Quantified Self Meets Social Media: Sharing of Weight Updates on Twitter

Yafei Wang
The Pennsylvania State University
University Park
PA 16802, USA
yxw184@ist.psu.edu

Ingmar Weber and Prasenjit Mitra
Qatar Computing Research Institute, HBKU
Tornado Tower, P.O. 5825 Doha, Qatar
{iweber,pmitra}@qf.org.qa

ABSTRACT
An increasing number of people use wearables and other smart devices to quantify various health conditions, ranging from sleep patterns, to body weight, to heart rates. Of these “Quantified Selves” many choose to openly share their data via online social networks such as Twitter and Facebook. In this study, we use data for users who have chosen to connect their smart scales to Twitter, providing both a reliable time series of their body weight, as well as insights into their social surroundings and general online behavior. Concretely, we look at which social media features are predictive of physical status, such as body weight at the individual level, and activity patterns at the population level. We show that it is possible to predict an individual’s weight using their online social behaviors, such as their self-description and tweets. Weekly and monthly patterns of quantified-self behaviors are also discovered. These findings could contribute to building models to monitor public health and to have more customized personal training interventions.

While there are many studies using either quantified self or social media data in isolation, this is one of the few that combines the two data sources and, to the best of our knowledge, the only one that uses public data.

CCS Concepts
• Human-centered computing → Empirical studies in HCI; Empirical studies in collaborative and social computing; • Applied computing → Sociology;

Keywords
Quantified Self; Twitter; Smart Scales

1. INTRODUCTION

During the last couple of years, the number of users who use “Quantified Self” (QS) health tracking devices has continuously increased.1 A survey of 1,202 U.S. adult consumers conducted in December 2014 found that 31% use a QS tool to track their health and fitness.2 In lockstep with this proliferation of QS tools, research output on all related aspects has also seen a dramatic increase. Google Scholar lists 74 publications with “Quantified Self” in the title for all years up to and including 2013.3 However, since 2014 alone, already 123 publications matching this criteria have been indexed.4 Most of this research, however, looks at QS data in isolation, separately from other data one might obtain for a user.

In this paper, we present results from an attempt to combine QS data with social media data. This link between two data sources is made possible as more and more users publicly share the QS data they generate. For our study, we use data from users who opted in to connect their internet-enabled Withings Smart Body Analyzer5 to their Twitter account. Figure 1 shows an anonymised example tweet. We analyze data for 897 Twitter users who (i) not only have auto-generated fitness tweets from Withings, and apps such as Fitbit, or Nike, but (ii) also have “normal” tweets not generated by fitness apps.

Figure 1: An anonymised example of a self-quantified English tweet generated by Withings’ smart scale.

We are interested to see if a user’s weight can be inferred from his or her general social media behavior on Twitter. Do users with a larger body weight somehow tweet differently? If a link can be found, then this opens up new opportunities as it hints at the actual health information using social media data. In particular, we envision that social media could be used as one building block, together with QS data and electronic health records, to devise more personalized, holistic interventions that take a user’s lifestyle into account.6

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DigitalHealth ’16 April 11–13, 2016, Montreal, Canada
© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ISBN 978-1-4503-4224-7/16/04...$15.00
DOI: http://dx.doi.org/10.1145/2896338.2896363

1http://nuviun.com/digital-health/quantified-self
For example, a doctor could be provided with information about the personality of the individual from their tweets or the influences on the individual from their social circles that need to be taken into account (and, in some cases, overcome) while understanding and prescribing diet and exercise interventions.

Combining QS data and social media data also helps to overcome one frequent shortcoming of health-related studies: a lack of individual-level ground truth. While county-level health statistics, such as obesity rates, are readily available, it is much harder to obtain a set of Twitter users with a known weight, ideally also traceable over time. Despite the obvious limitations due to selection bias, users who link their QS data to their public social media account still provide a valuable data set.

At the same time, there are technical challenges that need to be addressed for a successful data fusion. Social media data is famously noisy due to internet lingo, spam and bots, and data incompleteness resulting from API limitations. The QS data also has its share of issues as users share their scale with friends or they might weigh themselves both before and after eating a large meal or going to the rest room. The combined data creates additional challenges due to its heterogenous nature: a textual stream (and more) from normal tweets, and a time series of weight data from QS tweets. We help address some of these issues by describing a method to remove implausible weigh-in data points.

