Predicting Weight Goal Changes in a Self-Tracking App

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

Overview

We design a deep learning approach to predict the direction of change in weight goals as people progress along their weight loss journey. **Our model predicts: given someone changes their weight loss goal, did they make it harder or easier?**

Challenges

- Human behavior is dynamic and highly context-specific.
- Self-logged weight data is sparse and potentially unreliable.
- Highly variable frequency of logging makes weight interpolation very challenging

Input

Initial weight goal, first and last weight of initial goal, demographics (age, gender, initial BMI), length kept goal, number of times logged weight.

Output

0/1 - Whether next goal is above or below current goal

Dataset

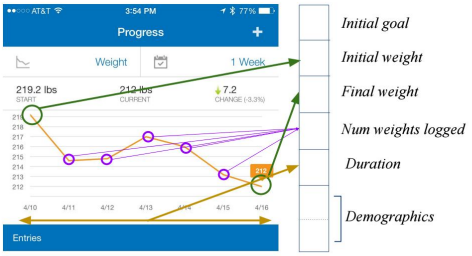
A proprietary dataset from MyFitnessPal, with 1.7M users (largest self-tracking dataset ever studied). We used data from the following tables:

- *Users* - Demographic data
- *Weight Goals* - Timestamped weight goals for users
- *Weight Logs* - Irregular logs of user weights

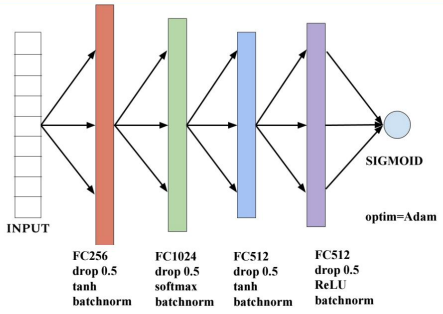
Data Filtering

- Initial goal kept for at least 1 month and at most 1 year
- Goal changed by at least 1 lb
- The objective was to lose weight
- Unreasonable outliers from self-logged data removed

Features



Model



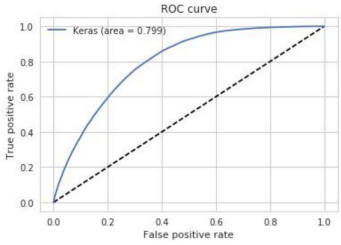
Cost function (BCE = binary cross-entropy):

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

Hyperparameter Search

- Fully connected (FC) layers, number of nodes: 256, 512, 1024
- Activation after each FC layer: ReLU, Tanh, Softmax
- Optimizer: RMSprop, Adam, and Adadelta

Results/Discussion



Training accuracy	Training size	Test accuracy	Test size
73.15%	226,233	73.75%	25,137

Deep learning is known to work well for natural perception data and structured outputs, but is rarely used to predict real-world human behavior.

Underlying data challenges - user personalities and habits affect how they will change their weight goals, but the user data is not sufficient to capture these factors

We also tried an LSTM model over the entire time-series of weights logged, re-sampling on a weekly basis and interpolating missing weeks. This performed poorly, perhaps due to the sparsity and inconsistency of self-logged weights.

Future Work

- Investigate RNN models for the timed logs.
- More complicated tasks - *when* someone might change their goals and by *how much*.