Published in final edited form as:

CSCW Conf Comput Support Coop Work. 2017 ; 2017: 1320–1333. doi:10.1145/2998181.2998336.

Assessing Behavioral Stages From Social Media Data

Jason Liu,
University of Waterloo

Elissa R Weitzman, and
Boston Childrens Hospital

Rumi Chunara
New York University

Abstract

Important work rooted in psychological theory posits that health behavior change occurs through a series of discrete stages. Our work builds on the field of social computing by identifying how social media data can be used to resolve behavior stages at high resolution (e.g. hourly/daily) for key population subgroups and times. In essence this approach opens new opportunities to advance psychological theories and better understand how our health is shaped based on the real, dynamic, and rapid actions we make every day. To do so, we bring together domain knowledge and machine learning methods to form a hierarchical classification of Twitter data that resolves different stages of behavior. We identify and examine temporal patterns of the identified stages, with alcohol as a use case (planning or looking to drink, currently drinking, and reflecting on drinking). Known seasonal trends are compared with findings from our methods. We discuss the potential health policy implications of detecting high frequency behavior stages.

Author Keywords

natural language processing; hierarchical classification; behavior; social media; health

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g., HCI); Miscellaneous; H.3.3. Information Storage and Retrieval; Information Storage and Retrieval

INTRODUCTION

It has been recognized that individuals post about a wide variety of health and health-behavior topics on social media data, which has been harnessed for many efforts. Often these data have been shown useful to generate precise and high-resolution temporal health related information. Substantial work in social computing has also focused on predicting behaviors from online communities [20] as well as studying shifts in ideation [16]. Thus, capturing detailed behavioral information is an important theme in social computing; methods from computing can be used to address the often complex, longitudinal and nuanced information that may be difficult to capture via traditional survey mechanisms that may also entail lengthy processing times. To-date, the main approaches that have been used to understand population-level temporal health trends from social media data derive from prediction and ecological correlations. The continuous and passive nature of social media, as well as mature technical methods of natural language and classification for processing short text documents, suggest that further advances with high public health relevance may also be possible, such as observing important characteristics about our behaviors. In particular, social media data can be crucial to identify more real-time behavior patterns relevant to health intervention and promotion efforts. Here we delve into a fundamental aspect of behavioral theory, and assess how harnessing computational approaches and social media data can go beyond correlations and survey-based approaches to provide added insight into how we understand behaviors.

The idea of behavior stages is rooted in psychological theories which have been applied extensively for better understanding our health-related risk factors. These theories including the transtheoretical model of behavior change (TTM) posit that shifts in behavior can be described in series as discrete changes, which are very important for understanding the process of behavior change. Foundations of this theory were developed specifically in the study of addictive behaviors, thus the theory has been designed around stages categorized as: precontemplation, contemplation, preparation, action, maintenance, and termination, each occurring over at least month-long time periods [53].

Recent literature has indicated challenges with this static and linear approach to how we perceive behavior. For example, Povey and others have explicitly raised concern about application of the TTM model to behaviors that may be fundamentally different in nature compared to these original scenarios upon which the model was originally based [51]. Results from Povey's study measuring stages of change with respect to dietary behaviors indicated that people who were actively making or maintaining a change had done so for a range of different periods of time, with no specific cut-off point being evident (opposed to previous categorizations which considered behavior changes as homogeneous, with specific ending points that occur after at least month-long time periods). Also particularly relevant for measures considered here (alcohol use), the same study indicated that answers from self-completion questionnaires used to place people in behavior stages are open to misinterpretation about how well people are aware of their activity behaviors. Other issues in categorization vis-à-vis existing approaches for measuring behavior stages have also been highlighted. Specifically, there is a strong issue of perceptions from survey-based measures; individuals could incorrectly perceive they have made healthy modifications, when in fact they have not. For example, bias in the term exercise which people have been shown to associate with more high-intensity physical activity only [4, 57]. Further, many of the existing studies on the stages of behavior have challenges in terms of representativeness. Given the modes of feasible participant enrollment, existing studies have generally focused on select populations; groups by age, gender, or those in a particular setting (e.g. a specific work-place) are considered, and generally these have been limited to on the order of 100–1000 participants. Thus there is a desire for studies which examine stage distributions in larger samples [41]. In sum, while the theory has been foundational in terms of understanding our behavior and changes, a predominantly linear and static model of behavior severely limits the ability to guide dynamic, adaptive interventions via mobile technologies and there is a need for contextual, dynamic and stratified understanding of behavior [56, 64].

In order to achieve these goals, reviews of the theory have consistently indicated the need for better ways to measure and quantify behavior [64, 56, 42]. To-date, assessment of which stage an individual is in has been assessed using mail, telephone or paper questionnaires, which are robust, yet can introduce self-report bias, can suffer from issues of recall and time commitment [67]. Given this format of data, specific rigid cut-off time periods are assigned to different stages. However, transitions between stages are dynamic, occurring at short and recurring time periods (as opposed to simple linear movement through the stages at defined time periods) highlighting the need for new data sources and metrics [51].

Further motivating the need for contextual, dynamic and stratified understanding of behavior, is the development of computerized and mobile health real-time and just-in-time behavior interventions has expanded rapidly in the last decade in the computer-supported cooperative work and social computing (CSCW) community in particular [30, 63]. In the field of social computing, social media data has also been shown to offer precise time information on health outcomes. These advances in social and wellness work using social media indicate that real-time data may address the challenge of better illuminating dynamic behavior stages and patterns at relevant time periods. Understanding what relevant dynamic stages are is a question that has not yet been examined. Ultimately the more dynamic and non-linear possibilities that new data types offer cannot map to stages defined through survey-based data gathering, however they can still reflect and illustrate important processes behind how we engage in behaviors. Given the understanding that behavior changes are not one-time efforts, new data sources and stage definitions will offer a novel way to demarcate small, every day efforts that represent our behaviors.

Here we examine this idea for the first time and show how the idea can be implemented it to a use case: alcohol-use behavior. In order to address this, in our study we develop a natural language processing pipeline, which is combined with domain knowledge to identify relevant alcohol-use behavior stages at high temporal resolution. We then demonstrate how this derived data can be used to assess temporal trends of real-time behavioral stages, that could not be resolved at the same scale with existing methods, for alcohol-use. We examine and validate results for key sub-groups and at potentially high-risk periods which are relevant for the development of behavioral interventions.

Specific contributions of this paper are:

- Detection of high-resolution behavioral stages from social media data using a machine learning pipeline
- Assessment of population-level temporal trends of high-resolution behavior stages of drinking alcohol as a use case, in the context of known epidemiological trends
- Examination of behavior stage trends stratified by groups and at high-risk time periods

Privacy and Ethics

Given the sensitivities of using personally-generated data, and data that is specifically related to health, we hold privacy and security of the data under utmost importance. We do not identify specific users in our study, and analysis results are de-identified and/or shown in aggregate. Given that we are only working with publicly available data, our work did not qualify for institutional review.

RELATED WORK

Online Data and Public Health Surveillance

Work in the CSCW community has presented the formative recognition that online data can be used to assess behavior change, initially starting with online community data [63]. As well, work in CSCW has pioneered assessment of the stages of change that members of an online addiction recovery group are in [32], as well as examining the relationship between (online) social structure and depression [26]. Other research regarding behaviors in CSCW has been focused on developing computational approaches to promote healthy behavior via content or design [38, 71, 55] and its potential utility for intervention and improvement and in [7] for assessing behavior around major life events. Generally, work in this area has focused on behavior over longer (month-scale) time periods. It is recognized that the Internet provides a candid and emotionally supportive network for communities with socially stigmatized illnesses, e.g., depression and behaviors that are not measured otherwise. By combining domain knowledge and data mining, previous work has identified features related to specific health afflictions and communities [8, 7].

Simultaneously in the public health surveillance field, natural language processing and supervised machine learning have been used with social media data to identify infectious disease events [62]. As well, meta-information from Twitter posts (such as affiliations, network sizes, etc.) alongside human curation have been used to identify subgroups [6, 34], and clustering and forecasting have been used to track high-resolution spatiotemporal trends [13, 23].

While the specific use-case presented here is alcohol use, we present a novel pipeline that brings together the high-resolution and candid nature of social media data to better understand our behaviors. We do this, by for the first time, discerning short term behavior stages. While previous work has assessed behaviors via features such as online activity, emotion and linguistic style, here we use natural language processing to assess stages directly from what people are posting about, in high-temporal resolution.