Apart from technical challenges, the ethical challenges are at least as daunting. Though legally “public” data, tweets still often contain information that many would consider “private”. A tweeter expects his tweets to be viewed by his followers and the general public and not to be used and highlighted by data scientists around the world. Such concerns are amplified in the domain of medical data.

We believe that tackling these challenges is well worth the effort, especially as our initial results are promising. Concretely, we find that social media data can predict a user’s average weight with an R-squared around .25. We also find that the QS data collected through the Twitter stream is valuable by itself for population-level health analysis. For example, the result lets us paint a picture of weight transition across the year – yes, it goes up over Christmas and New Year – and of “dieting morale” across the week – users weigh themselves most often on weekends, which, ironically, is when they are least likely to do something about their weight. Given that we would not have been able to collect this data otherwise, this is a further advantage of looking at the intersection between QS and social media.

2. RELATED WORK

To the best of our knowledge, there is very little prior work that combines social media and QS data at the level of the individual. The vision described by Estrin [6] definitely includes a combination of data sources but no study on such data seems to have been performed to date. Vickey and Breslin [12] report a system-level study of how fitness app data is shared on Twitter, but they do not include a user-level study that links a user’s normal Twitter data with their fitness data. The StudentLife Project[8][14] uses a mobile phone app to collect detailed activity data, which is linked to academic performance. This data also includes Facebook profiles, though these are not part of the publicly shared data set.

Studies related to obesity and weight loss using social media typically take a population-level, public health approach. Culotta [5] used geo-tagged tweets, and Abbar, et al., [1] used food-related tweets to predict geographical differences in obesity and diabetes. Though including “normal life”, like food-related tweets, in their analysis, they use county-level data as “ground truth” for obesity. By using quantified self data, we can obtain weight-related information at the individual level.

A body of work also studies specialized social media, such as online weight loss forums [9]. Particular attention has been given to predicting weight loss from interactions in the social network. Chomutare, et al., [4] showed that high levels of activity in online obesity communities and being connected to several disparate sub-communities were both predictive of weight loss. They observed that the network structure properties were more useful in predicting weight loss than the biographical information associated with the users. Li, et al., [8] take this a step further by studying the problem of recommending a “good” friend within the context of a weight-loss app. Brindal, et al., [3] showed that the inclusion of a social networking platform did not have additive effects with respect to weight loss or retention. However, these inclusions resulted in patients using their weight loss system for a longer duration. However, in their experiments, greater use of a weight tracker tool resulted in greater loss. Though our work only looks at the combination of social media and QS for data collection, their work provides evidence for benefits for health interventions as well.

3. DATA COLLECTION

To collect our data, we use the Twitter Streaming API for three weeks in Oct 2015 collecting tweets containing keywords, “lb” or “kg”. Note that this broad pattern captures data both weight-related QS tweets and other tweets in several languages. We also use the Topsy API to gather all obtainable historical tweets containing these keywords. These tweets were then filtered such that only tweets generated by “WiTwit” were kept.

For each unique user who generated these tweets, we obtained (i) (up to) 3,200 of their most recent tweets, (ii) self-generated profile known as “bio”, (iii) lists of their friends and followers, and (iv) the bios of their friends and followers.

A large fraction of users in our dataset only tweeted their weight or other automatic fitness tweets. These specially created accounts, potentially as a sort of personal fitness log, were not of interest for us as, apart from a time series of weigh-ins, there was no other social media data to better understand the user. We therefore imposed an additional filter by requiring (i) each user to have at least 10 “normal”

---

[1] We even observed an instance where, apparently, the weight of a cat was recorded.

[2] One of our data source, Topsy, is no longer supported since Dec. 2015.

[3] https://dev.twitter.com/streaming/overview


[6] WiTwit” is the “source” field used in Tweets generated by WiThings’ smart scale.
Table 1: List of patterns identifying automatic fitness tweets in the “source” field of a tweet’s JSON file.

