Predicting Weight Goal Changes in a Self-Tracking App CS230 Project Final Report, Spring 2018

Category - Healthcare

Mitchell Gordon (mgord)
Sharon Zhou (sharonz)
Shushman Choudhury (shushman)

Abstract

Hundreds of millions of people use self-tracking apps to manage their health. A key feature in most of these apps is the ability to set goals. While the psychological theory and effectiveness of goals have been studied extensively in psychological and biomedical literature, there is extremely limited understanding of how goals are used in today's self-tracking apps in practice. In this paper, we present the first large-scale model of the weight loss goals that people set in practice. Specifically, we ask: given someone changes their weight loss goal, do they make their new goal harder or easier compared to their initial goal? Our primary input set is a sparse timeseries of self-logged weights, as well as demographic data and the initial goal itself. We experiment with two methods: LSTM using an interpolated timeseries of weights logged, and fully connected feed-forward with hyperparameter search using hand-picked features. We find that the feed-forward model was most effective, with a test accuracy of nearly 74%.

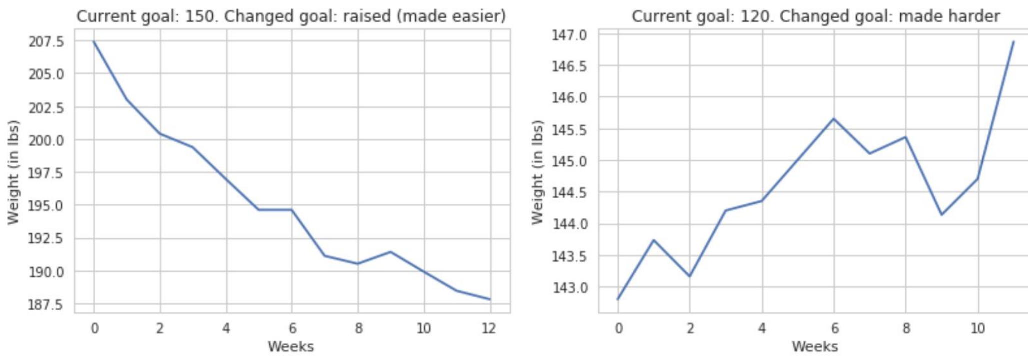
1 Introduction

Hundreds of millions of people use activity tracking applications and devices to manage their health. Core to most of these trackers is the ability to set and to change health and dietary goals. While goal-setting theory is well-studied in psychological and medical literature [4], current understanding about how people use health goals in practice in today's self-tracking apps, including how people choose to change their health goals, is extremely limited. We present a large-scale study of hundreds of thousands of weight loss goal changes in MyFitnessPal. Studying this popular self-tracking application allows for an unprecedented opportunity to create detailed models that can predict and understand real-world human behavior at scale. We study anonymized data about 251,370 people who log their weight over time and set weight loss goals in the MyFitnessPal smartphone app.

Specifically, we focus on what happens when someone chooses to change their weight loss goal. Along people's weight loss journeys, people set different weight goals depending on their progress. Predicting what new goals people set is important because it is an essential first step to constructing and timing interventions to help people with weight loss. In this paper, we construct a deep network that predicts, given someone changes their weight goal, will they make their goal easier or harder. We predict this behavior based on their demographics and past weight and behavior. We leverage the largest dataset of weight logs from MyFitnessPal to train this model. A pair of examples that illustrate the challenges with this problem is depicted in Figure 1.

Unlike many deep learning papers that aim to learn biological or natural processes, in this paper in are aiming to model human behavior. A core challenge of predicting and modeling human behavior in comparison to other domains is that it is highly variable and dependant upon an enormous number of

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.