Alcohol Surveillance

Currently, the National Survey on Drug Use and Health (NS-DUH) provides information on multiple alcohol consumption outcomes in the United States. For example, outcomes include the average number of drinks consumed during a week, and the number of binge drinking episodes in the last 30 days [36]. The NSDUH is sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), an agency in the U.S. Department of Health and Human Services (DHHS) and has been administered at an approximately an annual frequency. While there are varying social norms around alcohol use and individuals may consume alcohol without disease ramifications, it is an important behavior to monitor. Alcohol use is a leading contributor to global patterns of morbidity and mortality.

Previous work using Twitter data for alcohol surveillance has included a feasibility study for classifying Twitter data related to alcohol [1] and identifying locations of alcohol consumption [27]. While these studies have been important to show the potential for social media data to be used in this important problem, the work has focused on demonstrating performance of various classifiers for alcohol-related classification in general and geo-spatial distributions, neither of which have examined stages of alcohol use behavior or transitions through them.

Hierarchical Classification

In order to identify individuals engaging in a specific behavior we develop a hierarchical pipeline. Here we motivate and illustrate how this method is used to ascertain behavior stages from social media data.

A range of statistical and machine learning techniques have been applied to text categorization. Multi-label classification is a supervised learning problem where an instance may be associated with multiple labels. When the labels in a data set belong to a hierarchical structure, then we call the task hierarchical classification. Many real-world information sources have a complex hierarchical structure (e.g. in bioinformatics, image processing, patents, Web pages), thus these methods have received increasing attention. In hierarchical classification, a separate model is used to distinguish a second-level category from other categories of the top-level. This is in contrast to the flat non-hierarchical case, in which a model distinguishes a second-level category from all other second-level categories even of different parent categories.

A common approach to multi-label classification is to perform problem transformation, whereby a multi-label problem is transformed into one or more single-label (i.e. binary, or multi-class) problems. An alternative is to directly make multi-label predictions; such as by using decision trees. The former is most common, and in the case of transforming a multi-label problem into multiple binary problems is called binary relevance. In binary relevance, d binary classification problems are considered, one for each class variable $C_1, \cdots, C_d$, and a classifier is independently learned for each class variable. While this approach has low computational complexity, and existing classification techniques can be applied, there are known challenges. Binary relevance does not directly model correlations that exist between labels in the training data. Due to this loss of information, bias can manifest in the results or reduce overall predictive performance [54].

To overcome limitations of binary relevance, chain classifiers [74] may be used to preserve computational efficiency of the binary relevance approach while incorporating class dependencies. A chain classifier consists of d binary classifiers that are linked in a chain such that each classifier incorporates the class predicted by the previous classifiers as additional attributes. Thus the feature vector for each binary classifier, L_i, is filtered via the labeled data from all previous classifiers in the chain. In other words, each classifier in the chain is trained to learn the association of label l_i given the features augmented with all previous binary predictions in the chain, $l_1 l_2, \cdots, l_{i-1}$. For any document with feature vector $\mathbf{x}$, the overall classification thus predicts: $P(l_i | \mathbf{x}, l_1, l_2, \cdots, l_{i-1})$.

Implementation of a chain classifier consists of the following two steps: 1) obtain dependency structure for the classes and then 2) based on the obtained structure, build a classifier chain. Classifier chains have been receiving greater attention as the approach is an appealing method for tackling the multi-label classification task. To determine the actual elements in the chain (*behavioral stages* in this study), in some cases, the structure in step 1 can be obtained from a previously learned graph structure. In other cases categories are defined manually, or through exploratory cluster analysis, or geographic boundaries [39]. Often in domain-specific questions, domain expertise can be used to isolate a known hierarchical structure, and then the classification problem can be decomposed into a set of smaller problems corresponding to hierarchical splits in the tree. In the web-search domain, hierarchies such as X versus Y classification errors can result for highly related categories (e.g., a page about Furniture/Table might be confused with Furniture/Desk), thus category specific features should improve accuracy. Also in cases of imbalance between data distribution in each class, a hierarchical classification scheme has been preferred over a flat scheme [22]. Accordingly in the work here, we address an imperative social computing issue: balancing computational automation and subject expertise by working with domain experts to ascertain the appropriate classification hierarchy that we then implement with machine learning.

There have been uses of hierarchical structure for classifying large, heterogeneous collections of web content. The earliest work was focused on hierarchical classification of web content. This was appropriate given the highly structured, large amounts of text, spanning many categories such as 135 topical categories on one page. A full review of hierarchies in web pages has been explored [19]. Many classification approaches in social media have ignored hierarchical structure, treating each category or class separately, and in effect flattening the structure. For example, for using Twitter data in the health-domain, isolating posts that describe an individuals circumstances or experiences (first-person), are of particular importance. While there have been approaches to isolating these first-person posts, including Part of Speech Templates [29], as we are aware, hierarchical classification has to-date not been applied towards this aim. In sum, the social media medium provides a natural opportunity to garner behavior stage information and hierarchical classification offers an apt approach by which to achieve this goal.

DATA AND APPROACH

The idea behind our approach is to, from any set of Twitter data, identify posts that illustrate relevant stages of engaging in alcohol use. We develop a hierarchical classification model for social media data and evaluate the results in comparison to patterns of alcohol use derived from survey measures. This hierarchical separation increases the efficiency of our classification effort, as features for each level can be learned separately. We pinpoint also for whom, and when these patterns are manifest.

Data

Rich streams of dynamic data from personally generated digital sources now enable unprecedented opportunity to understand health for sizeable populations globally. Although the real-time nature of social media has been harnessed for event-based infectious disease surveillance, to-date these daily digital footprints have not been directly used to monitor behaviors and their shifts. Given its scale and reach amongst significant populations in the United States, and its tight coupling with our daily lives (wherein people discuss their every-day happenings), Twitter is an appropriate data source from which to examine alcohol-related behavior. Finally, Twitter is the largest social media platform with a freely available API.

Two sets of data were used in this study. The first was a month of data collected in June 2015 via a real-time stream from the Twitter API. In sum there were 804,000 Tweets total, from 228,405 users. We used data from the API to demonstrate detection of behavior stages from a publicly available source and the entire data stream. For identification of finer-resolution trends, we used a larger set of data which was obtained via GNIP (http://www.gnip.com), which provides access to the full Twitter Firehose. As the amount of data decreases at each level of the classification, we used this higher volume source in order to examine behavior stages for further validation during a specific event New Years Eve and New Years holiday which occurs over a narrow-time window further limiting the amount of related data. This data set included 4,839,870 Tweets and spanned from December 30, 2015 to January 3, 2016. In all cases we limited data collection to English language Tweets from USA.

Stage Determination

Previous work using Twitter in disease surveillance has demonstrated that simple content analysis via keywords can conflate Tweets that report self-infection (or a behavior of the self, in this case), with those that express awareness about a topic. In particular for alcohol, we qualitatively observed that many alcohol-related posts are related to alcohol advertising opposed to alcohol use. Thus here, part of our hierarchy was geared towards distinguishing someone talking about their own alcohol behavior (hereafter referred to as First person) versus other alcohol-related posts such as advertisements. In order to determine the subsequent stages in our hierarchy, we collaborated with domain experts. Alongside these experts in behavior we performed qualitative analysis of first-person alcohol-related Twitter data, and we identified temporal stages that are relevant to alcohol consumption. Resulting stages were determined to be (with example Tweets):

- **Looking**: Planning or looking to drink

  "who wanna get drunk tonight?"

  "I need a drink baad"

- **Current**: Currently engaging in drinking

  "out drinking with the girls!"

  "Drinking a beer with dad"

- **Reflecting**: Reflecting on alcohol consumption

  "last night was crazy, never drinking again XD"

  "so hungover…"

These stages, for the first time, represent ascertainment of real-time behavioral stages. The stages align with the premise of existing behavioral theories, in that they are relevant to behavioral choices which have a natural temporal ordering. As well from a health surveillance perspective, they enable an understanding of behavioral stages and choices at a higher time resolution than previous behavior stage theories could uncover.

Model

Our resulting hierarchical classification scheme is illustrated in Figure 1. We assume that at any given time, each individual Tweet belongs to one of a set of mutually exclusive and exhaustive latent stage classes.

