Discovering Persuasion Profiles Using Time Series Data

Igor Kulev
EPFL / IC / IIF / LIA
Batiment IN, Station 14
CH - 1015 Lausanne (Switzerland)

Pearl Pu
EPFL / IC / GR-SCI-IC / GR-PU
Batiment BC, Station 14
CH - 1015 Lausanne (Switzerland)

Boi Faltings
EPFL / IC / IIF / LIA
Batiment IN, Station 14
CH - 1015 Lausanne (Switzerland)

Editor: Leslie Pack Kaelbling

Abstract

Recommender systems have significantly changed the way people look for products online. Instead of formulating a query, collaborative filtering technology is capable of suggesting relevant movies, books, and other items to a user without explicit effort. Many researchers are interested in applying similar personalization techniques to the next wave of recommender systems, those that recommend diets and physical activity improvements. However, there is more at stake here. If handled improperly, the system risks suggesting behavior changes (or called interventions) that can harm a user. For example, if someone is not used to walking 10K steps a day, a recommendation of 8000 steps more may lead to injuries. In this research work, we are interested in learning the effect of interventions that were designed and administered to a population of users. Similar to traditional recommender systems where existing users’ preferences were analyzed and synthesized into patterns of “taste”, we would like to know how a given intervention changes the users’ current behavior habits. For example, an intervention may persuade some users to increase their level of activeness (we call them responders), cause some users to do the opposite (non-responders), and some to improve their activeness only temporarily (temporarily responders). The effect of the intervention is thus regarded the persuasiveness of the intervention. Our research aim is to discover the persuasion profile of a given intervention. Previous methods on intervention related research have focused on prediction, for example, predicting how users will change after an intervention (growth model), or predicting the event that has caused the change (change point analysis). Our focus is on designing effective recommender systems. To demonstrate how our system works, we use a dataset of 45 users over the period of 10 days. Their calorie expenditure was measured at the rate of once every minute, which poses a challenge to our modeling and clustering techniques. This work is in its early stage. Even though the conclusions are interesting, our future work includes applying the same technique to larger datasets.

Keywords: Probabilistic Graphical Models, Mixture Models, Time Series Clustering, Behavior Change

©2016 Igor Kulev, Pearl Pu and Boi Faltings.
1. Introduction

With the advance of technology in fitness trackers and their increasing popularity, the amount of physical activity data increases rapidly. This data, e.g. steps, distance, elevation, and energy expenditure, presents an opportunity for us to gain a better insight into people’s health and behavior patterns.\footnote{Data privacy is an extremely important issue, but its beyond the scope of this paper.} More importantly, it can help us understand how effective an intervention is and whether it leads to desired outcome. The ability to help users increase physical activities is important because insufficient physical activity is a key risk factor for noncommunicable diseases (NCDs) such as cardiovascular diseases, cancer and diabetes (World Health Organization). However, most people need a concrete plan for improvement. If the recommendation is too ambitious, it may harm them. If it is too weak, it may not lead to desired improvement. So our eventual goal is to produce a personalized but injury-free recommender system. This aligns well with the goals and objectives of personal health, which is about providing interventions focused on the individual needs of the patient (McCallum, 2012).

However, a personalized approach does not imply designing an intervention per user. If the method is too individualized, the analysis of previous dataset will not offer any insights to new users. We thus formulate our problem as discovering the patterns of behavior change after a given intervention per subpopulation. We call that the persuasion profile of an intervention. For example, we may discover that a subpopulation of people positively respond to the intervention and become more active in the morning. We may obtain the behavior change patterns for several subpopulations. Our additional goal is to predict how the existing behavior pattern of a user may put him into different types of responders after the intervention. For example, people who are inactive during work time because they work in the office might be more encouraged by the intervention to increase their activity levels during the lunch break. Knowing the conditional probability of observing a behavior change pattern given an observed behavior pattern can help us predict how the intervention will affect a new person for whom we have only pre-intervention observations. In the prediction process we first recognize the behavior pattern of the new user and predict the behavior change pattern associated to the similar people who behave according to the same behavior pattern. Our ultimate goal is to recommend the most beneficial intervention to a new person without the try-and-error process.

