

Predicting Continued Participation in Online Health Forums

Farig Sadeque

Dept. of Computer and Information Sciences
University of Alabama at Birmingham
Birmingham, AL 35294-1170
farigys@uab.edu

Thamar Solorio

Dept. of Computer Science
University of Houston
Houston, TX 77204-3010
solorio@cs.uh.edu

Ted Pedersen

Dept. of Computer Science
University of Minnesota, Duluth
Duluth, MN 55812-3036
tpederse@d.umn.edu

Prasha Shrestha

Dept. of Computer Science
University of Houston
Houston, TX 77204-3010
pshrestha3@uh.edu

Steven Bethard

Dept. of Computer and Information Sciences
University of Alabama at Birmingham
Birmingham, AL 35294-1170
bethard@uab.edu

Abstract

Online health forums provide advice and emotional solace to their users from a social network of people who have faced similar conditions. Continued participation of users is thus critical to their success. In this paper, we develop machine learning models for predicting whether or not a user will continue to participate in an online health forum. The prediction models are trained and tested over a large dataset collected from the support group based social networking site `dailystrength.org`. We find that our models can predict continued participation with over 83% accuracy after as little as 1 month observing the user's activities, and that performance increases rapidly up to 1 year of observation. We also show that features such as the time since a user's last activity are consistently predictive regardless of the length of the observation period, while other features, such as the number of times a user replies to others, decrease in predictiveness as the observation period grows.

1 Introduction

Online social networks have established themselves as an integral part of the human interaction in the

21st century. Along with the most popular online social networking services like Facebook, Google+ and the micro-blogging website Twitter, there are other online social networks that are tailored to fit more specific purposes. Among them are the support group based social networks that provide help to individuals with physical or mental afflictions through the sharing of personal experiences and expert advice on a single platform.

Though many aspects of online social networks have attracted the attention of researchers, there is little research to date on computational assessments of engagement among users of online health forums. These forums provide us with information that is unique to this kind of social network: the networks are largely based on the emotional support among the users, so being able to successfully track a user's engagement on these services has the potential for greater impact than the more general social networks. If a support group platform can accurately predict when a user is thinking about leaving, they can take targeted actions to make a more favorable environment for the user, and thus maintain consistent emotional support for their other users. This kind of support is key the health and well-being of the users.

Predicting user engagement is related to concept of *churn prediction* in telecommunication networks (Ngonmang et al., 2012; Mozer et al., 1999; Das-

gupta et al., 2008) (predicting when a user will leave one service provider for another), where the motivation is that winning a new customer is more expensive than retaining an existing customer (Hadden et al., 2007). Similarly, retaining an engaged user in a health forum is likely to be easier than engaging a new user in the group. However, users of online support groups are not typically moving between providers, but rather deciding whether or not to continue to use an online support group at all.

In this work, we make the following contributions:

1. We develop models that observe a user for a single month and can predict whether the user will continue participating in the support group in the future with more than 83% accuracy, and can identify users that will leave the group with more than 88% precision and 80% recall.
2. We show that performance on predicting continuing participation rises as the observation period grows beyond one month, rising sharply up to nine months, and then more gradually up to 24 months.
3. We demonstrate a variety of features that are important for this prediction task and show that how often a user replies to others and the time elapsed since their last support group activity are some of the strongest features.
4. We find that the relative importance of the different features changes over time.

We believe this is the first work to look in detail at predicting engagement as it evolves over time in online health forums.

In this work, we focus on the contents of a user’s posts and replies and their timeline of activities, rather than the network of friendship links (which is sparse in forum-based social networks as compared to friendship-based social networks like Facebook). We also focus on *active participation*, such as initiating a thread or posting a reply, and not on *passive participation*, such as simply viewing the forum (since such passive information is only available to administrators of the support group service).