HIERARCHICAL CLASSIFICATION

Active Learning and Labeling Data

In order to label data for training classifiers in the pipeline, Tweets were labeled using Amazon Mechanical Turk (AMT). Labels were obtained for each classifier in the pipeline through a staged mechanism. We used an active learning approach to increase the efficiency for labeling our data, as on average only 2% of a random sample of tweets would be alcohol related. This means that rather than labeling all Tweets for each classifier all at once, which could be very costly for getting positive examples, we requested data for labels in an iterative manner, in order to target labeling at data that could specifically increase the future performance of our model. To do so, we first train a support vector classifier (described below) to all available labeled data and then select all unlabeled Tweets that were close to the decision boundary, essentially points where the $P(l_i j\mathbf{x}) \approx 0:5$, and submit those tweets to Amazon Mechanical Turk for classification. This also prevents us from requesting labels for Tweets that were obviously irrelevant to classify. Table 1 shows the distribution of the labeled data we collected.

For the labeling procedure, we asked AMT workers to first classify Tweets as alcohol related; if the Tweet was determined to be alcohol related, the worker was asked to classify if the same Tweet was First person; if the Tweet was categorized as alcohol related and first person, the worker was asked to label it according to the behavior stage categories: looking, current, or reflecting. All instructions are as follows:

- **Alcohol**: The text is related to alcohol but not necessarily indicative of consumption, such as related to a beer promotion, bar/club event or drunk driving related.
- **First Person**: This is for text that discusses the writer themselves is consuming alcohol.

- **Behavior Stages**: These are labels describing which stage of drinking alcohol the text represents: looking to drink alcohol, currently drinking (current), or reflecting on drinking

While Tweets could potentially exhibit evidence of multiple stages (e.g. both looking and current stages), we found that only a small percentage of tweets did, so each stage was formulated as a binary task for simplicity. For quality control we also removed any labels which were generated in significantly lower time than average following established protocol [18].

Feature Selection

At each classification stage in the pipeline, we used a randomized grid search in order to find optimial hyperparameters for each prior distribution. The grid search allows us to identify parameters values sampled from distributions rather than using specific values that may not contribute to model performance [5]. We considered hyperparameters for different normalization schemes for vectorization, the range of n-grams, along with regularization constants and tuning parameters specific to our model. From our data set, we found that the best features were derived primarily by using term frequency-inverse document frequency (tf-idf) features extracted from the words of each Tweet collected into phrases. We found that using user features (e.g. number of friends, followers, etc) did not significantly increase the performance of any algorithm.

A 5-fold stratified cross validation was used at every stage of classification to evaluate our pipeline during the randomized grid search. Tokenization was implemented using a Twitter-specific tokenizer [43]. This tokenizer allows us to separate hashtags, Retweet tokens, URLs, and emoticons and tokenize them correctly. Identified optimal parameters for the pipeline involved using the top 200,000 tokens generated by taking (1,3)-grams of word tokens.

Classifier Details

We implemented the classifiers using the scikit-learn package in Python [48] and a randomized grid search was also used to optimize all hyperparameters of the classifiers at each stage of the hierarchy. For classification at each stage we considered a logistic regression (LR), support vector machines (SVM) and random forest (RF) approach. Parameter space for the regularization parameter was a uniform distribution across $(10^{-4}, 10^{4})$. For each of the classifiers (alcohol, first person, and behavioral stages), after completing a hyperparameter grid search on each approach we decided to use logistic regression (LR) with L2 regularization. Regularization coefficient was $C = 0.01$. Logistic regression is a well-established classification algorithm which is appropriate if the data can be assumed to be linearly separable, efficient to run and output is easily interpretable as a probability [35], all applicable here. To balance efficiency and performance, we found that a good limit was to use the top 200,000 features. Performance of the classifiers are recorded in Table 3. Table 2 illustrates the top five features for each hierarchy stage taken from the best performing L2 regularized logistic regression model, along with the feature strength (β). Most of the features are intuitive of the classification results (e.g. “tonight”, “wanna” for the looking category and “was”, “when” for reflecting). While “got drunk last” had a negative weight for the reflecting category, it is possible this feature is generally not used in a context truly reflective of first-person alcohol consumption.

RESULTS

Here we provide, first, a qualitative description of our temporal findings, and then summarize key metrics from the results that demonstrate the novel social computing of this approach.

Behavior Stages Over Time

In each of Figures 2, 3 and 6, the time series component is the normalized count of Twitter posts (posts in the specified category divided by all posts even prior to alcohol classification). Normalizing by the total number of posts provides a point prevalence estimate, which is a metric comparable to how latent variables are represented in behavioral science. The final measure is thus computed using a multiplicative decision rule giving the probability of a post, X, being classified in any behavioral stage as:

$$P(l_{stage})=P(l_{alc}|X)P(l_{fp}|X,l_{alc})P(l_{stage}|X,l_{alc},l_{fp}) \quad (1)$$

Where $P(l_{alc})$ is the probability of a label of alcohol-related, $P(l_{fp})$ the probability of a first-person label, and $P(l_{stage})$ the probability of a given behavioral stage label. Shaded areas represent the 95% confidence interval.

Descriptively, we find that current drinking increases over days in the week, peaking on the weekend, with the highest proportion of current drinking out of all Tweets on Saturday. Dissecting this trend by each of the behavioral stages (in Figure 2), and examining on a daily scale, current and looking to drink stages peak on Saturday, while reflecting peaks later, on Sunday. As the baseline proportion of Tweets in each stage category vary (overall current has the highest proportion), in order to understand the temporal significance of each stage, we computed which day of week has highest probability for each stage: $P(day|l_{stage})$, which is summarized in Table 4. The table shows how drinking behavior changes over the course of the week, with the looking to drink stage happening with highest probability during the week starting on Tuesday, currently drinking stage happening with highest probability on the weekend, followed by the reflecting stage after the weekend Monday through Wednesday.

We also examine how overall drinking patterns on Twitter differ between weekdays and weekends. Overall, people are talking about alcohol most after midnight (even though Currently drinking peaks in the evening — an important distinction). On weekends the total amount of alcohol conversation is larger, and thus the sum of all categories peaks earlier compared to weekdays (Figure 3).

The real-time nature of social media data enables us to look at detailed temporal trends over the course of the week as well. In Figure 4 we see that the proportion of current drinking stage posts peaks generally at about 8pm each day of the week, with another peak in the early morning (when overall Twitter posts are lowest). The proportion of Looking to drink posts starts to increase in the evening each day, until approximately midnight. Finally, reflecting is overall the lowest proportion of posts, but overall peaks in the early morning especially Friday through Saturday.

Validation of Trends

Given the importance of monitoring temporal trends of alcohol use, there have been some research studies that have explicitly surveyed people about their drinking behavior over a typical week or day time-period [2, 11]. While these studies have included limited population sizes, these are the best ground-truth that can be compared to the high-resolution findings from social media.

High-risk Time Periods—We examined temporal trends of alcohol behavioral stages just prior to and after the New Years holiday. New Years is one of the most universal holidays and is known to present an important period for alcohol surveillance. We found clear patterns of Looking to drink in the earlier hours of new year's eve, and peaked levels of Current alcohol consumption in the hours before midnight on New Years Eve and again early in the morning of January 1. These patterns occurred during the week, deviating from average weekly trends (Figure 6). Absolute values of the social media proportions were set by calibration; we aligned our data to align with previous work ascertaining alcohol Tweets at 2.1% of all posts (in a descriptive manner, with no disaggregation by first-person) [1]. Following this, while absolute proportion values are set by calibration, it is the transition between stages (highlighted in green, red and blue in Figure 6) that interestingly indicate an anomaly or event is occurring.

These patterns illustrate a conservative investigation: the New Year trends we report are situated within a holiday period in which the baseline for consumption may be high. Many workplaces are closed, and there is a high density of social events with alcohol during and following Christmas. Viewed in this fashion, findings of temporal peaks around New Years are especially robust.

Findings by Time—The high-resolution views into social phenomena that are enabled by social media, are not resolvable when relying on data from other mediums. Surveys are robust and can garner structured information, however resource limitations, recall and information biases impede collection high-resolution data comparable to what is possible through social media. Specifically in relation to alcohol-related behaviors, patterns of consumption are important to resolve to better understand patterns of behavior at specific time-scales and discern temporal relations with other phenomena such as traffic fatalities and homicide, in which alcohol use is known to play a significant role.