(a) (b)

Figure 1: Two illustrative examples of weight loss timeseries (with the interpolation described above). (a) The user was making limited progress on their goal, but not as fast as they might have wanted. This resulted in them making their goal easier, which we would expect. (b) Despite doing well on their goal, this user in this chunk ends up making their goals harder, as opposed to easier (raised their goal further), which is counterintuitive. We would have expected them to make their goal easier in this case.

psychological and contextual factors in a person's life, which are often impossible to directly measure. However, by intelligently selecting features, we are able to capture both direct health measurements as well as features that may indirectly indicate one's psychological state or life context. Another challenge is that this dataset is entirely self-logged, meaning it is very sparse and there are significant variations in the data. We discuss this more later on.

The input to our algorithm includes the following features: initial weight goal, initial weight, final weight, number of weight logs during this interval, duration of this interval, and demographic data (gender, location, age). We then used a feedforward neural network to output the whether the next weight loss goal was raised (made easier) or lowered (made harder) from the previous goal (binary classification, 0 or 1). We find that we are able to predict whether someone makes their weight loss goal harder or easier with over 73% test accuracy.

2 Related Work

The most relevant paper is from the 2017 WWW conference, where researchers applied CNNs and RNNs to self-logging data in order to infer individual traits related to wellness (BMI, alcohol intake), lifestyle (food logging, scale use), and behavior (compliance with medical interventions, amount of activity on a fitness app) [6]. This work saw variable results (53%-85% accuracy), but established that self-logging data, previously seen as unhelpful, is actually useful with the help of neural networks. Motivated by this work, we examine self-logging data to predict how individuals change their weight loss goals in light of their past behavior.

Some researchers have tried to predict related human behaviors to our task, such as predicting future self-logged actions from a small set of options (e.g. exercise, drink water) [3], predicting social interactions [5], and predicting process behavior [2]. However, these works used either significantly less data without deep learning or did not examine self-logging data. Finally, most researchers in this area have used quantitative methods to analyze food data, but at much smaller scale and without deep learning [1, 7, 8].

3 Dataset Description

We use a **proprietary dataset from MyFitnessPal** containing self-logged weight entries and goals from 1.7 million users. This is largest dataset of weight logs studied to date. The dataset includes the following tables:

1. *Users*: includes information about users, including demographic data and initial goal weights.
2. *Weight Goals*: includes logs of weight goals across all users, from which we can extract chunks of weight goals to preprocess for input and output.
3. *Weight Logs*: includes logs of weights across all users, from which we can draw time series data on weight logs.

4 Methods

We experimented with two difference models: feed-forward with hand-picked features, and an LSTM using an interpolated timeseries of weight logs. We find that the feed-forward model was the most effective, for reasons we discuss below.

4.1 Feed-forward neural network

We used a fully connected neural network with 5 hidden layers on a subset (50,000 rows, approximately 15%) of our data. Our first hidden layer has 256 nodes with Tanh activation, our second has 1024 nodes with softmax activation, our third has 512 nodes with Tanh activation, and our fourth has 512 with ReLU, and a final sigmoid activation layer. We used dropout (0.5, after trying 0.1 and 0.2) and batchnorm for all hidden layers. We use binary cross entropy (BCE) as a loss function and Adam as the optimizer. BCE specifically is used for binary classification, as it strongly penalizes examples that are classified incorrectly.

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log(\hat{y}_i + (1 - y_i) \cdot log(1 - \hat{y}_i) \tag{1}$$

These hyperparameters (number of hidden units in each layer, activation function, and optimization function) were found in a hyperparameter search across {256, 512, 1024} hidden units, {ReLU, Tanh, Softmax} activation functions, and {RMSprop, Adam, Adadelta} optimization functions. We trained on 100 epochs and set aside 10% of the data for validation. We reached an accuracy of 73.15% on our training set and 73.75% on our test set.

4.1.1 Features

We hand-picked features that aim to capture both health and psychological states for each user. Figure 3 shows the features we used in this model. The features are:

1. Initial goal: a user's first goal weight goal, normalized by their weight.
2. Initial weight: the first logged weight of their initial goal.
3. Last weight: the last logged weight of their initial goal.
4. Duration: the number of days they kept their initial goal before changing it.
5. Demographics: age and gender (as a one-hot encoding).