2 Data

Our data is collected from DailyStrength¹, one of the largest support group based online social networks with more than 500 support groups based on the physical and mental ailments of its users. Users in these support groups can either *post*, creating a new thread on a new topic, or they can *reply* to a thread that someone else has created. We focused on 20 support groups: Acne, ADHD (Attention Deficit Hyperactivity Disorder), Alcoholism, Asthma, Back Pain, Bipolar Disorder, Bone Cancer, COPD (Chronic Obstructive Pulmonary Disease), Diets and Weight Maintenance, Fibromyalgia, Gastric and Bypass Surgery, Immigration Law, Infertility, Loneliness, Lung Cancer, Migraine, Miscarriage, Pregnancy, Rheumatoid Arthritis and War in Iraq.

We crawled all of the posts (thread initiations) and replies (to existing threads) for these support groups from the earliest available post until the end of September 2013. The posts and replies were downloaded as HTML files, one per thread, where each thread contains an initial post and zero or more of replies. The HTML files were parsed and filtered for scripts and navigation elements to create XML files containing only the users, dates, posts and replies. Each extracted post and reply was part-of-speech tagged using the Stanford part-of-speech tagger (Manning et al., 2014) and was tagged for emotion words by matching it against the Linguistic Inquiry and Word Count (LIWC)² lexicon. We also collected the user profile pages of all the users who took part in any form of activity in any of these 20 support groups. Finally, we filtered out the users with the most incomplete profiles, where they were missing both age and gender. These users do not appear in the train, development or test sets, but the replies they post on other users’ posts who are not filtered out contribute to the participation prediction task of those users. We also filtered out the user *DS*, the only administrative user in DailyStrength.

Table 1 provides an overview of the resulting dataset. The largest support group among the 20 is Gastric and Bypass Surgery with 21507 posts and 158020 replies and the smallest is the Bone Cancer support group with only 40 posts and 51 replies. The amount of individual activity also varies greatly as there are people who posted or replied only once in their lifetime, and there are

¹<http://www.dailystrength.org>

²<http://www.liwc.net/>

Support groups	20
Posts	110316
Replies	788119
Users	39905

Table 1: Summary of the data collected from DailyStrength

people who have more than 5000 posts or replies.

3 Model

Our goal is to build a model that can observe the past activity of a user on the forum and predict whether the user will continue to participate in the forum in the future. Formally, we would like to construct a model:

$$m_{\Delta t}(u) = \begin{cases} 1 & \text{if } \exists a \in A : a.u = u \wedge \\ & a.t > u.t + \Delta t \\ 0 & \text{otherwise} \end{cases}$$

where u is a user, Δt is an amount of time which we call the *observation period*, A is the set of all activities (from any user at any time) such as posting or replying to a post, $a.u$ is the user whose activity it was, $a.t$ is the time of the activity, and $u.t$ is the time at which the user account was created. Intuitively, m should predict 1 iff Δt time has elapsed since the user created their account and there is any new participation (posting or replying) any time in the future after that.

We treat this as a supervised classification problem, and represent a user based on all of his/her activities in the forum during the observation period:

$$rep_{\Delta t}(u) \stackrel{\text{def}}{=} \{a \in A : a.u = u \wedge a.t \leq u.t + \Delta t\}$$

The following sections describe classifier features that are derived from this representation.

3.1 Activity features

These features gather information of a user’s activity on DailyStrength. In general, we would expect users who are more active during the observation period to also be more likely to continue to participate in the future.

PostCount The number of threads a user has initiated on the DailyStrength website over the observation period.

ReplyCount The number of replies a user has posted to other users’ posts on the DailyStrength website over the observation period.

SelfReplyCount The number of replies a user has posted to their own posts over the observation period.

OtherReplyCount The number of replies a user has received to their posts from other users over the observation period.

3.2 Time features

These features provide a look into the timing of a user’s participation on DailyStrength. In general, we would expect users who are participating frequently throughout the observation period to be more likely to participate in the future.

TimeGap1 The number of days between the point at which the user created their DailyStrength account and their first activity (post or reply). This is a measure of how long it took a user to start actively participating in the community.

TimeGap2 The number of days from the time of the last post or reply of a user to the end of the observation period. This is a measure of how long the user has been idle since their last activity.

AvgDays The average number of days between any two sequential activities (posts or replies) by the user during the observation period. This is a measure of how often a user is idle.

