full text] [doi: 10.1016/j.jsat.2010.11.002] [Medline: 21185683]
4. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
5. Substance Abuse and Mental Health Services Administration. SAMHSA. 2015. Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvUm7BiF]
6. Quanbeck A, Chih M, Isham A, Gustafson D. Mobile delivery of treatment for alcohol use disorders: a review of the literature. Alcohol Res 2014;36(1):111-122 [FREE Full text] [Medline: 26259005]
7. Litvin EB, Abrantes AM, Brown RA. Computer and mobile technology-based interventions for substance use disorders: an organizing framework. Addict Behav 2013 Mar;38(3):1747-1756. [doi: 10.1016/j.addbeh.2012.09.003] [Medline: 23254225]
8. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. J Am Med Assoc Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
9. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions—a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
10. Greidanus E, Everall RD. Helper therapy in an online suicide prevention community. Br J Guid Counc 2010 May;38(2):191-204. [doi: 10.1080/03069881003600991]
11. Gilat I, Tobin Y, Shahar G. Responses to suicidal messages in an online support group: comparison between trained volunteers and lay individuals. Soc Psychiatry Psychiatr Epidemiol 2012 Dec;47(12):1929-1935. [doi: 10.1007/s00127-012-0508-7] [Medline: 22491905]
12. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. J Biomed Inform 2013 Dec;46(6):998-1005 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.011] [Medline: 24025513]
13. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol Sci Pract 2016 Nov 17;24(1):27-45. [doi: 10.1111/cpsp.12173]
14. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
15. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
16. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. Austin, TX: The University of Texas at Austin; 2015. The development and psychometric properties of liwc2015 URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2018-04-24] [WebCite Cache ID 6yvUyNkq1]
17. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. Lancet 2009 Jun 27;373(9682):2223-2233. [doi: 10.1016/S0140-6736(09)60746-7] [Medline: 19560604]
18. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
19. Dennis M, Scott CK. Managing addiction as a chronic condition. Addict Sci Clin Pract 2007 Dec;4(1):45-55 [FREE Full text] [Medline: 18292710]
20. Kelly JF, White WL. Recovery management and the future of addiction treatment and recovery in the USA. In: Kelly JF, White WL, editors. Addiction Recovery Management: Theory, Research and Practice (Current Clinical Psychiatry). Totowa, NJ: Humana Press; 2010:303-316.
21. Pew Research Center. Demographics of Mobile Device Ownership and Adoption in the United States URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-02-14] [WebCite Cache ID 6xEcZNiMF]
22. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. Br Med J 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
23. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
24. Heisler M, Vijan S, Makki F, Piette JD. Diabetes control with reciprocal peer support versus nurse care management: a randomized trial. Ann Intern Med 2010 Oct 19;153(8):507-515 [FREE Full text] [doi: 10.7326/0003-4819-153-8-201010190-00007] [Medline: 20956707]