4.2 RNN

We also implemented an RNN with 3 hidden layers. The first hidden layer was an LSTM with 500 hidden units and Tanh activation, two fully connected layers with 512 units and ReLU activation, and a final sigmoid activation layer. We applied dropout to each layer (0.2, after trying 0.1 and .5). We use binary cross entropy (BCE) as a loss function and Adam as the optimizer.

5 Results and Discussion

As mentioned earlier, we tried to use an LSTM that took as input the time series data of the weight logs. We divided the weight goal period into weekly intervals. For most datapoints, the users did not log their weights each week, so we had to do interpolate their weight logs on a weekly scale. We tried

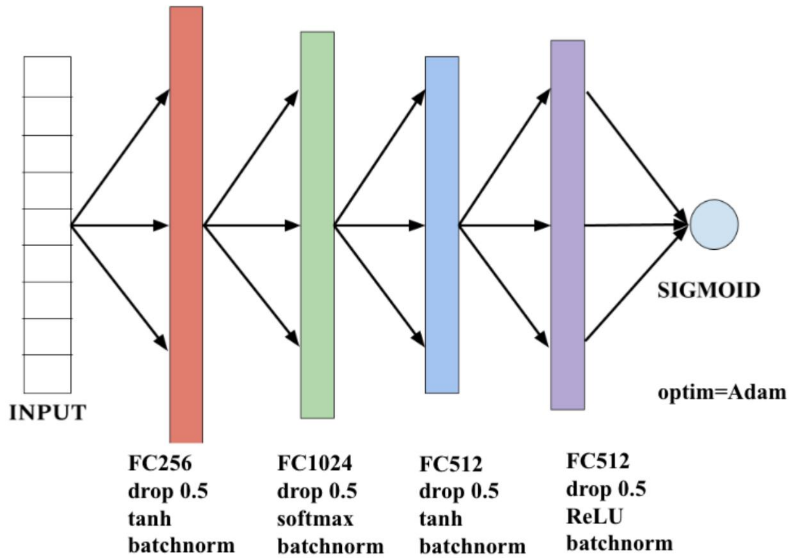


Figure 2: The final feed-forward model that we used for our project.

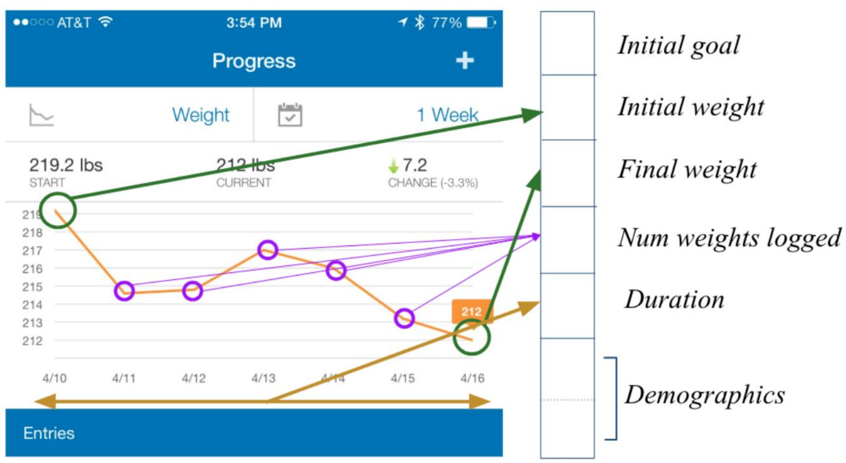


Figure 3: A schematic of the features that we used for our feed-forward network. We select an effective subset of the dataset features which capture the relevant aspects of weight loss while avoiding the sparsity or irregularity of the weight log time series data which we initially tried with the LSTM.

a few LSTM models but unfortunately the accuracy was not sufficiently higher than random guessing to encourage futher exploration. The most probable cause of this is that the interpolation is too noisy. Most users log weights infrequently and at irregular intervals, which probably does not allow for a reliable time-series of data. We also did linear interpolation of weights but it is unclear what the appropriate method to do that is. There is also a high variability in the duration of weight goal periods, between 4 to 53 weeks. Though LSTMs are capable of supporting variable length time-series input data, we think the high variability of the dimensionality is another confounding factor here.