3.3 Personal features

These features are gathered from a user’s account information page. Since providing age, gender, location and a profile photo are all optional during the DailyStrength account creation process, many users are missing one or more of these pieces of information. In general, we would expect users with more complete profiles to be more likely to continue to participate.

Age The user’s age.

Gender The user’s gender, either *male*, *female* or *unknown*.

HasLocation A binary feature representing whether or not the user has provided their location.

HasImage A feature representing whether or not the user has provided a profile photo, and if provided, whether it is a stock photo or a user-provided one.

3.4 Content features

These features examine the content of the text in the posts and replies of a user. In general, we would expect users with longer posts to be more likely to continue to participate than users with short posts.

PosUnigrams The total number of words over the observation period that were identified as positive emotions by the LIWC lexicon.

NegUnigrams The total number of words over the observation period that were identified as negative emotions by the LIWC lexicon.

TotalUnigrams The total number of words a user posted over the observation period. This includes all the words (including stop words), not only the emotion words.

Question The total number of questions the user has asked over the observation period in either posts or replies. Questions were identified by looking for sentences ending in question marks.

Url The total number of URLs a user has posted over the observation period.

4 Experiments

For all of the following experiments, we divided the users in our corpus into train, development and test sets with a 60-20-20 ratio, that is, we used 60% of the users to train our prediction model, 20% of the users as the development set and the remaining 20% of the users to test the model. Users were partitioned into each of these sets randomly. These sets do not change with the changes in the observation period, thus giving us the opportunity to compare the results from all observation periods.

We trained classifiers based on WEKA v3.6.11 (Hall et al., 2009), a widely used machine learning toolkit. We normalized all of the aforementioned features to the range $[0, 1]$ to make feature weights more comparable and interpretable. We initially explored several classifiers: naïve Bayes, logistic regression, support vector machines and J48 (a decision-tree based classifier). Logistic regression outperformed the other three in evaluations on the

development set, so all results reported here use logistic regression.

We rely on several different performance measures to evaluate our models. First, we report simple classification *accuracy*, the fraction of users for which we correctly predicted whether they would continue participating or leave the forum. We compare this to the *baseline accuracy*, the accuracy of a model that predicts that no users will continue to participate in the forum, and we report *error reduction* of the model accuracy relative to this baseline accuracy. We also report performance measures on the task of identifying users who will not participate in the future: *precision*, the fraction of the users that our model predicted would stop participating who did in fact stop, *recall*, the fraction of the users known to have stopped participating that our model predicted would stop, and *F-measure*, the harmonic mean of precision and recall. Formally:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F\text{-measure} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

where TP is the number of *true positives* (users the model predicted would stop participating and did in fact stop), TN is the number of *true negatives* (users the model predicted would continue participating and did in fact continue), FP is the number of *false positives* (users the model predicted would stop participating but actually continued participating) and FN is the number of *false negatives* (users the model predicted would continue participating but actually stopped participating).

4.1 Can continued participation be predicted?

Our first research question is whether a user’s continued participation on the forum can be predicted given the features we developed in Section 3. To test this, we consider an observation period of 1 month and train and test the corresponding classifier. The first row of Table 2 shows the results. Our model achieves 83.06% accuracy, compared to the 50.48% accuracy of the baseline model. For the task of identifying just those users that have

Period	Baseline	Accuracy	Error Reduction	Precision	Recall	F-measure
1	50.48	83.06	65.81	88.3	80.2	84.0
3	63.53	85.69	60.76	92.0	86.4	89.1
6	72.01	87.69	56.02	94.7	89.0	91.7
9	77.75	89.12	51.10	94.9	91.4	93.1
12	82.19	90.71	48.01	96.3	92.7	94.5
15	85.09	92.03	46.54	97.1	93.8	95.4
18	86.96	92.29	40.87	97.3	94.0	95.6
21	87.65	92.34	37.98	97.7	93.8	95.7
24	87.84	92.34	37.01	97.8	93.7	95.7

Table 2: Performance across different observation periods (Period, in months), in terms of baseline accuracy (Baseline, %), model accuracy (Accuracy, %), error reduction of the model over the baseline (ErrRed, %), precision (%), recall (%) and F-measure (%). Precision, recall and F-measure are on the task of identifying users who will not participate in the future.

stopped participating, we achieve 88.3% precision and 80.2% recall. These high performance numbers suggest that while our models are still imperfect, our features are capturing a large proportion of the information necessary to predict continued participation.