Drawing inference about temporal patterning has been continually challenged by insufficient and homogeneous samples from surveys. Further, social drinking problems might be viewed as deviations from normative behavior, and for this reason, greater understanding of behavior during anomalous times such as holidays is important. Survey research is challenged by this level of granularity and recall, although incidence rates of problems (emergency department visits, traffic accidents) have indicated that there are changes in alcohol behaviors at these times (as measured through clinical intoxication data) [31]. Some historical studies using survey methodology have uncovered general patterns at the level of days. For example, one group surveyed 500 adults of the City of Boston regarding their temporal drinking patterns [2]. This work showed that Tuesday represents the day of the week with the lowest frequency of alcohol-related traffic fatalities, while Saturday has the highest [2]. More recent findings support these temporal trends as well [21]. Alcohol consumption patterns over the course of a single day are also important to consider for a host of bio-psychosocial reasons. Evening alcohol consumption can affect glucose control in diabetics [68]; daytime consumption can interfere with work performance [33]; driving and accidents are associated with morning consumption [49]. Understanding patterns of alcohol use at the hourly level can also be affected by alcohol dependence [65] and school class schedules [73]. In a survey of 16,086 US adults, the times at which lowest drinking rates were reported were between midnight and 6 am [11].

It is not possible to precisely compare data from these two sources (social media and survey data) due to differences in the included populations. However, overall trends align; Tuesday is the lowest consumption and alcohol-posting day and Saturday the highest [2]. Over the course of a day, we also see that all first-person alcohol posts decrease in the morning, after which they start to rise (Figure 3). Weekdays on average have a lower proportion of Tweets that are first-person alcohol compared with weekends. Weekend posts also continue through the night moreso and slightly later than weekdays (Figure 2). These results correspond to general expected alcohol consumption trends; for example the times at which lowest drinking rates were reported in a survey of 16,086 US adults, were between midnight and 6 am [11].

Transitions by Group—It is always important to recognize, especially in epidemiological modeling, that inference to the entire population should only be made once the population that the data represents is understood. For example, more Black non-Hispanic and Hispanic people who are online use Twitter than White non-Hispanics, and representation from those over 65 and the very young is low [23]. Identifying specific populations the data represents is thus useful for better understanding of findings. As well, subgroups are also of importance for surveillance; some may lend themselves better to certain mediums (age, college students, etc.). Regarding alcohol, for example, there have been noted differences in consumption between genders [72]. From Twitter data there has been a lot of work to infer latent features of users; for example by gender and age [70]. For simplicity here, we used a lookup table that has demonstrated good performance for classification of names. The table has more than 40,000 first names linked with most common associated gender. Names have been classified by native speakers and covers the vast majority of names in USA amongst other countries (the lookup table is available at: https://github.com/cstuder/genderReader/blob/master/gender.c/nam_dict.txt which is released under the GNU Free Documentation License).

Tweets were classified as from users that are male, female or other. The other category included users that were organizations, bots, or people with unconventional names; they were not included in our gender-stratified analysis. This analysis resulted in Tweets from 404,191 others, 222,253 male and 176,425 female. The hierarchical classification results were then examined separately for male and female groups (Figure 5) and results are discussed below.

In Figure 5 we examined trends by gender. In this figure, for each of men and women, the time series component is the normalized count of Twitter posts in the current drinking category divided by all first person alcohol posts for that gender. In other words, this figure shows the proportion of currently drinking tweets relative to first person tweets by each gender. Normalizing by the total number of first person posts here provides a point prevalence estimate which accounts for the different overall posting trends by different groups. The final measure at each day for each gender g is thus computed as: $P(l_{curr}|l_{alc}, l_{fp}, l_g, X)$ where $P(l_g)$ is the probability of a label of gender g (computed via the lookup table as described above).

Overall, while the time series over the course of the week was similar for male and female groups, we found that proportion of currently drinking Tweets by men were generally higher than the proportion for women. Again, while we can not exactly compare the survey and social media results due to population differences, not exactly comparable the trends generally align with gender differences in alcohol consumption that have been uncovered via surveys. In these, it has been found that women and men differ little in the probability of currently drinking versus abstaining, but men consistently exceed women in typical drinking quantities and rates [72].

Error

There are common error propagation issues in chained classifiers that have been recognized, including that chained classifiers can break the machine learning assumption that test data is representative (identically distributed) with training data [60]. This issue has been categorized as attribute noise in studies of real-world data [75]. Several factors have been identified to contribute to this error: 1) length of the chain, 2) order of the chain, 3) dependency among labels and 4) accuracy of the classifiers

While a theoretical examination of errors in classifier chains is beyond the scope of this work, here we describe how we have attempted to reduce and understand the errors in our results. Factors regarding the selected labels in the chain, their ordering and chain length have been addressed by first reducing the number of stages in the hierarchy to those that are most important by working with domain experts. Error from accuracy of the classifiers can be identified by computing the marginal from the product in Equation 1.

Thus, in regards to error in determination of the behavior stages and their ordering, our approach naturally results in the most appropriate ordering of classifier levels. As well, dependencies (links between labels) were set based on discussion with the domain experts, and in practice this results in more discriminating topics higher in the hierarchy, such that subsequent levels become more specialized.

DISCUSSION

Summary of Findings

Our work illuminates a high-resolution view into behavior stages over time that has simply not been possible to assess previously. We first ascertained what these novel stages at high-resolution would be, then resolved them from social media data, assessed trends from these new stages in comparison to existing epidemiological knowledge and finally examined these trends stratifying by relevant subgroups and at relevant times. Through this process we demonstrate some behavioral trends related to alcohol use at the hourly and weekly level that are known but have not previously been shown quantitatively at scale. While the trends observed link to known behavior patterns of alcohol use, the approach presented here, discerning high-resolution behavioral stages from social media data, has the potential to be used for novel applications such as uncovering unknown anomalies in behavior and linking changes in behavior to specific times and places.

Addition to Existing Work

This work adds to previous research in a number of ways. First, in general for computational public health temporal surveillance efforts, this work goes beyond ecological correlations and starts to assess real-time measures of our behaviors, and how we can learn how these diverge for different subgroups and at key times. Practically, temporal patterns are important for public health departments to assess in order to anticipate and deploy focused interventions. Methodologically, our work also describes an novel approach for resolving nuanced personal features from social media data (via hierarchical classification) which to-date we have not seen implemented for social media data.

Overall, this work builds on research in the social computing community. The community has recognized that social networking or social media data can be used to assess behavior change, and here we show this at higher-resolution, relevant to time periods at which we all make behavior choices and changes. The approach demonstrated here is also shows how empirical data can be garnered to inform research in the development and design of systems to instill behavior change and real-time interventions [63]. A predominantly linear and static model of behavior that is currently available, severely limits the ability to guide dynamic, adaptive interventions via mobile technologies [56].

Implications for Behavior Surveillance and Change

The approach designed here can enable prospective monitoring of events that can potentially affect behavioral patterns. Indeed, besides monitoring events that have a known or expected effect (such as the New Year’s holiday), by prospectively monitoring data in this detailed manner, novel unanticipated events can be identified through transitions in behavioral stages (as in Figure 6). Public health departments could use this approach to recognize unexpected events that have changed behavior, and for which they could then follow up on with qualitative analysis to better understand mechanisms and causes. As well with the precise knowledge crystallized here, interventions to targeted at most vulnerable populations and times (such as groups that are most likely to actually engage in a behavior, or at times when most people decide to start or stop) can be better designed. As well, these novel behavior stages can be used to design real-time online and off-line tools and interventions. This is of particular importance for behavioral surveillance because most behaviors are modifiable. The active learning approach applied here to determine what kinds of posts are related to a particular behavior can also assist with recognition of new events. As data from an unexpected event surfaces on social media, it would get labeled and improve classifier performance in recognizing intricacies of the novel occurrence. Thus the pipeline can learn features that might not be seen otherwise. Classification time for the data was very reasonable (all analyses performed on a MacBook) suggesting the method could be used with efficiency in real-time and with larger data sets for dynamic public health interventions.

The machine learning pipeline and hierarchical classification applied here, while demonstrated for one particular type of behavior, can be readily applied to other target behaviors based on the modular nature of the approach. For example, we would only have to retrain the parent (alcohol) node of our classifier rather than the whole model, for simple expansion to surveillance of other important behaviors. Or, retrain the child (behavior stage) nodes as the number or types of stages change.

