Discourse Analysis of User Forums in an Online Weight Loss Application

Lydia Manikonda[1], **Heather Pon-Barry**[1], **Subbarao Kambhampati**[1], **Eric Hekler**[2]
David W. McDonald[3]

[1]School of Computing, Informatics, and Decision Systems Engineering, Arizona State University
[2]School of Nutrition and Health Promotion, Arizona State University
[3]The Information School, University of Washington
{lmanikon, ponbarry, rao, ehekler}@asu.edu, dwmc@uw.edu

Abstract

Online social communities are becoming increasingly popular platforms for people to share information, seek emotional support, and maintain accountability for losing weight. Studying the language and discourse in these communities can offer insights on how users benefit from using these applications. This paper presents a preliminary analysis of language and discourse patterns in forum posts by users who lose weight and keep it off versus users with fluctuating weight dynamics. Our results reveal differences about how the types of posts, polarity of sentiments, and semantic cohesion of posts made by users vary along with their weight loss pattern. To our knowledge, this is the first discourse-level analysis of language and weight loss dynamics.

1 Introduction and Related Work

Obesity is a major public health problem; the number of people suffering from obesity has risen globally in the last decade (Das and Faxvaag, 2014). Many of these people are trying to lose weight as the multifactorial diseases such as metabolic syndromes, respiratory problems, coronary heart disease, and psychological challenges are all closely associated with obesity (Rippe et al., 1998; Must et al., 1999). More obese people are trying to lose weight by using weight-loss applications and other people interested in using these applications are trying to avoid gaining weight. Many internet services are becoming increasingly popular for supporting weight loss as they provide users with the opportunities to seek information by asking questions, answering questions, sharing their experiences and providing emotional support. Also, the internet provides many attributes that can help people feel more comfortable with openly expressing their problems and concerns (Ballantine and Stephenson, 2011; Hwang et al., 2010).

Most of the existing studies (Saperstein et al., 2007; Johnson and Wardle, 2011; Hwang et al., 2010; Ballantine and Stephenson, 2011; Leahey et al., 2012; Das and Faxvaag, 2014) focused on why people participate in online weight loss discussion forums and how the social support can help them to lose weight. These studies are conducted from the perspective of medical and psychology domains, where the data are collected via interviews or a small set of online forum data that are manually analyzed by human experts. Their primary focus is on measuring the social support by collecting views/opinions of people through surveys; less attention is given to understanding the natural language aspects of users' posts on these online communities. Unlike choosing a small subset of a dataset, our work is novel in automating the process of language analysis that can handle a larger dataset. Automating the process can also help classify the user type based on the language efficiently. This work also considers weekly check-in weights of users along with the study of their language.

In this paper, we study the user's language in correlation with their weight loss dynamics. To this end, we analyze a corpus of forum posts generated by users on the forum of a popular weight loss application. The forum from which we obtained the data is divided into several threads where each thread consists of several posts made by different users. From the overall dataset we identify two preliminary patterns of weight dynamics: (1) users who lose weight and successfully maintain the weight loss (*i.e.*, from one week to the next, weight is lost or weight remains the same) and (2) users whose weight pattern fluctuates (*i.e.*, from one week to the next, weight

changes are erratic or inconsistent). While there are many possible groupings that we could have utilized, we chose this grouping because of the known problems with "yo-yo" dieting compared to a more steady weight-loss. We study how the user's language in these two groups varies by measuring the semantic cohesion and sentiment of posts made by them.

Our main contributions include understanding the types of posts users make on different threads with a main focus on question-related posts, the type of language they use by measuring the semantic cohesion and sentiment by correlating with users' weight loss patterns. From the empirical analysis we find that users who lose weight in a fluctuating manner are very active on the discussion forums compared to the users who follow a non-increasing weight loss pattern. We also find that users of non-increasing weight loss pattern mostly reply to the posts made by other users and fluctuating users post more questions comparatively. Both the users from these two clusters differ in terms of the way their posts cohere with previous posts in the threads and also in terms of the sentiment associated with their posts.

2 Dataset

We obtain a text corpus of online discussion forums from *Lose It!*, a popular mobile and web-based weight loss application. Along with the text corpus, we also obtain weekly weight check-in data for a subset of users. The entire corpus consists of eight different forums that are subdivided into conversation topic threads. Each thread consists of several posts made by different users. The forum data in our corpus consists of 884 threads, with a median length of 20 posts per thread. The posts were made between January 1, 2010 and July 1, 2012. We identify the subset of users for whom we have weight check-in data and who made at least 25 weight check-ins during this time period. This results in a total of 2,270 users.