4.2 How long must a user be observed?

Our second research question aims to determine the optimal observation period for predicting continued participation. For this experiment, we created 9 observation periods: 1 month, 3 months, 6 months, 9 months, 12 months, 15 months, 18 months, 21 months and 24 months. We then evaluated models trained on these different evaluation periods to see how performance increased or decreased.

Table 2 shows the results. Model accuracy always rises as the observation period grows longer, ranging from 83.06% at 1 month to 92.34% at 24 months. However, the biggest gains are in the shorter periods, with the model increasing accuracy by 7.65% between 1 and 12 months, but only by 1.63% between 12 and 24 months. The performance of the baseline model also increases with the size of the observation period, so that after 24 months 87.84% of all users will not return.

For the task of identifying just those users that have stopped participating, we observe that precision and recall also both rise as the observation period grows, with precision making moderate gains, from 88.32 at 1 month to 97.8 at 24 months, and recall making larger gains, from 80.20 at 1 month to 93.7 at 24 months. As with accuracy, the biggest gains are between 1 and 3 month observation periods.

Overall, these results suggest that observing a user for even 1 month gives reasonable performance, observing for 12 months gives noticeably better performance, and observing for longer than 12 months gives diminishing returns.

4.3 Which features are most important?

Our third research question aims to prioritize our features based on how useful they are to the task of predicting continued participation. To investigate this, we turn to the coefficients (weights) for the independent variables (features) in our logistic regression, which represent the importance of each variable in the classification model. The larger the absolute value of the coefficient, the bigger the impression of that variable on the output. The sign indicates positive or negative effect of that variable on the result, where a negative value means that the feature is associated with continued participation, while a positive value means that the feature is associated with stopping participation.

Table 3 shows the weights of the features obtained from the test data for a 1-month observation period. The most important features (the features with the highest absolute values) are the number of times the user has replied to other users (ReplyCount), the time since the user’s last activity (TimeGap2), the time between creating a DailyStrength account and the user’s first post (TimeGap1) and the content (Unigram) features. The least important features are mostly the ones aimed at measuring completeness of the profile (Age, Gender, etc.), suggesting that profile completeness is not a good predictor of continued participation. However, the presence of a profile photo

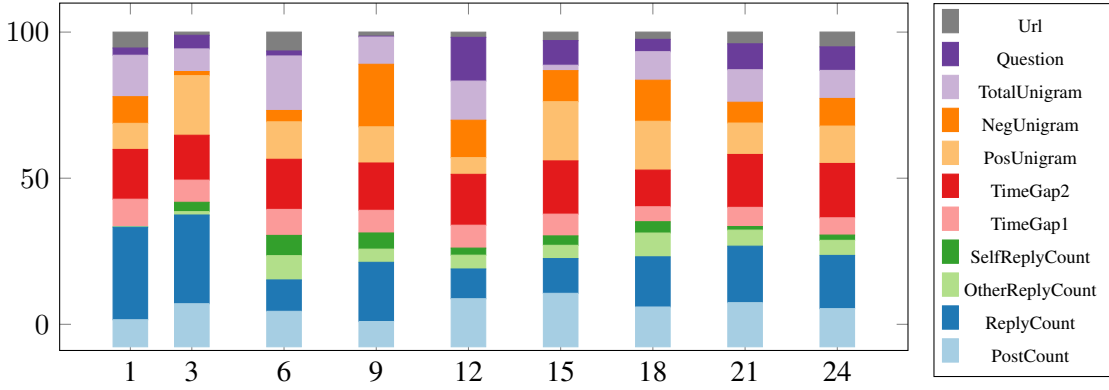


Figure 1: Relative importance of features over the different observation periods. The height of a bar segment represents the absolute value of the weight of the feature, scaled so that the sum of the feature weights is 100%.