The second challenge we face is that our data is sampled at a relatively high frequency. We have developed a method to treat this dense data. We consider the raw sensor data as fine-grained in contrast to coarse-grained data that contains much less information. For example, the total number of steps on a daily basis tells us whether some person is active or not, but the minute by minute measurement of his calorie expenditure or step counts during the day tells us additionally when he is most active. We are interested in extracting high-level concepts, i.e., people’s daily routines from low-level sensory signals. It is due to this fine-grained approach we are able to discover whether an intervention has an effect to users in the morning, during lunchtime, or in the evening.

We propose a mixture model to discover the useful persuasion patterns from an existing dataset. Mixture model is a popular and powerful method for finding the latent structure in a heterogeneous population. In our data we assume that for each person there is one hidden
behavior pattern and one hidden behavior change pattern “determining” her behavior before and after the intervention. The mixture model proposed in this paper discovers the two types of patterns and for each person assigns a partial membership to the behavior patterns and the behavior change patterns. In this way we use more information to represent people’s behavior, and we could generate more accurate predictions for new people.

2. Related work

Mixture models have already been used successfully to understand the different kinds of behavior patterns emerging in the subpopulations. A latent class or growth mixture modeling approach seems to be the most appropriate method for fully capturing information about interindividual differences in intraindividual change taking into account unobserved heterogeneity (different groups) within a larger population (Jung and Wickrama, 2008). Growth mixture modeling (GMM) is a method for identifying multiple unobserved subpopulations, describing longitudinal change within each unobserved sub-population, and examining differences in change among unobserved sub-populations (Ram and Grimm, 2009). GMM has been used in many areas of developmental research including the study of criminology, alcohol use, parenting, and reading difficulties (Curran et al., 2010). Also, it has been used successfully to assess intervention effects in longitudinal randomized trials (Muthén et al., 2002).

Another method that can be used to analyse the behavior change in time series data and detect (non)responders is change point analysis. Change point analysis is the process of detecting distributional changes within time-ordered observations (Matteson and James, 2014). Change point analysis has been successfully used to detect transitions from one activity to another such as sit, stand and walk (Khan et al., 2016). In our domain, change point analysis could be used to identify responders and nonresponders by detecting events of behavior change after the intervention in people affected by the intervention. However, we are more interested to predict whether people will respond to the intervention or not using their current behavior patterns. The methods used in change-point prediction (Hirade and Yoshizumi, 2012) are unable to discover the different behavior change patterns as a result of the intervention.

GMM is also not suitable for our problem. To the best of our knowledge, GMM has been used to model just the post-intervention behavior associated to a trajectory defined by only few observations. For example, in (Muthén et al., 2002), the trajectory is defined by the child’s aggressive behavior in the classroom for grades 1-7. In our case, we want to model the difference between the post-intervention and the pre-intervention behavior associated to a trajectory defined by many observations. This means that we use the pre-intervention trajectory as a baseline for estimating the intraindividual change. For example, in calorie expenditure data we are interested whether people increased or decreased their activity levels at different times during the day after the intervention. GMM is also not able to discover the primary behavior patterns which are predictors of the behavior change patterns after the intervention. However, in our approach, we simultaneously discover both types of patterns. The behavior patterns can be considered as latent features used to cluster people according to their behavior changes. There are also other clustering methods that simultaneously learn feature representations and cluster assignments and they are
applied in different domains (Law et al., 2004; Xie et al., 2015). The Deep Embedded Clustering (DEC) method has been successfully applied on image and text corpora and the results showed significant improvement over state-of-the-art methods (Xie et al., 2015). The mixture model that we propose can be applied on periodic time series data to estimate the impact of interventions on the periodic behavior. In our model the time of intervention is a known information and the baseline is not the behavior of a control group, but the behavior before the intervention.