25. DeAndrea DC, Anthony JC. Online peer support for mental health problems in the United States: 2004-2010. Psychol Med 2013 Nov;43(11):2277-2288 [FREE Full text] [doi: 10.1017/S0033291713000172] [Medline: 23410539]
26. Cunningham JA, van Mierlo T, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. Patient Educ Couns 2008 Feb;70(2):193-198. [doi: 10.1016/j.pec.2007.10.003] [Medline: 18022340]
27. Klaw E, Dearmin Huebsch P, Humphreys K. Communication patterns in an online mutual help group for problem drinkers. J Community Psychol 2000;28(5):535-546. [doi: 10.1002/1520-6629(200009)28:5<535::AID-JCOP7>3.0.CO;2-0]
28. Chuang KY, Yang CC. Interaction patterns of nurturant support exchanged in online health social networking. J Med Internet Res 2012;14(3):e54 [FREE Full text] [doi: 10.2196/jmir.1824] [Medline: 22555303]
29. Arnstein P, Vidal M, Wells-Federman C, Morgan B, Caudill M. From chronic pain patient to peer: benefits and risks of volunteering. Pain Manag Nurs 2002 Sep;3(3):94-103. [Medline: 12198640]
30. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
31. Huh J, Patel R, Pratt W. Tackling dilemmas in supporting "the whole person" in online patient communities. Proc SIGCHI Conf Hum Factor Comput Syst 2012;2012:923-926 [FREE Full text] [doi: 10.1145/2207676.2208535] [Medline: 24634893]
32. Mares M, Gustafson DH, Glass JE, Quanbeck A, McDowell H, McTavish F, et al. Implementing an mHealth system for substance use disorders in primary care: a mixed methods study of clinicians' initial expectations and first year experiences. BMC Med Inform Decis Mak 2016 Sep 29;16(1):126 [FREE Full text] [doi: 10.1186/s12911-016-0365-5] [Medline: 27687632]
33. De Vries RA, Zaga C, Bayer F, Drossaert CH, Truong KP, Evers V. Experts get me started, peers keep me going: comparing crowd- versus expert-designed motivational text messages for exercise behavior change. : EAI; 2017 Presented at: International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain p. 155-162. [doi: 10.1145/3154862.3154875]
34. Klemm P. Effects of online support group format (moderated vs peer-led) on depressive symptoms and extent of participation in women with breast cancer. Comput Inform Nurs 2012 Jan;30(1):9-18. [doi: 10.1097/NCN.0b013e3182343efa] [Medline: 22240564]
35. Lieberman MA, Golant M, Winzelberg A, McTavish F, Gustafson DH. Comparisons: professionally-directed and self-directed internet groups for women with breast cancer. Int J Self Help 2004;2(3):219-235. [doi: 10.2190/GE85-J31W-XJV7-LB9L]
36. Johnson K, Richards S, Chih MY, Moon TJ, Curtis H, Gustafson DH. A pilot test of a mobile app for drug court participants. Subst Abuse 2016;10:1-7 [FREE Full text] [doi: 10.4137/SART.S33390] [Medline: 26917964]
37. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
38. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Transl Behav Med 2015 Sep;5(3):335-346 [FREE Full text] [doi: 10.1007/s13142-015-0324-1] [Medline: 26327939]
39. Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing adaptive treatment strategies in substance abuse research. Drug Alcohol Depend 2007 May;88 Suppl 2:S24-S30 [FREE Full text] [doi: 10.1016/j.drugalcdep.2006.09.008] [Medline: 17056207]
40. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:2098-2110 [FREE Full text] [doi: 10.1145/2858036.2858207] [Medline: 29082385]
41. Kornfield R, Toma CL, Shah DV, Moon TJ, Gustafson DH. What do you say before you relapse? How language use in a peer-to-peer online discussion forum predicts risky drinking among those in recovery. Health Commun 2017 Aug 09:1-10. [doi: 10.1080/10410236.2017.1350906] [Medline: 28792228]
42. Brancu M, Jobes D, Wagner BM, Greene JA, Fratto TA. Are there linguistic markers of suicidal writing that can predict the course of treatment? A repeated measures longitudinal analysis. Arch Suicide Res 2016 Jul 02;20(3):438-450. [doi: 10.1080/13811118.2015.1040935] [Medline: 26219609]
43. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France p. 3267-3276. [doi: 10.1145/2470654.2466447]
44. Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno-Tempini ML, et al. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. 2014 Presented at: ACL Workshop on Computational Linguistics and Clinical Psychology; June 27, 2014; Baltimore, MD p. 27-36. [doi: 10.3115/v1/W14-3204]
45. Bak JY, Lin CY, Oh A. Self-disclosure topic model for classifying and analyzing Twitter conversations. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October 25-29, 2014; Doha, Qatar p. 1986-1996. [doi: 10.3115/v1/D14-1213]