Feature	Weight
ReplyCount	-10.652
TimeGap2	5.727
TotalUnigram	4.772
TimeGap1	3.172
NegUnigram	-3.069
PosUnigram	3.002
Url	1.834
HasImage	-0.845
Question	0.827
AvgDays	0.809
PostCount	0.526
Age	-0.346
Gender=unknown	0.202
Gender=female	-0.110
Gender=male	0.093
HasLocation	-0.092
OtherReplyCount	-0.051
SelfReplyCount	0.001

Table 3: Weights of the features for a 1-month observation period

(HasImage) did make a small contribution to the model.

The signs of the weights of the features reveal the direction of predictiveness. The TimeGap1 and TimeGap2 weights are positive, indicating that longer gaps between activities predict someone leaving the forum. PostCount is positive while ReplyCount is negative, suggesting that people who only post will likely leave the forum, while people who reply to others will likely stay. Posting questions and URLs are associated with leaving the forum, along with higher usage of positive uni-

grams, while higher usage of negative unigram is associated with continued participation.

4.4 Does feature importance change over time?

Our fourth research question asks whether the importance of features is consistent across all observation periods, or whether some features become more or less important than others as the observation period grows.

Figure 1 shows the percentage importance of the eleven most significant features over the different observation periods. Features like TimeGap1 and TimeGap2 are fairly stable in importance over time, with TimeGap1 accounting for 5-9% of the weight and TimeGap2 accounting for 12-18%. ReplyCount is a very strong feature, accounting for as much as 30% in the 1 and 3 month observation periods, but it receives a lower weight for longer observation periods (as little as 10% in the 12 month period). SelfReplyCount and OtherReplyCount, which had almost no weight in the 1 month model, increase in importance for longer observation periods. The other features have less consistent patterns. For example, content features (TotalUnigram, NegUnigram, PosUnigram, Question, Url) account for around 40% of the model weights for most observation periods, but the distribution of weight across these 5 features is erratic over time.

As another measure of feature importance over time, Table 4 shows the increase in accuracy over the baseline majority class model for models trained using only a single feature. Note that the baseline model’s accuracy increases for longer observation periods (because more users leave), so

OP	BL	PC	RC	OR	SR	TG1	TG2	AvD	Age	Gen	Loc	Img	Pos	Neg	TUn	Que	Url
1	50.48	4.91	6.55	2.97	2.13	0.00	31.7	4.72	0.36	-0.20	0.00	11.9	5.10	4.32	4.98	4.03	0.10
3	63.53	3.05	3.75	2.43	1.42	0.00	20.7	0.00	0.14	0.00	0.00	0.00	3.06	2.65	2.73	2.52	0.30
6	72.01	1.53	1.89	1.40	0.77	0.00	14.9	0.00	0.00	0.00	0.00	0.00	1.60	1.43	1.50	1.46	0.11
9	77.75	0.81	1.03	0.74	0.24	0.00	10.8	0.00	0.00	0.00	0.00	0.00	0.73	0.72	0.82	0.83	0.07
12	82.19	0.13	0.29	0.11	-0.13	-0.20	7.76	-0.20	-0.20	-0.20	-0.20	-0.20	0.15	0.09	0.13	0.03	-0.18
15	85.09	0.43	0.47	0.37	0.27	0.15	6.16	0.14	0.14	0.14	0.14	0.14	0.33	0.25	0.35	0.36	0.25
18	86.96	0.23	0.20	0.18	0.05	0.00	4.92	0.00	0.00	0.00	0.00	0.00	0.15	0.10	0.09	0.13	0.02
21	87.65	0.08	0.14	0.00	0.00	0.00	4.43	0.00	0.00	0.00	0.00	0.00	0.07	0.07	0.09	0.19	0.02
24	87.84	0.08	0.19	0.01	-0.01	0.00	4.42	0.00	0.00	0.00	0.00	0.00	0.08	0.07	0.09	0.20	0.01