3. Model

Our goal is to understand how the intervention affects people’s daily activities, whether they increased or decreased the amount of physical activities at each moment during the day. Two types of information are relevant for our analysis: the person’s daily activity pattern before the intervention and the person’s activity change pattern after the intervention. This information is hidden and we would like to discover it using the observed data. The number of "calories burned" indicates the amount of physical activity performed by the individual at a given time. The person’s daily activity pattern relates each minute of the day to a number of calories burned. There is a finite number of daily activity patterns and before the intervention each person performs his daily activities according to one of these patterns (unknown). The person’s activity change pattern relates each time of the day to the change of the number of calories burned after the intervention. There is also a finite number of activity change patterns and after the intervention each person changes his daily activities according to one of these patterns (unknown).

The model proposed in this paper is able to simultaneously discover the daily activity patterns and the activity change patterns. Also, it discovers the conditional probability of observing an activity change pattern given an observed daily activity pattern. This can be useful to predict how a particular person will change his daily activities after the intervention if she is given the treatment. Our model is applied on physical activity time series data obtained from $N$ people. Each observation in the time series gives the number of calories burned by a particular person in a particular time relative to the beginning of the day. The $i$-th observation for the $n$-th person made before the intervention can be defined by a tuple $(X_{a,n,i}, Y_{a,n,i})$. $X_{a,n,i}$ stores the temporal information and $Y_{a,n,i}$ stores the calorie expenditure information. In a similar way we can define the $i$-th observation for the $n$-th person made after the intervention $(X_{b,n,i}, Y_{b,n,i})$. We will assume that there are $M$ observations for each person before and after the intervention.

In our model there are two latent variables $C_{a,n}$ and $C_{b,n}$ associated to every person. The first latent variable indicates the person $n$’s daily activity pattern, and the second latent variable indicates the person $n$’s activity change pattern. There are $K_a$ different daily activity patterns and $K_b$ different activity change patterns. The relationship between the latent and observed variables is given on Fig. 1. In the model we assume that the observations before the intervention are generated from the following probability distribution:

$$
p(Y_{a,n,i}|X_{a,n,i}, C_{a,n}, \beta_a, \Sigma) = N(Y_{a,n,i}|\beta_{a,C_{a,n}}, X_{a,n,i}, \Sigma)
$$

where $\beta_{a,i}$ represents the regression coefficients associated to the $i$-th daily activity pattern. The observations after the intervention are generated from the following probability
Figure 1: The probabilistic graphical model of the person’s behavior under intervention. The observed variables are given in one-lined white circles, the latent variables are given in double-lined white circles and the model parameters are given in one-lined gray circles. In our model, the behavior of the person before the intervention (the relationship between $X_a$ and $Y_a$) depends on his behavior pattern ($C_a$), and the behavior of the person after the intervention (the relationship between $X_b$ and $Y_b$) depends both on his behavior pattern ($C_a$) and his behavior change pattern ($C_b$).

distribution:

$$p(Y_{b,n,i}|X_{b,n,i}, C_{a,n}, C_{b,n}, \beta_a, \beta_b, \Sigma) = \mathcal{N}(Y_{b,n,i}|(\beta_a C_{a,n} + \beta_b C_{b,n}) X_{b,n,i}, \Sigma)$$  \hspace{1cm} (2)$$

where $\beta_{b,i}$ represents the regression coefficients associated to the $i$-th activity change pattern. For simplicity, in our model the covariance matrix $\Sigma$ is common for both distributions. The prior probability distribution over $C_{a,n}$ is defined by the model parameter $\pi$, $p(C_{a,n} = k|\pi) = \pi_k$, so that $\sum_{k=1}^{K_a} \pi_k = 1$. The prior distribution over $C_{b,n}$ given $C_{a,n}$ is defined by the model parameter $\tau$, $p(C_{b,n} = l|C_{a,n} = k, \tau) = \tau_{k,l}$, so that for each $k$, $\sum_{l=1}^{K_b} \tau_{k,l} = 1$. The marginal log-likelihood of our model is:

$$\log p(Y_a, Y_b|X_a, X_b, \beta_a, \beta_b, \Sigma, \pi, \tau) = \sum_{n=1}^{N} \log \sum_{k=1}^{K_a} \tau_k \left[ \prod_{i=1}^{M} \mathcal{N}(Y_{a,n,i}|(\beta_a C_{a,n} X_{a,n,i}, \Sigma) \right]$$  \hspace{1cm} (3)$$

$$\sum_{i=1}^{N} \tau_{k,l} \prod_{j=1}^{M} \mathcal{N}(Y_{b,n,j}|(\beta_a + \beta_b l) X_{b,n,j}, \Sigma)$$  \hspace{1cm} (4)$$