46. Balani S, De Choudhury M. Detecting and characterizing mental health related self-disclosure in social media. : ACM; 2015 Presented at: ACM Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, Korea p. 1373-1378. [doi: 10.1145/2702613.2732733]
47. Wang YC, Burke M, Kraut R. Modeling self-disclosure in social networking sites. : ACM; 2016 Presented at: 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; February 27-March 02, 2016; San Francisco, CA, USA p. 74-85. [doi: 10.1145/2818048.2820010]
48. Houghton DJ, Joinson AN. Linguistic markers of secrets and sensitive self-disclosure in Twitter. 2012 Presented at: 45th Hawaii International Conference on System Science (HICSS); Jan 4-7, 2012; Maui, Hawaii USA p. 3480-3489. [doi: 10.1109/HICSS.2012.415]
49. De Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. 2014 Presented at: AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI URL: https://pdfs.semanticscholar.org/2db7/15a479c8961d3020fe906f7bedfa0311b937.pdf
50. Milne DN, Pink G, Hachey B, Calvo RA. Clpsych 2016 shared task: Triaging content in online peer-support forums. In: Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2016 Presented at: Workshop on Computational Lingusitics and Clinical Psychology; June 16, 2016; San Diego, California, USA p. 118-127. [doi: 10.18653/v1/W16-0312]
51. Cohan A, Young S, Yates A, Goharian N. Triaging content severity in online mental health forums. J Assoc Inf Sci Technol 2017 Sep 25;68(11):2675-2689. [doi: 10.1002/asi.23865]
52. Chuang KY, Yang CC. Informational support exchanges using different computer-mediated communication formats in a social media alcoholism community. J Assoc Inf Sci Technol 2013 Oct 23;65(1):37-52. [doi: 10.1002/asi.22960]
53. Campbell SW, Kelley MJ. Mobile phone use among Alcoholics Anonymous members: new sites for recovery. New Media Soc 2008 Dec;10(6):915-933. [doi: 10.1177/1461444808096251] [Medline: 22973420]
54. Witkiewitz K. Predictors of heavy drinking during and following treatment. Psychol Addict Behav 2011 Sep;25(3):426-438 [FREE Full text] [doi: 10.1037/a0022889] [Medline: 21480681]
55. Chung C, Pennebaker JW. The psychological functions of function words. In: Fiedler K, editor. Frontiers of Social Psychology. Social communication. New York, NY, US: Psychology Press; 2007:343-359.
56. Zimmermann J, Wolf M, Bock A, Peham D, Benecke C. The way we refer to ourselves reflects how we relate to others: associations between first-person pronoun use and interpersonal problems. J Res Pers 2013 Jun;47(3):218-225. [doi: 10.1016/j.jrp.2013.01.008] [Medline: 25904163]
57. Anderson B, Goldin PR, Kurita K, Gross JJ. Self-representation in social anxiety disorder: linguistic analysis of autobiographical narratives. Behav Res Ther 2008 Oct;46(10):1119-1125 [FREE Full text] [doi: 10.1016/j.brat.2008.07.001] [Medline: 18722589]
58. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: Detecting influenza epidemics using twitter. : Association for Computational Linguistics; 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 27-29, 2011; Edinburgh, UK p. 1568-1576.
59. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for Twitter data against established measures of suicidality. JMIR Ment Health 2016 May 16;3(2):e21 [FREE Full text] [doi: 10.2196/mental.4822] [Medline: 27185366]
60. Quanbeck A, Gustafson DH, Marsch LA, Chih MY, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
61. Sohn S, Wang Y, Wi C, Krusemark EA, Ryu E, Ali MH, et al. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. J Am Med Inform Assoc 2017 Nov 30:-. [doi: 10.1093/jamia/ocx138] [Medline: 29202185]
62. Cantú-Paz E, Newsam S, Kamath C. Feature selection in scientific applications. 2004 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 22-25, 2004; Seattle, WA, USA. [doi: 10.1145/1014052.1016915]
63. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. Pattern Recognit Lett 2010 Oct;31(14):2225-2236. [doi: 10.1016/j.patrec.2010.03.014]
64. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
65. Tong S, Koller D. Support vector machine active learning with applications to text classification. J Mach Learn Res 2002;2:45-66. [doi: 10.1162/153244302760185243]
66. Apté C, Weiss SM. Data mining with decision trees and decision rules. Future Gener Comput Syst 1997 Nov;13(2-3):197-210. [doi: 10.1016/S0167-739X(97)00021-6]
67. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
68. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321-357. [doi: 10.1613/jair.953]

69. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
70. Wang Y, Mehrabi S, Mojarad MR, Li D, Liu H. Retrieval of Semantically Similar Healthcare Questions in Healthcare Forums. : IEEE; 2015 Presented at: International Conference on Healthcare Informatics (ICHI); October 21-23; Dallas, TX, USA p. 517-518. [doi: 10.1109/ICHI.2015.97]
71. Bickmore T, Giorgino T, Green N, Picard R. Special issue on dialog systems for health communication. J Biomed Inform 2006 Oct;39(5):465-467 [FREE Full text] [doi: 10.1016/j.jbi.2006.02.002] [Medline: 16546453]
72. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. J Am Med Assoc Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
73. Kasckow J, Zickmund S, Rotondi A, Mrkva A, Gurklis J, Chinman M, et al. Development of telehealth dialogues for monitoring suicidal patients with schizophrenia: consumer feedback. Community Ment Health J 2014 Apr;50(3):339-342. [doi: 10.1007/s10597-012-9589-8] [Medline: 23306676]
74. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
75. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
76. Nakagawa T, Inui K, Kurohashi S. Dependency tree-based sentiment classification using CRFs with hidden variables. : Association for Computational Linguistics; 2010 Presented at: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics; June 2-4, 2010; Los Angeles, California, USA p. 786-794.
77. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas L, Wingood G, et al. A computational future for preventing HIV in minority communities: how advanced technology can improve implementation of effective programs. J Acquir Immune Defic Syndr 2013 Jun 1;63 Suppl 1:S72-S84 [FREE Full text] [doi: 10.1097/QAI.0b013e31829372bd] [Medline: 23673892]

Abbreviations

AUC: area under curve
BoW: Bag-of-Words
LIWC: Linguistic Inquiry and Word Count
ROC: receiver operating characteristics
SUD: substance use disorder
SVM: support vector machines
TF-IDF: Term Frequency-Inverse Document Frequency

Edited by G Wadley, R Calvo, M Czerwinski, J Torous; submitted 15.02.18; peer-reviewed by Y Wang, J Colditz, K Loveys; comments to author 22.03.18; revised version received 04.04.18; accepted 05.04.18; published 12.06.18

Please cite as:
Kornfield R, Sarma PK, Shah DV, McTavish F, Landucci G, Pe-Romashko K, Gustafson DH
Detecting Recovery Problems Just in Time: Application of Automated Linguistic Analysis and Supervised Machine Learning to an Online Substance Abuse Forum
J Med Internet Res 2018;20(6):e10136
URL: http://www.jmir.org/2018/6/e10136/
doi: 10.2196/10136
PMID: 29895517