<table>
<thead>
<tr>
<th>Type</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Weight Loss</td>
<td>WiTwit</td>
</tr>
<tr>
<td>Other Weight Loss</td>
<td>Lose It!, SimpleWeight</td>
</tr>
<tr>
<td>Fitness</td>
<td>RunKeeper, Fitbit, Nike, Runmeter, Runtastic, Nike+ GPS, iSmoothRun</td>
</tr>
</tbody>
</table>

Table 2: A Support Vector Machine model with linear kernel for predicting a person’s average weight using their tweets and self-description. The top-15 features for each direction are shown here.

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Coef.</th>
<th>Feature</th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio_LIWC_pronoun</td>
<td>90.71</td>
<td>Tweet_LIWC_social</td>
<td>-15.10</td>
</tr>
<tr>
<td>Bio_LIWC_auxverb</td>
<td>89.57</td>
<td>Tweet_LIWC_present</td>
<td>-18.10</td>
</tr>
<tr>
<td>Bio_LIWC_verb</td>
<td>86.24</td>
<td>Tweet_LIWC_past</td>
<td>-19.55</td>
</tr>
<tr>
<td>Bio_LIWC_english</td>
<td>79.47</td>
<td>Tweet_LIWC_article</td>
<td>-20.01</td>
</tr>
<tr>
<td>Bio_LIWC_english</td>
<td>76.12</td>
<td>Tweet_LIWC_bio</td>
<td>-13.48</td>
</tr>
<tr>
<td>Bio_LIWC_verb</td>
<td>76.09</td>
<td>Tweet_LIWC_english</td>
<td>-15.56</td>
</tr>
<tr>
<td>Tweet_LIWC_english</td>
<td>75.62</td>
<td>Tweet_LIWC_verb</td>
<td>-15.50</td>
</tr>
<tr>
<td>Bio_LIWC_incl</td>
<td>72.23</td>
<td>Tweet_LIWC_english</td>
<td>-15.50</td>
</tr>
<tr>
<td>Bio_LIWC_cogmech</td>
<td>71.20</td>
<td>Tweet_LIWC_incl</td>
<td>-15.26</td>
</tr>
<tr>
<td>Bio_LIWC_article</td>
<td>69.83</td>
<td>Tweet_LIWC_cogmech</td>
<td>-13.91</td>
</tr>
<tr>
<td>Bio_LIWC_conj</td>
<td>66.49</td>
<td>Tweet_LIWC_article</td>
<td>-13.48</td>
</tr>
<tr>
<td>Tweet_LIWC_bio</td>
<td>65.97</td>
<td>Tweet_LIWC_conj</td>
<td>-11.00</td>
</tr>
</tbody>
</table>

In order to quantify and interpret the effects of different w(i) to w(i+1) [in pounds] between days d(i) and d(i+1) is “plausible” if |w(i)−w(i+1)| ≤ 4+d(i)−d(i+1). In words, we allow for up to 4lb of weight fluctuation within one day and 1lb for each day passed. Note that 1lb of body fat is roughly equivalent to 3,500kcal. Though larger fluctuations are possible, especially due to excessively storing or losing of liquid, we decided to err on the side of caution, rather than including too much erroneous data. Users with more than three plausibility violations were excluded. Some users reported suspiciously low or high some weights such as 12 lbs or 400 lbs. Users whose average weight is either smaller than 100 lbs or larger than 300 are treated as outliers and excluded from our analysis. As a point of reference, the average of the individual average weights was 178.4lb. Note that this is very close to the 2012 average weight of adults in North America of 177.9lb, indicating that our data may not be as odd and biased as one may imagine. After applying the data filtering explained above, we use the remaining 391 users to build a prediction model.

In order to capture and summarize a user’s social media content, we utilize two existing dictionaries that have undergone psychometric validation. The first is the Linguistic Inquiry and Word Count (LIWC) dictionary, with 64 categories, and the second is the PERMA dictionary with ten categories. Both dictionaries map terms to a set of categories such as “social”, “health” and “body” in the case of LIWC, or “positive emotion”, “engagement” and “meaning” in the case of PERMA. For example, PERMA maps the term “distract” to “negative emotion”; and LIWC maps the term “brother” to “social”. We applied this mapping both to a user’s normal tweets and their bio. For boosting the model performance, we also add Bag-of-Word features.