The interesting feature of this weight loss application is that users are encouraged to set goals to regularly log their weight, diet, and exercise. For a subset of users, Lose It! has provided a weekly weight "check-in", an average of the user's weight check-ins during the week, for the January 1, 2010 through July 1, 2012 period. This allows us to juxtapose the weekly weights of the users with their posts on the discussion forums.
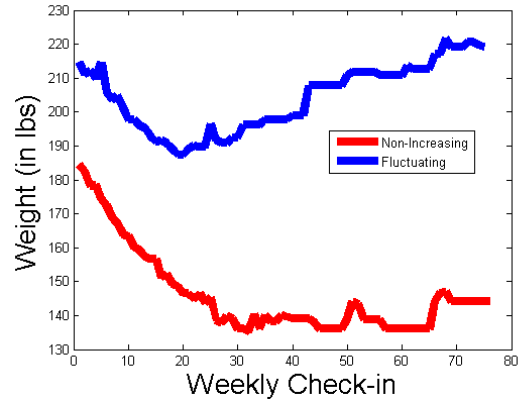

Figure 1: Example weight loss patterns from two individual users: non-increasing (bottom line), and fluctuating (top line). The x-axis ranges from the 1st through the 80th weekly check-in; the y-axis shows the weight, measured in lbs.

We partition the users into two groups based on their dynamic weight loss patterns: a *non-increasing* group and a *fluctuating* group.

1. **Non-increasing**: For each week j, the user's check-in weight w_j is less than or equal to their past week's weight w_{j-1}, within a small margin Δ. That is, $w_j \leq (1 + \Delta)w_{j-1}$.

2. **Fluctuating**: If the difference between two consecutive weekly check-in weights do not follow the non-increasing constraint, users are grouped into this category.

We empirically set $\Delta = 0.04$ to divide the users in our dataset into two groups of similar size. To illustrate the two patterns of weight change, Figure 1 shows the weekly weight check-ins of two individual users, one from each group. This grouping is coarse, but is motivated by studies (Kraschnewski et al., 2010; Wing and Phelan, 2005) acknowledging that approximately 80% of people who set out to lose weight are successful at long-term weight loss maintenance, where successful maintenance is defined as losing 10% or more of the body weight and maintaining that for at least an year. In the future for further analysis, we aim to separate users less coarsely, e.g., users who maintain their weight neither gaining nor losing weight, users who lose weight and maintain it and finally, users who gain weight.

2.1 Characteristics of Online Community

The Lose It! application helps users set a personalized daily calorie budget, track the food they are eating, and their exercise. It also helps users to stay motivated by providing an opportunity to

connect with other users who want to lose weight and support each other. Example snippets (paraphrased) from forum threads are shown below. The *"Can't lose weight!"* thread demonstrates users supporting each other and offering advice. The *"Someday I will"* thread highlights the complex relationship between text, semantics, and motivation in the forums.

Example thread: "Can't lose weight!"

User 1: *"I gained over 30 lbs in the last year and am stressed about losing it. I eat 1600 calories a day and burn more than that in exercise, but I havent lost any weight. I am so confused."*

User 2: *"You've only been a member for less than 2 months. I suggest you relax. Set your program to 1 pound weight loss a week. Adjust your habits to something you can live with. . . long term."*

User 3: *"You sound just like me. I think your exercise is good but maybe you are eating more than you think. Try diligently logging everything you consume."*

User 1: *"Thanks for the suggestions! I am going to get back to my logging."*

Example thread: "Someday I will. . . "

User 1: *"Do a pull-up :-)"*

User 2: *". . . actually enjoy exercising."*

User 3: *"Someday I will stop participating in the lose it forums, but obviously not today."*

User 4: *"I hope you fail :-)"*

3 Empirical Analysis

In this section, we present preliminary observations on how the language and discourse patterns of forum posts vary with respect to weight loss dynamics. As an initial step, part-of-speech (POS) tagging is performed on all forum posts using the Stanford POS Tagger (Toutanova et al., 2003).

From the weekly check-in data we identified the number of users and the number of posts from each weight-loss pattern cluster which are shown in Table 1. We see that the average number of posts by fluctuating users is greater than the average number of posts by non-increasing users.

	Weight Pattern	
	Non-increasing	Fluctuating
# Total users	1127	1143
# Forum users	29	68
# Forum posts	99	1279
Posts per user	3.5	18.2
Words per post	49.1	77.3

Table 1: Statistics of users and forum posts.

This suggests that fluctuating users are more active in participation. Our data also suggest that posts made by non-increasing users are shorter compared to those made by fluctuating users.

3.1 Asking Questions

Previous studies (Bambina, 2007; Langford et al., 1997) revealed that people on online health communities mainly engage in two activities: (*i*) seeking information, and (*ii*) getting emotional support. People usually ask questions to other community members or just browse through the community forums to get information while seeking information. Below is an example (paraphrased) showing how a users ask and respond to questions.