Table 4: Accuracy gain over baseline (BL) over observation periods (OP) when a classifier is trained using only a single feature: PostCount (PC), ReplyCount (RC), OtherReplyCount (OR), SelfReplyCount (SR), TimeGap1 (TG1), TimeGap2 (TG2), AvgDays (AvD), Age, Gender (Gen), HasLocation (Loc), HasImage (Img), PosUnigram (Pos), NegUnigram (Neg), TotalUnigram (TUn), Question (Que), Url

the absolute gains over the baseline always correspondingly decrease. TimeGap2 (TG2) always gives the largest increase in accuracy on its own, as much as 31.7% at a 1 month observation period, and is the only feature that continues (by itself) to give gains over the baseline all the way out to 24 months. ReplyCount (RC) is the next best feature by itself, achieving 6.55% improvement over the baseline at a 1 month observation period, but dropping to less than a 1% improvement by 12 months. The content features PosUnigram (Pos), NegUnigram (Neg), TotalUnigram (TUn) and Question (Que) each achieve a 4-5% improvement over the baseline for a 1 month observation period, but drop below a 1% improvement by 9 months. The personal features generally achieve very little on their own, except for HasImage (Img), which is very useful at 1 month (giving a 11.9% improvement), but giving no improvement for any other observation period.

5 Related Work

Though we are not aware of other models that can observe an online support group user over time and predict their continued participation, there are several works analyzing related problems in other types of social networks. Ngonmang et al. (2012) has worked on a similar problem on a French online blog network called Skyrock³. They neither used the contents in the users' posts, nor analyzed the users' behavior. Rather, they used the friendship relationship among the users to predict future participation. There has also been some works on Usenet newsgroups (Joyce and Kraut, 2006; Ar-

guello et al., 2006) where the models take a single post of a user and predict whether or not it will receive a reply. Significant predictors for this task include whether or not the message is cross-posted, the topical coherence of the message with the newsgroup, whether the user posts a question, whether the user is a newcomer to the newsgroup, and the use of third person pronouns. Lampe and Johnston (2005) have shown the effect of feedback on a new user in Slashdot, a news and discussion site. They calculated the user's first comment score and likelihood of getting a second comment by the user based on the feedback the comment received from the other users. They also introduced the time gap between two activities of a user as an indicator of socialization. Mahmud et al. (2014) analyzed word usage to predict social engagement behavior in Twitter. They used psycholinguistic word categories from LIWC, and showed how these categories influence reply and retweet behavior of users. Chen and Pirolli (2012) have analyzed factors that influenced users' engagement in the Occupy Wall Street movement. Danescu-Niculescu-Mizil et al. (2013) has used linguistic change as a predictor of user lifespan in social networks. They conducted their experiments on BeerAdvocate and RateBeer and attempted to find the lifespan of a user early in his or her career.

There are also some studies that include DailyStrength as a data source. Wiley et al. (2014) examined the characteristics of ten different online social networking sites to find impacts of these characteristics on the discussions of pharmaceutical drugs among the users and DailyStrength was one of these websites. Sarker et al. (2015) performed

³<http://www.skyrock.com>

a study on automatic monitoring of Adverse Drug Reactions (ADRs) using user-posted data on social media, and DailyStrength was a data source for their study.

6 Discussion

Our findings have several implications for social interaction in online health forums. This is the first study that attempts to predict continued participation of users in such support groups. Though the model is not perfect, it produces results with high accuracy, precision and recall. The high precision and recall has greater significance in this experiment, as they represent our model's correctness in identifying the people who leave the group after a certain observation period. Identifying these people early in their lifecycle will help social health platforms identify users that are not being fully served, allowing the platforms to analyze the reason for the departure and create a more favorable environment for everyone.

This is also the first study that examines the effect of different lengths of observation period to determine the minimum amount of time required to accurately predict future participation. With a 12-month observation period, we can predict continued engagement with high accuracy, precision and recall, though even at a 1-month observation period, performance is good.