In the learning phase we want to find optimal parameter values for $\beta_a$, $\beta_b$, $\Sigma$, $\pi$ and $\tau$ so that the marginal log-likelihood is maximized. There is no closed form solution for this optimization problem. We use the Expectation Maximization (EM) algorithm to find (local) maximum likelihood parameters (Dempster et al., 1977). EM is an iterative method that alternates between performing expectation and maximization step. In the expectation step, the learning algorithm creates a function for the expectation of the log-likelihood using the current estimate for the parameters. In the maximization step, the learning algorithm
computes parameters maximizing the expected log-likelihood found in the expectation step. In the expectation step of our algorithm we calculate the posteriors over $C_a$ and $C_b$:

$$v_{n,k,l} = p(C_{b,n} = l|C_{a,n} = k, X_{b,n}, Y_{b,n}, \beta_a, \beta_b, \Sigma, \tau) = \frac{\tau_{k,l} \prod_{j=1}^{M} \mathcal{N}(Y_{b,n,j}|(\beta_{a,k} + \beta_{b,l}) X_{b,n,j}, \Sigma)}{\sum_{p=1}^{K_b} \tau_{k,p} \prod_{j=1}^{M} \mathcal{N}(Y_{b,n,j}|(\beta_{a,k} + \beta_{b,p}) X_{b,n,j}, \Sigma)}$$ (5)

$$u_{n,k} = p(C_{a,n} = k|X_{a,n}, Y_{a,n}, X_{b,n}, Y_{b,n}, \beta_a, \beta_b, \Sigma, \pi, \tau) = \frac{\pi_k \prod_{i=1}^{M} \mathcal{N}(Y_{a,n,i}|\beta_{a,k} X_{a,n,i}, \Sigma) \sum_{l=1}^{K_b} \tau_{k,l} \prod_{j=1}^{M} \mathcal{N}(Y_{b,n,j}|(\beta_{a,k} + \beta_{b,l}) X_{b,n,j}, \Sigma)}{\sum_{p=1}^{K_a} \pi_p \prod_{i=1}^{M} \mathcal{N}(Y_{a,n,i}|\beta_{a,p} X_{a,n,i}, \Sigma) \sum_{l=1}^{K_b} \tau_{p,l} \prod_{j=1}^{M} \mathcal{N}(Y_{b,n,j}|(\beta_{a,p} + \beta_{b,l}) X_{b,n,j}, \Sigma)}$$ (6)

In the maximization step, we maximize the following function with respect to $\beta_a$, $\beta_b$, $\Sigma$, $\pi$ and $\tau$:

$$\sum_{n=1}^{N} \sum_{k=1}^{K_a} \left[ \log \pi_k + \sum_{i=1}^{M} \log \mathcal{N}(Y_{a,n,i}|\beta_{a,k} X_{a,n,i}, \Sigma) \right]$$

$$+ \sum_{l=1}^{K_b} \left[ \log \tau_{k,l} + \sum_{j=1}^{M} \log \mathcal{N}(Y_{b,n,j}|(\beta_{a,k} + \beta_{b,l}) X_{b,n,j}, \Sigma) \right]$$ (7)

The quality of the solution depends a lot on the initial parameter values. We use the random restart approach for escaping a local maximum. A disadvantage of our method is that we need to define $K_a$ and $K_b$ in advance. In order to find the optimal $K_a$ and $K_b$ for some dataset, we use $k$-fold cross-validation. We choose the model that produces the highest log-likelihood on the held-out data (Shalizi, 2013).