Example thread: "New user"

User 1: *"Did anyone upgrade to the premium app? What do you like about it?"*

User 2: *"I upgraded to the premium. I LOVE the functionality to log food in advance. I can track and set goals that are not related to weight like how much I sleep, how much water I drink, etc."*

User 3: *"I upgraded my account to premium too. I really liked the added features because it helped me keep track of my steps and participate in challenges."*

We are interested in knowing whether users in the two clusters are actively involved in posting questions. We deem a forum post to be a question if it meets one of these two conditions:

1. **Wh-question words:** If a sentence in the post starts with a question word: *Wh-Determiner (WDT), Wh-pronoun (WP), Possessive wh-pronoun (WP$), Wh-adverb (WRB).*

2. **Punctuation:** If the post contains a question mark symbol ('?').
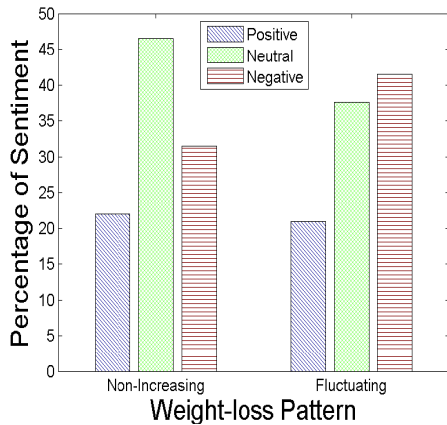
Figure 2: Proportion of sentiments for the two weight-loss patterns. For non-increasing users, percentage of posts with *Positive*, *Neutral* and *Negative* sentiments are: *22%*, *46.5%* and *31.5%* respectively. For fluctuating users, the percentage of posts with *Positive*, *Neutral* and *Negative* sentiments are: *20.9%*, *37.6%* and *41.5%* respectively.

We computed the ratio of question-oriented posts made by each user in the two clusters. After averaging these ratio values across all the users in each cluster separately, we found that on average, **32.6%** of the posts made by non-increasing users were questions ($SE = 0.061$). And, **37.7%** of the posts made by fluctuating users were questions ($SE = 0.042$). This shows that on an average fluctuating users post relatively more number of questions than the non-increasing users.

3.2 Sentiment of Posts

Analyzing the sentiment of user posts in the forums can provide a suprisingly meaningful sense of how the loss of weight impacts the sentiment of user's post. In this analysis, we report our initial results on extracting the sentiments of user's posts. In order to achieve this, we utilized the Stanford Sentiment Analyzer (Socher et al., 2013). This analyzer classifies a text input into one of five sentiment categories—from *Very Positive* to *Very Negative*. We merge the five classes into three: *Positive*, *Neutral* and *Negative*. In future, we may consider specific (health and nutrition) sentiment lexicons.

We analyzed the sentiment of posts contributed by the users from the two clusters. As shown in Figure 2, posts of users belonging to the non-increasing cluster are more *neutral* whereas the posts made by users from the fluctuating cluster are mainly of *negative* sentiment. This gives an interesting intuition that the fluctuating group of users might require more emotional support as they use more negative sentiment in their posts.

3.3 Cohesion with Previous Posts

Cohesion is the property of a well-written document that links together sentences in the same context. Several existing models measure the cohesion of a given text with applications to topic segmentation or multi-document summarization (Elsner and Charniak, 2011; Barzilay and Lapata, 2005; Soricut and Marcu, 2006). In this analysis, we want to find out if there is any correlation between the cohesiveness of posts made by users and their pattern of weight loss. We are mainly interested in measuring the similarity of a user's post with respect to the previous posts in a thread. This can help identify users who elaborate on previous post versus those who shift the topic.

We focus on content words: verbs and nouns (part-of-speech tags *VB*, *VBZ*, *VBP*, *VBD*, *VBN*, *VBG*, *NN*, *NNP*, *NNPS*). Next, we use WordNet (Miller, 1995) to identify synonyms of the content words. Then, we compute similarity between the current post and previous posts of other users in the thread, in terms of commonly shared verbs and nouns including synonyms. In our current, preliminary analysis, we consider this similarity score to be the measure of cohesion.

In this step, we consider all posts that are not thread-initial. To approximate whether a post is cohesive, we compare the nouns and verbs of the current post to the list of nouns and verbs (plus synonyms) obtained from the previous posts of the thread. Our analysis finds that posts made by fluctuating users have an average cohesion score of **0.42** ($SE = 0.008$), whereas posts made by non-increasing users have an average cohesion score of **0.51** ($SE = 0.027$). This suggests that non-increasing users may be more focused when participating in forums whereas the fluctuating users are more prone to make posts that have less in common with the previous posts in a thread.