The types of behaviors this work would be amenable to extending to can be informed by historical work in behavioral sciences. Theories of behavior change, as described above, have been used to describe the different phases involved in the acquisition and maintenance of addictive behaviors (e.g. drinking alcohol, and smoking) and generally on a long (month-scale) timescale of progressing through said stages. The TTM has started to be applied to other behaviors such as physical activity, however based on the data types available, relevant dynamic stages of these behaviors that represent relevant behavior changes over daily and shorter timescales have not been captured. These shorter timescales are important, as we each make choices every day regarding which healthy behaviors to choose or not. Using passively sourced data such as from social media can offer this high-resolution. As well this type of data also offers the opportunity to learn about behaviors which otherwise may not be reported for various reasons (e.g. recall difficulties, willingness to disclose in survey situations, etc.). Thus, beyond topics such as alcohol use, other relevant dynamic and sensitive behaviors that could be investigated with this approach include tobacco and other drug use, as well as physical activity.

Limitations and Future Work

We acknowledge limitations of our research. First, our findings are limited to the population who uses Twitter. However, our work in stratifying the data by specific groups (e.g. gender) can be extended alongside other inference methods for assessing characteristics of the Twitter population such as age, to better understand the reference population. Our data is also limited by access to public data only. We cannot access private posts. This makes it difficult to perform a systematic study of any differences in disclosure on public versus private data, however the sheer size of the data considered and further dissection to understand the reference population can help address this issue. The performance of classification implemented here can always be improved with added labelled data (especially in the later stages of our classification hierarchy, wherein there is less data). In general we find that F1 and AUC scores reported in Table 3 are reasonable (with better performance in the earlier classifiers for which there was more data), and given that findings described in this study correspond to expected domain knowledge, we did not further pursue improved classification accuracy for this work.

In the future, geo-spatial information which is available via Twitter, for example at the state or city level (either directly or through inference), can be incorporated in order to assess how alcohol is being consumed in specific regions. Methods used here can also be applied to detection of anomalous behaviors. Although we knew to examine behavior stages at New Years, there may be other time periods that prospectively can be identified and further investigated for reasons behind unexpected behavioral shifts. Finally, with enough data, user-level trends can be evaluated for personalized risk assessment. There are also many practical implications for this work. In the future we envision that this approach can be used to create a modular framework that can be applied to a variety of health surveillance topics and inform priority setting by national and state governments, by detecting trends at a timescale and with a scope that are otherwise simply not feasible. In terms of use, this work can also be combined with work that is used to assess triggers and mediators to identify potential change moments [30]. As well, given enough data, mature methods such as probabilistic graphical models can be applied to make more sophisticated predictions from social media data at individual and population-levels. Given that this work shows that we can resolve relevant behavioral stages from passively mined social media data. Important next steps would be to start assessing transitions between stages. This could be done at an individual or population level, both with important preventative implications. Challenges for achieving this involve developing criteria to map an individual and not post to a stage (at the individual level) and garnering enough aggregate (at the population level).

While we demonstrate specific temporal trends and those stratified by group (gender) here, we must also take care to best define exactly what these trends are showing and be aware that there can be confounding variables unaccounted for. Specifically, the trends illustrated in Figures 2–6 are showing social media posting behavior regarding alcohol. We do not link or validate if the posts also represent drinking behavior off-line. Further, there can be confounding variables (beyond overall posting behavior in Twitter, by which we already normalize the data) that affect the displayed trends. For example, while we account for overall posting trends on Twitter by gender (Figure 5), women and men could post about alcohol differently for other reasons that we can not measure; unmeasured confounders are a consistent issue even in survey-based research. Models of behavior have been foundational in public health, making it possible to identify and reach groups of at-risk individuals. Doing so helps us address the core question of any personalized or precision medicine effort; if there are subgroups of patients with a given disorder or disease who respond differentially to the available treatments.

CONCLUSION

In this paper, we offer the first approach to assess behavior stages from social media data, with alcohol as a use case. The nature of social media, its real-time availability and close coupling to our individual experiences, has been shown to have potential for predictions and correlation relationships with important public health topics. However, parsing the data to identify specific and relevant behaviors is non-trivial. To this end, we generated a hierarchical classification scheme, which combines natural language processing, machine learning and domain-specific knowledge for the first time, to articulate first-person behavior stages from Twitter data. The method is efficient through multiple means: hierarchical classification on its own decreases the amount of labels needed compared to flat classification. Further, we use an active learning approach to train the most relevant data. We develop this pipeline using alcohol engagement as a use case. This enables us to articulate different stages of alcohol behavior (looking, current and reflecting). We then examine these stages for hourly and daily time periods. We observe that the highest probability day for different stages shifts from looking to drink Tuesday through Friday, to currently drinking on the weekend, to reflecting about drinking Monday through Wednesday. We can observe known differences in gender alcohol-use, and see behavior changes at a high-risk holiday time period. We assess the resulting patterns against available survey data at the same time resolution and existing knowledge of alcohol behavior trends. In general the modular nature of the pipeline affords us, in the future, the ability to branch out into different domains, and easily apply this approach to different surveillance tasks. Given that unhealthy behaviors adversely impact quality of life for populations worldwide, not just observing their incidence, but also the stages of these behaviors, especially at high resolution can enable their close monitoring and potentially be useful to target effective and real-time interventions.

Acknowledgments

This work was supported by grant R21 AA023901-01 from the National Institutes of Alcohol Abuse and Alcoholism at the National Institutes of Health.

References

1. Aphinyanaphongs, Yin, Ray, Bisakha, Statnikov, Alexander, Krebs, Paul. Text classification for automatic detection of alcohol use-related tweets: A feasibility study. Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on; IEEE; 2014. p. 93-97.
2. Argeriou, Milton. Daily alcohol consumption patterns in Boston: some findings and a partial test of the Tuesday hypothesis. Journal of studies on alcohol. 1975; 36(11):1578–1583. [PubMed: 556]
3. Balani, Sairam, De Choudhury, Munmun. Detecting and characterizing mental health related self-disclosure in social media. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems; ACM; 2015. p. 1373-1378.
4. Baranowski, Tom. Validity and reliability of self report measures of physical activity: an information-processing perspective. Research Quarterly for Exercise and Sport. 1988; 59(4):314–327. (1988).
5. Bergstra, James, Bengio, Yoshua. Random search for hyperparameter optimization. The Journal of Machine Learning Research. 2012; 13(1):281–305. (2012).
6. Cavazos-Rehg, Patricia A., Krauss, Melissa, Fisher, Sherri L., Salyer, Patricia, Grucza, Richard A., Bierut, Laura Jean. Twitter Chatter About Marijuana. Journal of Adolescent Health. 2015; 56(2): 139–145. (2015). [PubMed: 25620299]
7. Chancellor, Stevie, Lin, Zhiyuan Jerry, Goodman, Erica L., Zerwas, Stephanie, De Choudhury, Munmun. Quantifying and Predicting Mental Illness Severity in Online Pro-Eating Disorder Communities. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; 2016a.
8. Chancellor, Stevie, Pater, Jessica Annette, Clear, Trustin, Gilbert, Eric, De Choudhury, Munmun. # thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; ACM; 2016b. p. 1201-1213.
9. Coppersmith, Glen, Dredze, Mark, Harman, Craig. Quantifying mental health signals in Twitter. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; 2014. p. 51-60.
10. Coughlin, Steven S. Recall bias in epidemiologic studies. Journal of clinical epidemiology. 1990; 43(1):87–91. (1990). [PubMed: 2319285]