The results for our feedforward network with our selected features are stated in Table 1. The final number of datapoints that satisfy our numerous filtering criteria is quite small, but our accuracy is reasonable considering we are trying to predict highly uncertain contextual human behaviour with very noisy data.

Train Size	Train Accuracy	Test Size	Test Accuracy
226,233	73.15%	25,137	73.75%

Table 1: Feedforward performance details

6 Conclusion/Future Work

We designed a feed-forward model to predict the nature of changes that people make to their weight goals as they progress along their weight loss journeys. A significant portion of our work involved deciding numerous filtering criteria to have relatively less noisy data to train our model with, processing the data records of the MyFitnessPal app to create a dataset in a format that was compatible with deep learning libraries, and then creating the features for our feed-forward network to use.

This course project gave us a number of broader insights beyond just the experience of working on a real-world problem. Deep learning is known to work quite well for natural perception data and in structured settings like web applications where there is a lot of data available. In our case, though our source of data is an mobile app, we are essentially predicting human behaviour, which is an area where the success of deep learning has not been as significant. Furthermore, there were some underlying challenges in our dataset, in that how someone changes their weight goal is not just a function of their progress and broad demographics but also their personality and behaviour to some extent. The user data that we have from the app is perhaps not expressive enough to capture this aspect of users, which was also a challenge.

There are two particularly interesting directions of future work. The first would be to figure out how to use RNNs to do well on this task, because it clearly has a temporal element. This might be done by curating the data better or by trying to use more sophisticated models. The other direction would be to attempt more complicated tasks like predicting when the user would change their weight goal and by how much.

7 Contributions

- **Sharon Zhou** - Led model formulation, implementation, training/GPUs, hyperparameter search; helped with project formulation, data understanding, data filtering and pre-processing, poster, and write-ups. PM'd the team for TA involvement and scheduling time to work.
- **Mitchell Gordon** - Led project formulation, data understanding, data filtering and pre-processing (acquired dataset through his lab); discovered the most effective model layers which we use for our final feed-forward network without hyperparameters, implementation, poster, and write-ups.
- **Shushman Choudhury** - Led the poster; helped with project formulation, data understanding, data filtering and pre-processing, model formulation, and write-ups.

References

[1] Yong-Yeol Ahn, Sebastian E Ahnert, James P Bagrow, and Albert-László Barabási. Flavor network and the principles of food pairing. *Scientific reports*, 1:196, 2011.

[2] Joerg Evermann, Jana-Rebecca Rehse, and Peter Fettke. Predicting process behaviour using deep learning. *Decision Support Systems*, 100:129–140, 2017.

[3] Takeshi Kurashima, Tim Althoff, and Jure Leskovec. Modeling interdependent and periodic real-world action sequences. *arXiv preprint arXiv:1802.09148*, 2018.

[4] Edwin A Locke and Gary P Latham. *A theory of goal setting & task performance.* Prentice-Hall, Inc, 1990.

[5] Nhathai Phan, Dejing Dou, Brigitte Piniewski, and David Kil. A deep learning approach for human behavior prediction with explanations in health social networks: social restricted boltzmann machine (srbm+). *Social Network Analysis and Mining*, 6(1):79, 2016.

[6] Tom Quisel, Luca Foschini, Alessio Signorini, and David C Kale. Collecting and analyzing millions of mhealth data streams. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1971–1980. ACM, 2017.

[7] Charles Spence. Breakfast: The most important meal of the day? *International journal of gastronomy and food science*, 8:1–6, 2017.

[8] Ingmar Weber and Palakorn Achananuparp. Insights from machine-learned diet success prediction. In *Biocomputing 2016: Proceedings of the Pacific Symposium*, pages 540–551. World Scientific, 2016.