In order to quantify and interpret the effects of different LIWC [10], or “positive emotion”, “engagement” and “meaning” in the case of PERMA. For example, PERMA maps the term “distract” to “negative emotion”; and LIWC maps the term “brother” to “social”. We applied this mapping both to a user’s normal tweets and their bio. For boosting the model performance, we also add Bag-of-Word features.

In order to quantify and interpret the effects of different LIWC [10], or “positive emotion”, “engagement” and “meaning” in the case of PERMA. For example, PERMA maps the term “distract” to “negative emotion”; and LIWC maps the term “brother” to “social”. We applied this mapping both to a user’s normal tweets and their bio. For boosting the model performance, we also add Bag-of-Word features.

In order to quantify and interpret the effects of different LIWC [10], or “positive emotion”, “engagement” and “meaning” in the case of PERMA. For example, PERMA maps the term “distract” to “negative emotion”; and LIWC maps the term “brother” to “social”. We applied this mapping both to a user’s normal tweets and their bio. For boosting the model performance, we also add Bag-of-Word features.

In order to quantify and interpret the effects of different LIWC [10], or “positive emotion”, “engagement” and “meaning” in the case of PERMA. For example, PERMA maps the term “distract” to “negative emotion”; and LIWC maps the term “brother” to “social”. We applied this mapping both to a user’s normal tweets and their bio. For boosting the model performance, we also add Bag-of-Word features.

In order to quantify and interpret the effects of different LIWC [10], or “positive emotion”, “engagement” and “meaning” in the case of PERMA. For example, PERMA maps the term “distract” to “negative emotion”; and LIWC maps the term “brother” to “social”. We applied this mapping both to a user’s normal tweets and their bio. For boosting the model performance, we also add Bag-of-Word features.
Table 3: Results of predicting a person’s average weight (in lb) using social media information.

<table>
<thead>
<tr>
<th>Tweet Only</th>
<th>R</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Baseline</td>
<td>-0.14</td>
<td>27.89</td>
<td>34.67</td>
</tr>
<tr>
<td>Language Split</td>
<td>0.23</td>
<td>23.73</td>
<td>30.06</td>
</tr>
<tr>
<td>Gaussian Process</td>
<td>0.50</td>
<td>23.67</td>
<td>29.57</td>
</tr>
<tr>
<td>Gaussian Process + BoW</td>
<td>0.45</td>
<td>20.00</td>
<td>28.04</td>
</tr>
<tr>
<td>SVM (Linear Kernel)</td>
<td>0.34</td>
<td>26.65</td>
<td>33.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tweet + Bio</th>
<th>R</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Split</td>
<td>0.07</td>
<td>40.99</td>
<td>62.32</td>
</tr>
<tr>
<td>Gaussian Process</td>
<td>0.38</td>
<td>23.84</td>
<td>30.26</td>
</tr>
<tr>
<td>Gaussian Process + BoW</td>
<td>0.55</td>
<td>22.96</td>
<td>28.99</td>
</tr>
<tr>
<td>SVM (Linear Kernel)</td>
<td>0.34</td>
<td>26.81</td>
<td>33.47</td>
</tr>
</tbody>
</table>

Table 4: Quantified-Self and Google search behavior on different weekdays.

<table>
<thead>
<tr>
<th>Month</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
<th>Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Withings weigh-ins</td>
<td>54k</td>
<td>54k</td>
<td>54k</td>
<td>53k</td>
<td>47k</td>
<td>63k</td>
<td>52k</td>
</tr>
<tr>
<td>Fitness tweets</td>
<td>9.4k</td>
<td>9.6k</td>
<td>9.3k</td>
<td>9.2k</td>
<td>8.9k</td>
<td>8.8k</td>
<td>9.0k</td>
</tr>
<tr>
<td>“bmi”</td>
<td>27.1</td>
<td>28.5</td>
<td>28.2</td>
<td>27.8</td>
<td>25.0</td>
<td>21.6</td>
<td>22.8</td>
</tr>
<tr>
<td>“Weight loss”</td>
<td>36.1</td>
<td>35.9</td>
<td>34.8</td>
<td>33.4</td>
<td>31.5</td>
<td>31.5</td>
<td>34.1</td>
</tr>
<tr>
<td>“diet”</td>
<td>92.0</td>
<td>90.5</td>
<td>87.7</td>
<td>85.0</td>
<td>78.5</td>
<td>78.3</td>
<td>86.9</td>
</tr>
</tbody>
</table>

Figure 2: Weight changes for different months, relative to a user’s mean weight.