4 Conclusions and Future Work

In this paper, we analyze how the language changes based on the weight loss dynamics of users who participate in the forum of a popular weight-loss application. Specifically, this analysis revealed four interesting insights about the two types of users who lose weight in a non-increasing manner and who lose weight in a fluctuating manner. Firstly, fluctuating users are more active in participation compared to the other set of users. Secondly, fluctuating users post more question-

oriented posts compared to the non-increasing users. Thirdly, non-increasing users contribute posts that are more cohesive with respect to the previous posts in a given thread. Fourthly, posts contributed by fluctuating users have more negative sentiment compared to the posts made by non-increasing users. This observation hints that fluctuating users may need more emotional support to continue using this weight loss application and lose weight in an effective manner.

While this work is preliminary, our analyses provide a valuable early "proof of concept" for providing insights on how user behavior within online weight loss forums might impact weight outcomes. These sorts of analyses, particularly when replicated, could provide valuable insights for developing refined online weight loss forums that might facilitate more effective interactions for weight loss. It could also provide valuable insights for improving behavioral theories about behavior change (Hekler et al., 2013).

In the future, we plan to focus on a larger corpus from an extended time period, aligned more closely with weekly check-in weight data. Other directions for consideration are the temporal aspect of forum posts and gender-based analyses of user behavior.

Acknowledgments

References

Paul W. Ballantine and Rachel J. Stephenson. 2011. Help me, I'm fat! Social support in online weight loss networks. *Journal of Consumer Behaviour*, 10(6):332–337.

Antonina D. Bambina. 2007. *Online Social Support: The Interplay of Social Networks and Computer-Mediated Communication*. Cambria Press.

Regina Barzilay and Mirella Lapata. 2005. Modeling local coherence: An entity-based approach. In *Proceedings of the Association for Computational Linguistics*, ACL '05, pages 141–148.

Anita Das and Arild Faxvaag. 2014. What influences patient participation in an online forum for weight loss surgery? *Interactive Journal of Medical Research*, 3(1).

Micha Elsner and Eugene Charniak. 2011. Disentangling chat with local coherence models. In *Proceedings of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1179–1189.

Eric Hekler, Predrag Klasnja, Jon E. Froehlich, and Matthew P. Buman. 2013. Mind the theoretical gap: Interpreting, using, and developing behavioral theory in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Kevin O. Hwang, Allison J. Ottenbacher, Angela P. Green, M. Roseann Cannon-Diehl, Oneka Richardson, Elmer V. Bernstam, and Eric J. Thomas. 2010. Social support in an internet weight loss community. *I. J. Medical Informatics*, 79(1):5–13.

Fiona Johnson and Jane Wardle. 2011. The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss programme: a retrospective analysis. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1):1–7.

J L Kraschnewski, J Boan, J Esposito, N E Sherwood, E B Lehman, D K Kephart, and C N Sciamanna. 2010. Long-term weight loss maintenance in the united states. *International Journal of Obesity*, 34(11):1644–1654.

Catherine Penny Hinson Langford, Juanita Bowsher, Joseph P. Maloney, and Patricia P. Lillis. 1997. Social support: A conceptual analysis. *Journal of Advanced Nursing*, 25(1):145–151.

Tricia M. Leahey, Rajiv Kumar, Brad M. Weinberg, and Rena R. Wing. 2012. Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. *Obesity*, 20(7):1413–1418.

George A. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.

Aviva Must, Jennifer Spadano, Eugenie H. Coakley, Alison E. Field, Graham Colditz, and Dietz William H. 1999. The disease burden associated with overweight and obesity. *JAMA*, 282(16):1523–1529.

James M. Rippe, Suellyn Crossley, and Rhonda Ringer. 1998. Obesity as a chronic disease: Modern medical and lifestyle management. *Journal of the American Dietetic Association*, 98(10, Supplement):S9 – S15.

S. L. Saperstein, N. L. Atkinson, and R. S. Gold. 2007. The impact of internet use for weight loss. *Obesity Reviews*, 8(5):459–465.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the EMNLP*, pages 1631–1642, October.

Radu Soricut and Daniel Marcu. 2006. Discourse generation using utility-trained coherence models. In *Proceedings of the COLING/ACL on Main Conference Poster Sessions*, COLING-ACL '06, pages 803–810.

Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the NAACL HLT - Volume 1*, pages 173–180.

Rena R. Wing and Suzanne Phelan. 2005. Long-term weight loss maintenance. *The American Journal of Clinical Nutrition*, 82(suppl):222S–5S.