STATS 710 – Sequential Decision Making with mHealth Applications

Note: A Reinforcement Learning System to Encourage Physical Activity in Diabetes Patients

Instructors: Susan Murphy and Ambuj Tewari *Scribe: Anurag Beniwal*

1 Introduction

The aim of the study is to assess effectiveness of automatically-tailored personalized feedback in increasing adherence in diabetic patients(only type 2 diabetes) to a personal physical activity regimen recommended by a diabetes specialist.The secondary objective is to improve glycemic control.

2 Design of study

It was a 26 weeks long study on 27 sedentary diabetes type 2 patients with smart phone and a Wifi connection (with a pedometer) and a personal plan for physical activity.

Inclusion criterion : HbA1c greater than 6.5% , sedentary lifestyle. People with other diseases that prohibit exercise were excluded from the study the patients were randomly assigned to the control and the test(personalized) groups.

3 Learning algorithm

The authors use a contextual bandit type of an algorithm.

3.1 Actions:

A message is sent daily and a summary message is sent every week.

- **Daily messages**

1. **Negative Feedback**: "You need to exercise to reach your activity goals. Please remember to exercise tomorrow."
2. **Positive feedback relative to self**:"You have so far achieved N% of your weekly activity goal. Your exercise level is in accordance with your plan . Keep up the good work."
3. **Positive feedback relative to others:**"You have so far achieved N% of your weekly activity goal . you are exercising more than the average person in your group. Keep up the good work."
4. **No Message**

- **Weekly summary messages**

  On most weeks the weekly summary message is ”Please remember to exercise this week to reach your exercise goals”. When patients achieve a significant exercise level and not more than once per 3 weeks they could receive the following messages.

1. **Maximal Increase:** Over the past week you increased your activity more than any previous week.

2. **Significant increase :**over the past week you increased your activity more than any previous week

3. **Maximal Social :** ”Last week you increased your physical activity more than any other participant in the experiment.

4. **Significant social :**”Last week you increased your activity more than any other participant in the experiment.

3.2 Initialization of algorithm:

1. On 20% of days no message was sent

2. A uniform random number between 0 and 1 is drawn .

3.
```
if the number is ¿ expected fraction of weekly activity at that day then
  the user will get the negative feedback message.
else
  one of the positive messages is sent with equal probability
end if
```

3.3 Features/Elements of context vector $x_{i,t}$:

1. **Activity Attributes :**

   - Number of minutes of activity in the last day
   - Cumulative number of minutes of activity this week
   - Fraction of activity goal
   - Fraction versus expected as this point in the week

2. **Demographics:** Age, Gender

3. **Feedback attributes:** Number of days since each feedback message was sent

4. **Interaction Features:** Interaction features were used . Example: Interaction between activity performed so far and time since each feedback message was sent , fraction of activity performed so far and time since each feedback message was given, Daily activity a day before feedback and activity performed so far ,Daily activity a day before feedback and time since each feedback was given etc.

3.4 Learning:

Let $x_{i,t}$ be the context vector for person i at time t and let $y_{i,t}$ denote the change in activity from day t to day t+1. A linear regression model is trained on all $x_{i,t}, y_{i,t}$ pairs to predict $y_{i,t}$ from $x_{i,t}$

$$Y_{i,t} = X_{i,t}\hat{\beta} + \epsilon \tag{1}$$

The learning algorithm is run everyday and the most up-to-date model is used for prediction
To predict appropriate action to be taken from the predicted response $\hat{y}_{i,t}$ Boltzmann sampling is performed with $T_{Boltzmann} = 5$ and the action is probabilistically selected with probability for selection of k^{th} arm/message

$$p(y_{i,t,k}) = \frac{exp(y_{i,t,k})}{\sum_k exp(y_{i,t,k})} \tag{2}$$

The use of boltzmann function induces exploration by itself as an action/message with predicted probability less than that of best arm could also be selected

Note 1: Their was single model for all users and data from all users was used to update the model
Note2: Inverse probability weighting was not used and could be potentially helpful to reduce the bias in estimates

4 Results and Insights

- **Effect of Different Messages over time** : It was found that inclusion of feature "time since each feedback was sent " provides historical context to the policy, allowing feedback to be dependent between days . Results showed that the time dependent feedback was more useful than the time independent one . For example, even though on average negative feedback reduced activity but it could still induce positive change if given before a positive feedback.
- **Variability in patient response**: K means clustering of users was done and response for each message type was measured for each of the clusters . It was found that different clusters exhibited different change in activity levels as a response to messages . It was found that females responded less to messages in general as compared to males whereas response did not differ significantly with age. This shows that demographic variables play an important part in personalizing actions.
- **Learning process over time**: It was found that learning algorithm improved over time and the coefficients became more stable over time. The R^2 value improved over time finally settling to 0.43 . There were few fluctuations in R^2 due to extreme weather events , which makes it extremely important to include external factors like weather forecast in features.
- **Effectiveness of the algorithm**: A linear regression was fit to assess the change in activity levels for each users separately and it was found that the average slope and standard error of slope for the treatment group were higher and lower respectively than both control group (No messages sent) and for the group to which messages were sent using the initial policy.This means that the algorithm lead to increase in activity levels in general.The rate of walking also improved for the treatment group

This note is a summary of [HFK$^+$16]

References

[HFK$^+$16] Irit Hochberg, Guy Feraru, Mark Kozdoba, Shie Mannor, Moshe Tennenholtz, and Elad Yom-Tov. A reinforcement learning system to encourage physical activity in diabetes patients. *arXiv preprint arXiv:1605.04070*, 2016.