11. Dawson, Deborah A. Temporal drinking patterns and variation in social consequences. Addiction. 1996; 91(11):1623–1635. (1996). [PubMed: 8972921]
12. De Choudhury, Munmun, Counts, Scott, Horvitz, Eric. Major life changes and behavioral markers in social media: case of childbirth. Proceedings of the 2013 conference on Computer supported cooperative work; ACM; 2013a. p. 1431-1442.
13. De Choudhury, Munmun, Counts, Scott, Horvitz, Eric. Predicting postpartum changes in emotion and behavior via social media. Proceedings of the 2013 ACM annual conference on Human factors in computing systems; ACM; 2013b. p. 3267-3276.
14. De Choudhury, Munmun, Counts, Scott, Horvitz, Eric J., Hoff, Aaron. Characterizing and predicting postpartum depression from shared facebook data. Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing; ACM; 2014. p. 626-638.
15. De Choudhury, Munmun, Gamon, Michael, Counts, Scott, Horvitz, Eric. Predicting Depression via Social Media. ICWSM; 2013.
16. De Choudhury, Munmun, Kiciman, Emre, Dredze, Mark, Coppersmith, Glen, Kumar, Mrinal. Discovering shifts to suicidal ideation from mental health content in social media. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; ACM; 2016a. p. 2098-2110.
17. De Choudhury, Munmun, Sharma, Sanket, Kiciman, Emre. Characterizing dietary choices, nutrition, and language in food deserts via social media. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; ACM; 2016b. p. 1157-1170.
18. Diakopoulos, Nicholas A., Shamma, David A. Characterizing debate performance via aggregated twitter sentiment. In. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; ACM; 2010. p. 1195-1198.
19. Dumais, Susan, Chen, Hao. Hierarchical classification of Web content. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval; ACM; 2000. p. 256-263.
20. Fast, Ethan, McGrath, William, Rajpurkar, Pranav, Bernstein, Michael. Augur: Mining Human Behaviors from Fiction to Power Interactive Systems. 2016 arXiv preprint arXiv:1602.06977 (2016).
21. Foster, Simon, Gmel, Gerhard, Estvez, Natalia, Bhler, Caroline, Mohler-Kuo, Meichun. Temporal patterns of alcohol consumption and alcohol-related road accidents in young Swiss men: seasonal, weekday and public holiday effects. Alcohol and alcoholism. 50(5):565–572. (????).
22. Ghazi, Diman, Inkpen, Diana, Szpakowicz, Stan. Hierarchical versus flat classification of emotions in text. Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text; Association for Computational Linguistics; 2010. p. 140-146.
23. Godin, Frderic, Zuallaert, Jasper, Vandersmissen, Baptist, De Neve, Wesley, Van de Walle, Rik. Beating the Bookmakers: Leveraging Statistics and Twitter Microposts for Predicting Soccer Results. KDD Workshop on Large-Scale Sports Analytics; 2014.
24. Goodman, Leo A. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika. 1974; 61(2):215–231. (1974).
25. Harford, Thomas C., Gerstel, Eva K. Age-related patterns of daily alcohol consumption in metropolitan Boston. Journal of studies on alcohol. 1981 (1981).
26. Homan, Christopher M., Lu, Naiji, Tu, Xin, Lytle, Megan C., Silenzio, Vincent. Social structure and depression in TrevorSpace. Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing; ACM; 2014. p. 615-625.
27. Hossain, Nabil, Hu, Tianran, Feizi, Roghayeh, Zheng, Danning, White, Ann Marie, Luo, Jiebo, Kautz, Henry. Precise Localization of Homes and Activities: Detecting Drinking-While-Tweeting Patterns in Communities. Tenth International AAAI Conference on Web and Social Media; 2016.
28. Joachims, Thorsten. Text categorization with support vector machines: Learning with many relevant features. Springer; 1998.
29. Lamb, Alex, Paul, Michael J., Dredze, Mark. Separating Fact from Fear: Tracking Flu Infections on Twitter. HLT-NAACL. 2013:789–795.

30. Lee, Yeoreum, Lim, Youn-kyung. Understanding the Roles and Influences of Mediators from Multiple Social Channels for Health Behavior Change. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; ACM; 2015. p. 1070-1079.
31. Levine, Harry Gene. What is an alcohol-related problem?(Or, what are people talking about when they refer to alcohol problems?). Journal of Drug issues. 1984; 14(1):45–60. (1984).
32. MacLean, Diana, Gupta, Sonal, Lembke, Anna, Manning, Christopher, Heer, Jeffrey. Forum77: An analysis of an online health forum dedicated to addiction recovery. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; ACM; 2015. p. 1511-1526.
33. Mangione, Thomas W., Howland, Jonathan, Amick, Benjamin, Cote, Jennifer, Lee, Marianne, Bell, Nicole, Levine, Sol. Employee drinking practices and work performance. Journal of studies on alcohol. 1999; 60(2):261–270. (1999). [PubMed: 10091965]
34. McIver, David J., Hawkins, Jared B., Chunara, Rumi, Chatterjee, Arnaub K., Bhandari, Aman, Fitzgerald, Timothy P., Jain, Sachin H., Brownstein, John S. Characterizing sleep issues using Twitter. Journal of medical Internet research. 2015; 17(6) (2015).
35. Menard, Scott. Applied logistic regression analysis. Vol. 106. Sage; 2002.
36. (SAMHSA). Mental Health Services Administration. Substance Abuse and Mental Health Services Administration (2014). Rockville, MD: Substance Abuse and Mental Health Services Administration; 2014. Results from the 2012 national survey on drug use and health: Summary of national findings. 2013. NSDUH Series H-46, HHS Publication No.(SMA) 134795
37. Mhurchu, Cliona Ni, Margetts, Barrie, Speller, Viv. Applying the stages-of-change model to dietary change. Nutrition reviews. 1997; 55(1):10–16. (1997). [PubMed: 9155212]
38. Munson, Sean, Lauterbach, Debra, Newman, M., Resnick, Paul, Morris, M. The Opportunity of Social Computing to Make Us Healthier Together. In. Workshop on CSCW Research in Healthcare: Past, Present, and Future. CSCW; 2010.
39. Muntaner, Carles, Chung, Haejoo, Benach, Joan, Ng, Edwin. Hierarchical cluster analysis of labour market regulations and population health: a taxonomy of low-and middle-income countries. BMC Public Health. 2012; 12(1):1. (2012). [PubMed: 22214479]
40. Murnane, Elizabeth L., Counts, Scott. Unraveling abstinence and relapse: smoking cessation reflected in social media. Proceedings of the 32nd annual ACM conference on Human factors in computing systems; ACM; 2014. p. 1345-1354.
41. Nigg, Claudio R., Burbank, Patricia M., Padula, Cindy, Dufresne, Robert, Rossi, Joseph S., Velicer, Wayne F., Laforge, Robert G., Prochaska, James O. Stages of change across ten health risk behaviors for older adults. The Gerontologist. 1999; 39(4):473–482. (1999). [PubMed: 10495586]
42. Nilsen, Wendy J., Pavel, Misha. Moving behavioral theories into the 21st century: technological advancements for improving quality of life. IEEE pulse. 2013; 4(5):25–28. (2013). [PubMed: 24056790]
43. O'Connor, Brendan, Krieger, Michel, Ahn, David. TweetMotif: Exploratory Search and Topic Summarization for Twitter. ICWSM. 2010
44. Owoputi, Olutobi, O'Connor, Brendan, Dyer, Chris, Gimpel, Kevin, Schneider, Nathan, Smith, Noah A. Improved part-of-speech tagging for online conversational text with word clusters. Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings; Association for Computational Linguistics; 2013.
45. OReilly, Gillian A., Spruijt-Metz, Donna. Current mHealth technologies for physical activity assessment and promotion. American journal of preventive medicine. 2013; 45(4):501–507. (2013). [PubMed: 24050427]
46. Paul, Michael J., Dredze, Mark. A model for mining public health topics from Twitter. Health. 2012; 11:16–6. (2012).
47. Pedregosa, Fabian, Varoquaux, Gal, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research. 2011a; 12:2825–2830. (2011).
48. Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, et al. Scikit-

learn: Machine learning in Python. The Journal of Machine Learning Research. 2011b; 12:2825–2830. (2011).