5.1 Trends in QS Behavior Across Days-of-Week

Table 4 reports the frequency of automatically generated weigh-in and fitness tweets in our dataset. For comparison, we also show information for Google search volumes\(^{19}\) for the three queries “BMI”, “weight loss” and “diet”, summarized for days-of-the-week from September 2015 to November 2015.

There are clear weekly patterns detectable for both the QS and the Google Trends data. Put simply, users are most aware of their weight on Saturdays with, by far, the largest number of weigh-ins. However, this is also the day where they are least likely to “take action” by (i) generating fitness tweets, or (ii) searching for diet-related information.

By contrast, such actions seem to be most likely on Mondays. Consistent observations were made by Weber and Achananuparp [15] who observed that the number of users logging their meals with MyFitnessPal is highest (lowest) on Mondays (Saturdays). Of those users that log their meals, the fraction consuming more than their self-set calorie goals is also lowest (highest) on Mondays (Saturdays). Based on this observation, Saturday is everyone’s “weigh-in day” but not “fitness day”.

5.2 Trends in QS Behavior Across Months-of-Year

In this subsection, we look at actual weight changes across a year. For each person, we compare her average weight within a given month to her global average weight. For each month, these weight changes are then averaged across all users. Mean and error bar of weight changes for different months are shown in Figure 2. We see that users gain weight during winter months, from a low in October to a high in January, before starting to lose weight again. Though the observed jump of just under 1lb during the holiday seasons might appear lower than intuition would suggest, this value


\(^{19}\)https://www.google.com/trends/explore
that such a data fusion between messy, general life style patterns, such as a weight increase over Christmas. We believe as a peak of weigh-in activity on Saturday, and monthly patient activity. Our data also capture weekly patterns, such as weight or not" can be inferred from their Twitter profile picture.

Additionaly, we wanted to see if there is a link between QS data and topical interest as observed through Google Trends. Figure 3 presents the mean and the error bar of Google search scores aggregated by month from 2005 to 2015. For all three of our search terms, search activity is highest in January, possibly due to New Year’s resolutions. Overall, the search volume changes more abruptly from December to January than the actual weight (see Figure 2). So whereas users slowly put on pounds from October to January, there appears to be a sudden change in weight loss intent from December to January – assuming that the selected Google search terms do indeed measure “weight loss intent”.

6. DISCUSSION AND LIMITATIONS

The data analyzed for this study does not come from a randomized trial or from a representative sample of the population. Users who choose to publicly tweet their weight are likely to differ from a “normal” user trying to lose weight, though (i) our population’s average weight and (ii) the weight gain over Christmas were surprisingly close to known values. Weber and Mejova [16] show that, with a certain amount of noise, a user’s body weight or at least classes such as “overweight or not” can be inferred from their Twitter profile pictures. We are not relying on such noisy labels but, basically, we are trading a loss in recall for an increase in precision.

A considerable fraction of users, 198 out of 391, had chosen Japanese as their interface language and, correspondingly, many of their tweets were not be in English. Google’s automatic translation might introduce errors, though for tasks such as sentiment analysis machine translation typically performs sufficiently well [2].

7. CONCLUSIONS

In this paper, we presented a study that combines quantified self data from internet-enabled smart scales with general social media data on Twitter. We used this combination of data sources to predict a user’s weight using only their social media activity. Our data also capture weekly patterns, such as a peak of weigh-in activity on Saturday, and monthly patterns, such as a weight increase over Christmas. We believe that such a data fusion between messy, general life style social media data and very accurate, longitudinal quantified self data has great potential to improve personalized health care.

8. REFERENCES


---

Figure 3: Google Trends search scores for different months from 2005 to 2015.