Our work has shown which features contribute the most to predict a user's continued participation. As we can see from the results, personal features covering demographics and profile completeness play little to no part in predicting user's engagement, whereas the other three categories have varied significance over time. The predictiveness of time based features, especially the time from account creation until a user's first activity and time since a user's last activity, are consistently predictive over all lengths of observation. The predictiveness of replies to other users' posts is very large for 1 and 3 month observation periods, but is a little less informative for larger observation periods. The predictiveness of content features (word count, negative/positive words, etc.) is generally good, though which of these features is most important varies somewhat over time.

Although the model we built produces high performance results, there are several opportunities to improve it further. As our results show that the usage of positive and negative unigrams has

some influence over the prediction task, we plan to expand these features by using additional psycholinguistic word categories to find other relations between emotional status and continued participation. We also plan to expand beyond word unigrams, which fail to account for phenomena such as negation (where *good* becomes *not good*), incorporating longer linguistic dependencies into these features. And we plan to use more linguistic features to capture explicit speech acts in the posts of the user that may indicate their intent to leave the forum. In addition, we plan to analyze the replies received by other users to find out whether there is a pattern that encourages or discourages a user's participation, such as a long duration between posting and receiving a reply, or the use of harsh or aggressive forms of language. Finally, we plan to explore machine learning formulations that will allow us to dynamically extend our observation period for a user to just the point at which we can confidently predict whether or not they will continue to participate.

7 Conclusion

In this paper we presented a study to determine what makes a user continue his or her participation in a support group based online health forum and how long we have to observe a user's activity to predict this accurately. We built a model that predicts continued participation of a user and we showed that this model is accurate with an observation period as little as one month. Increasing the observation period increases performance, though most of the gains are achieved by the end of 12 months. The model reveals that features like the time since a user's last activity and the number of times a user has replied to others are consistently strong predictors of continued participation. Our model forms a foundation for future research in modeling the evolution over time of user engagement in online health forums.

Acknowledgments

We would like to thank Binod Gyawali who did the initial data gathering and data coding.

References

Jaime Arguello, Brian S. Butler, Elisabeth Joyce, Robert Kraut, Kimberly S. Ling, Carolyn Rosé, and Xiaoqing Wang. 2006. Talk to me: Foundations

- for successful individual-group interactions in on-line communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 959–968, New York, NY, USA. ACM.
- Jilin Chen and Peter Pirolli. 2012. Why you are more engaged: Factors influencing twitter engagement in occupy wall street. In *Sixth International AAAI Conference on Weblogs and Social Media*.
- Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 307–318, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- Koustuv Dasgupta, Rahul Singh, Balaji Viswanathan, Dipanjan Chakraborty, Sougata Mukherjee, Amit A Nanavati, and Anupam Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, pages 668–677. ACM.
- John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta. 2007. Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10):2902–2917.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November.
- Elisabeth Joyce and Robert E. Kraut. 2006. Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11(3):723–747.
- Cliff Lampe and Erik Johnston. 2005. Follow the (slash) dot: Effects of feedback on new members in an online community. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*, GROUP '05, pages 11–20, New York, NY, USA. ACM.
- Jalal Mahmud, Jilin Chen, and Jeffrey Nichols. 2014. Why are you more engaged? predicting social engagement from word use. *CoRR*, abs/1402.6690.
- Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland, June. Association for Computational Linguistics.
- Michael Mozer, Richard H Wolniewicz, David B Grimes, Eric Johnson, and Howard Kaushansky. 1999. Churn reduction in the wireless industry. In *NIPS*, pages 935–941.
- Blaise Ngonmang, Emmanuel Viennet, and Maurice Tchente. 2012. Churn prediction in a real online social network using local community analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, ASONAM '12, pages 282–288, Washington, DC, USA. IEEE Computer Society.
- A. Sarker, A. Nikfarjam, K. O'Connor, R. Ginn, G. Gonzalez, T. Upadhaya, S. Jayaraman, and K. Smith. 2015. Utilizing social media data for pharmacovigilance: A review. *J Biomed Inform*, Feb.
- M. T. Wiley, C. Jin, V. Hristidis, and K. M. Esterling. 2014. Pharmaceutical drugs chatter on Online Social Networks. *J Biomed Inform*, 49:245–254, Jun.