49. Philip P, Vervialle F, Le Breton P, Taillard J, Horne James A. Fatigue, alcohol, and serious road crashes in France: factorial study of national data. Bmj. 2001; 322(7290):829–830. (2001). [PubMed: 11290636]
50. Ploderer, Bernd, Smith, Wally, Pearce, Jon, Borland, Ron. An Object-Centred Approach to Encourage Online Participation in the Context of Behaviour Change. Computer Supported Cooperative Work (CSCW). 2015; 24(1):39–64. (2015).
51. Povey R, Conner M, Sparks P, James R, Shepherd R. A critical examination of the application of the Transtheoretical Model's stages of change to dietary behaviours. Health Education Research. 14(5):641–651. (????).
52. Povey, Rachel, Conner, Mark, Sparks, Paul, James, Rhiannon, Shepherd, Richard. A critical examination of the application of the Transtheoretical Model's stages of change to dietary behaviours. Health Education Research. 1999; 14(5):641–651. (1999). [PubMed: 10510072]
53. Prochaska, James O., Velicer, Wayne F. The transtheoretical model of health behavior change. American journal of health promotion. 1997; 12(1):38–48. (1997). [PubMed: 10170434]
54. Read, Jesse, Pfahringer, Bernhard, Holmes, Geoff, Frank, Eibe. Classifier chains for multi-label classification. Machine learning. 2011; 85(3):333–359. (2011).
55. Riche, Yann, Mackay, Wendy. PeerCare: supporting awareness of rhythms and routines for better aging in place. Computer Supported Cooperative Work (CSCW). 2010; 19(1):73–104. (2010).
56. Riley, William T., Rivera, Daniel E., Atienza, Audie A., Nilsen, Wendy, Allison, Susannah M., Mermelstein, Robin. Health behavior models in the age of mobile interventions: are our theories up to the task? Translational behavioral medicine. 2011; 1(1):53–71. (2011). [PubMed: 21796270]
57. Sallis, James F., Saelens, Brian E. Assessment of physical activity by self-report: status, limitations, and future directions. Research quarterly for exercise and sport. 2000; 71(sup2):1–14. (2000).
58. Schaefbauer, Christopher L., Khan, Danish U., Le, Amy, Sczechowski, Garrett, Siek, Katie A. Snack Buddy: Supporting Healthy Snacking in Low Socioeconomic Status Families. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; ACM; 2015. p. 1045-1057.
59. Andrew Schwartz, H., Eichstaedt, Johannes, Kern, Margaret L., Park, Gregory, Sap, Maarten, Stillwell, David, Kosinski, Michal, Ungar, Lyle. Towards assessing changes in degree of depression through facebook. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Citeseer. 2014. p. 118-125.
60. Senge, Robin, Del Coz, Juan Jos, Hllermeier, Eyke. On the problem of error propagation in classifier chains for multi-label classification. Springer; 2014. p. 163-170.
61. Sharma, Sanket S., De Choudhury, Munmun. Measuring and characterizing nutritional information of food and ingestion content in instagram. Proceedings of the 24th International Conference on World Wide Web; ACM; 2015. p. 115-116.
62. Signorini, Alessio, Segre, Alberto Maria, Polgreen, Philip M. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PloS one. 2011; 6(5):e19467. (2011). [PubMed: 21573238]
63. Sleeper, Manya, Acquisti, Alessandro, Cranor, Lorrie Faith, Kelley, Patrick Gage, Munson, Sean A., Sadeh, Norman. I Would Like To…, I Shouldn't…, I Wish I…: Exploring Behavior-Change Goals for Social Networking Sites. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; ACM; 2015. p. 1058-1069.
64. Spruijt-Metz, Donna, Hekler, Eric, Saranummi, Niilo, Intille, Stephen, Korhonen, Ilkka, Nilsen, Wendy, Rivera, Daniel E., Spring, Bonnie, Michie, Susan, Asch, David A., et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Translational behavioral medicine. 2015; 5(3):335–346. (2015). [PubMed: 26327939]
65. Stockwell, Tim, Hodgson, Ray, Edwards, Griffith, Taylor, Colin, Rankin, Howard. The Development of a Questionnaire to Measure Severity of Alcohol Dependence*. British Journal of Addiction to Alcohol Other Drugs. 1979; 74(1):79–87. (1979). [PubMed: 283831]

66. Tsoumakas, Grigorios, Katakis, Ioannis. Multi-label classification: An overview. Dept. of Informatics, Aristotle University of Thessaloniki; Greece: 2006. (2006)
67. Tucker, Michelle, Reicks, Marla. Exercise as a gateway behavior for healthful eating among older adults: an exploratory study. Journal of nutrition education and behavior. 2002; 34:S14–S19. (2002). [PubMed: 12047824]
68. Turner, Benjamin C., Jenkins, Emma, Kerr, David, Sherwin, Robert S., Cavan, David A. The effect of evening alcohol consumption on next-morning glucose control in type 1 diabetes. Diabetes Care. 2001; 24(11):1888–1893. (2001). [PubMed: 11679452]
69. Venerandi, Alessandro, Quattrone, Giovanni, Capra, Licia, Quercia, Daniele, Saez-Trumper, Diego. Measuring urban deprivation from user generated content. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; ACM; 2015. p. 254-264.
70. Volkova, Svitlana, Bachrach, Yoram, Armstrong, Michael, Sharma, Vijay. Inferring Latent User Properties from Texts Published in Social Media. AAAI. 2015:4296–4297.
71. Wang, Jing, Shih, Patrick C., Carroll, John M. Life after weight loss: design implications for community-based long-term weight management. Computer Supported Cooperative Work (CSCW). 2015; 24(4):353–384. (2015).
72. Wilsnack, Richard W., Vogeltanz, Nancy D., Wilsnack, Sharon C., Robert Harris, T. Gender differences in alcohol consumption and adverse drinking consequences: crosscultural patterns. Addiction. 2000; 95(2):251–265. (2000). [PubMed: 10723854]
73. Wood, Phillip K., Sher, Kenneth J., Rutledge, Patricia C. College student alcohol consumption, day of the week, and class schedule. Alcoholism: Clinical and Experimental Research. 2007; 31(7): 1195–1207. (2007).
74. Zaragoza, Julio H., Sucar, Luis Enrique, Morales, Eduardo F., Bielza, Concha, Larranaga, Pedro. Bayesian chain classifiers for multidimensional classification. IJCAI; Citeseer; 2011. p. 2192-2197.
75. Zhu, Xingquan, Wu, Xindong. Class noise vs. attribute noise: A quantitative study. Artificial Intelligence Review. 2004; 22(3):177–210.

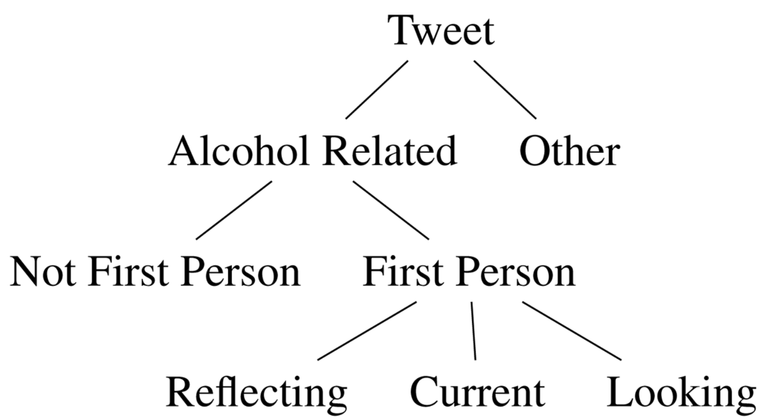

Figure 1.
Hierarchical classification scheme used to distinguish first-person, alcohol and further stages of drinking.

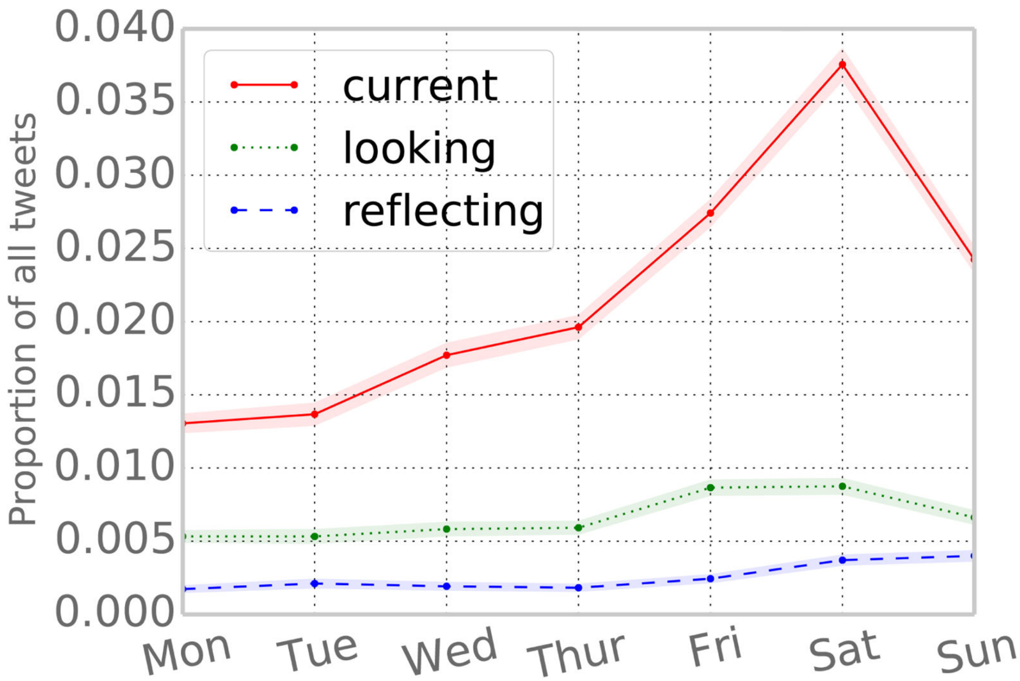

Figure 2.

Distribution of first person labels by day of week. Y-axis represents proportion of all Tweets. Day labels are aligned to midnight.

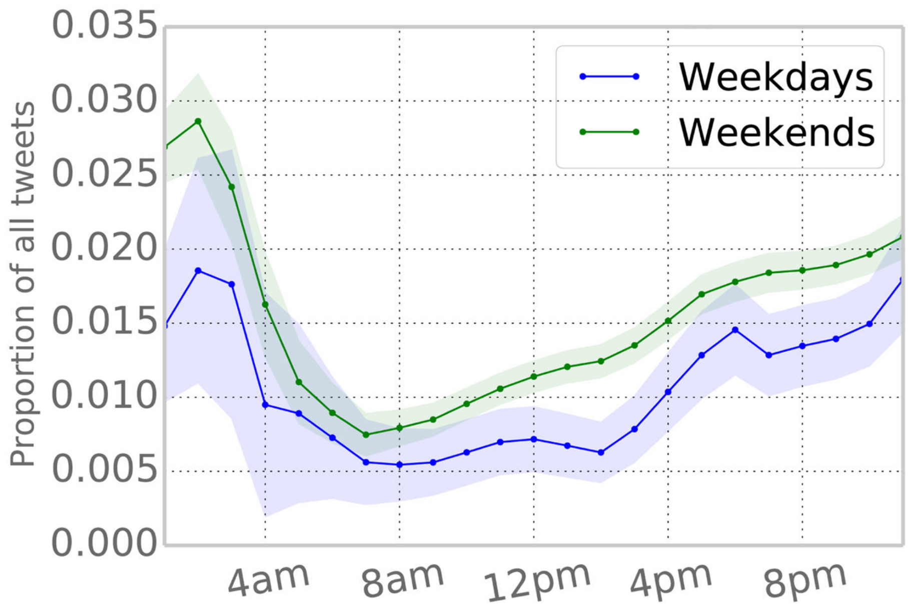

Figure 3.
Distribution of first person alcohol posts by hour. Y-axis represents proportion of all Tweets.

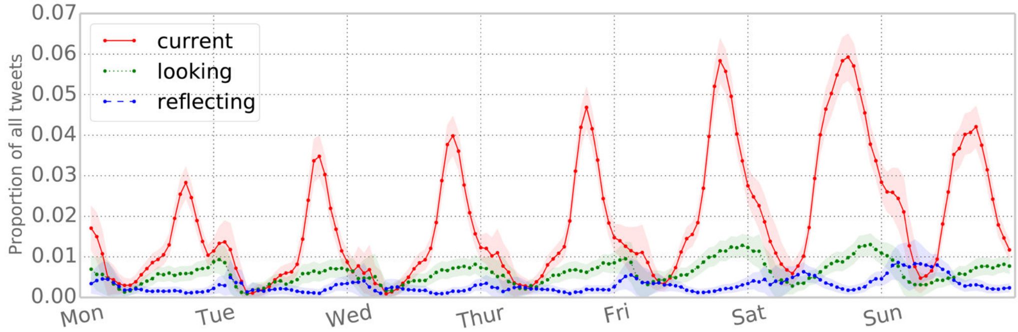

Figure 4.
Comparison between Twitter first-person alcohol stages during the week by hour. Day labels are aligned to midnight.

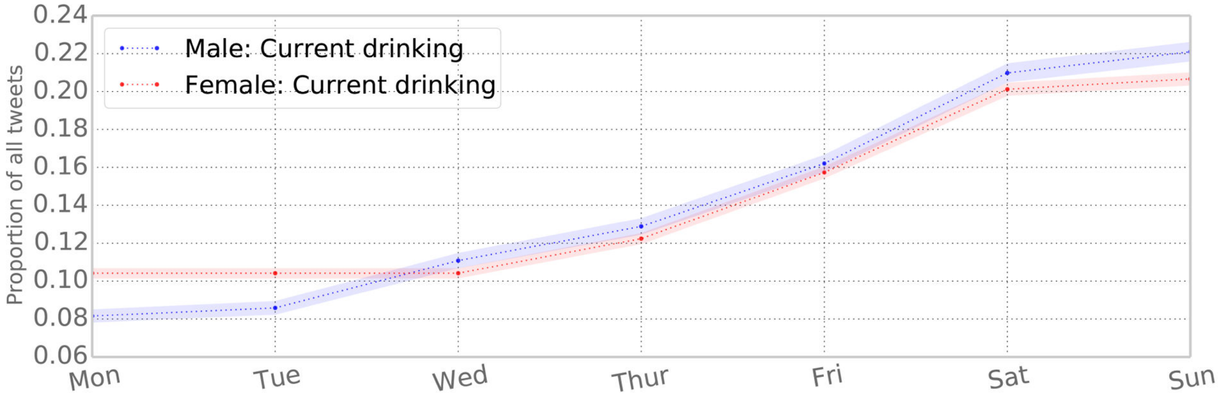

Figure 5.
Trends for men and women by hour week (current drinking)

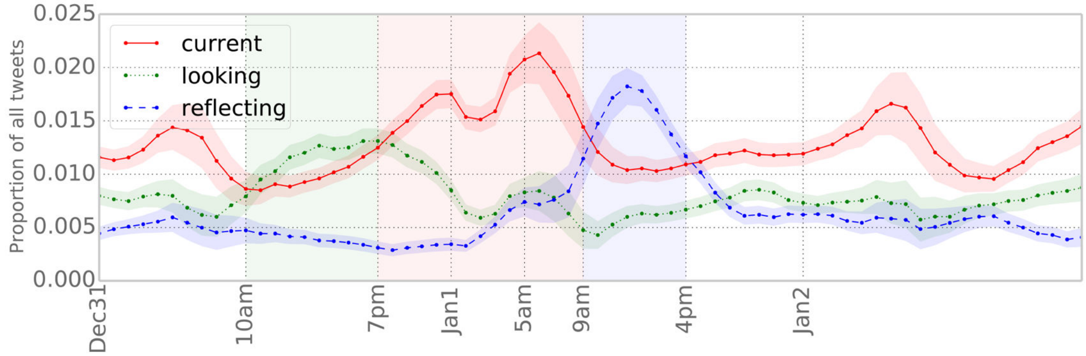

Figure 6.
First person alcohol stages through the New Years holiday period.

Table 1

Distribution of labeled data.

Labels	Alcohol Related	First Person
Pos	10295	6357
Neg	5355	3287

Current	Looking	Reflecting
3287	1676	1394

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Top features for each classification stage and their associated feature strength, β.

Class Label	Positive	β	Negative	β
alcoholic	drunk	52.44	coffee	−13.60
	beer	42.24	water	−13.59
	drinking	36.21	tea	−11.65
	wine	36.21	soda	−10.44
	drink	25.02	drunk no	−9.7
first person	i	16.55	you	−10.71
	drinking	16.10	she	−10.10
	i'm	12.70	he	−10.04
	drunk	10.04	people	−9.88
	with my	9.08	they	−9.45
future/looking	tonight	18.09	drunk	−17.43
	wanna	14.42	drinking	−11.45
	need	14.40	i drink	−8.12
	get drunk	11.86	drank	−6.45
	want	9.93	was	−6.45
present/current	drinking a	12.12	tonight	−10.57
	drunk	10.55	get drunk	−10.09
	last drink	7.59	was	−9.25
	I'm drunk	5.69	wanna	−9.00
	drink drink	5.84	who	8.96
past/reflecting	was	14.98	tonight	−8.14
	when	11.98	got drunk last	−7.75
	last night	9.91	drinking a	−6.63
	yesturday	6.52	wanna	−6.40
	never drinking	6.03	tomorrow	−4.73

Table 3

Error analysis (F1 score and area under curve) for each classification approach considered using 5-fold cross validation.

	LR		SVM		RF	
Class Label	**F1**	**AUC**	**F1**	**AUC**	**F1**	**AUC**
alcohol	.86	.87	.85	.85	.82	.84
first person	.76	.70	.72	.66	.77	.63
current	.72	.81	.68	.78	.70	.72
lookig	.64	.79	.54	.74	.53	.73
reflecting	.53	.77	.44	.60	.21	.55

Table 4

Highest probability day for each class.

Stage	High Probability Day
Looking	Tuesday, Wednesday, Thursday, Friday
Current	Saturday, Sunday
Reflecting	Monday, Tuesday, Wednesday