4. Results

4.1 Dataset

We evaluated our model on the HealthyTogether dataset. This dataset contains the calorie expenditure data of 48 users wearing Fitbit (a wearable accelerometer) for ten days (Chen and Pu, 2014; Yürüten et al., 2014). There is one time series for each person and each day. Each time series contains 1440 observations. Each observation gives the calorie expenditure during a single minute. The minimum calorie expenditure value per minute is 0.77 (resting metabolic rate). Any value larger than 0.77 means that the user did some activity in that minute. During the last five days, the users were under treatment, more specifically, they were using a mobile application, HealthyTogether, which enables them to participate in physical activities together with a partner, send each other messages, and earn badges.

Some of the users in the HealthyTogether dataset forgot to wear the Fitbit device in some days. From the dataset we excluded two users who had more than 3 days with no activity. Additionally, we excluded one user who had very significant increase of his
activities after the intervention (he was associated to the four time series with the largest calorie expenditure in the whole dataset). We can consider this person as an outlier in the data. We applied our model on the remaining 45 users in the HealthyTogether dataset.

4.2 Analysis

We began our analysis by exploring the relationship between the temporal information ($X_a$ and $X_b$) and the calorie expenditure ($Y_a$ and $Y_b$). For this purpose we applied kernel regression on all observations from the pre-intervention and the post-intervention data separately. These results are shown on Fig. 2. We can see that there is nonlinear relationship between the time of the day and the average number of calories burned at that time. For example, there are two main peaks of activity, around 13:00 and around 19:00. In our model, there is a linear relationship between $X_a$ and $Y_a$, and between $X_b$ and $Y_b$. In order to be able to model the nonlinear relationship that exists in the data, we apply nonlinear transformation on the temporal features (polynomial regression). We include polynomial bases in $X_a$ and $X_b$ and we transform the vector $[1, t]$ into the vector $[1, t, t^2, \ldots, t^d]$, where $t$ is the time relative to the beginning of the day. We still need to determine $d$. A standard approach is to treat $d$ as a hyperparameter, in addition to $K_a$ and $K_b$, and from all the possible models that correspond to different values for $d$, $K_a$ and $K_b$, to choose the model according to some model selection criteria. Because it is very time-consuming to use $k$-fold cross-validation method for this purpose, we choose the value of $d$ in advance using a more heuristic approach described below, and after that we use cross-validation only to determine the optimal $K_a$ and $K_b$.

We generated $X_a$ and $X_b$ using different values for $d$, and we applied separate linear regressions on the whole dataset. We had two criterias for choosing $d$: (1) how well we can approximate the functions obtained with the kernel regression, and (2) Bayesian Information Criterion. The results suggested to choose the model with $d = 10$ and in the rest of the analysis we use polynomials up to degree 10 to represent the temporal information in $X_a$. 

![Figure 2: Calorie expenditure during a single day. People are the most active around 12:00 and around 19:00.](image)
Figure 3: The average log-likelihood after two 10-fold cross-validations for models with different $K_a$ and $K_b$. As the number of behavior patterns and the number of behavior change patterns increase, the model fits better on the training data. However, we choose $K_a = 3$ or $K_b = 3$ because a larger number of patterns does not bring significant improvement on the held-out data.

and $X_b$. It is interesting that the functions that correspond to the pre-intervention and the post-intervention data on Fig. 2 are slightly different. After the intervention there is a small increase of the activity levels in the morning and a small decrease of the activity levels in the evening. Our method should give us more accurate information about the subpopulations that change their daily activities after the intervention.

We choose to evaluate models with up to five daily activity patterns and up to five activity change patterns. It is unlikely that we can extract more patterns from a small dataset with 45 people. For every combination of $K_a$ and $K_b$ we run 10-fold cross-validation. For each fold, we train 100 models with different initial parameter values and we choose the model with the highest log-likelihood. We performed the cross-validation two times in order to get more stable results. The average log-likelihoods on the training and the held-out data are shown on in Fig. 3. We can observe that the log-likelihood on the training data strictly increases as we increase the number of patterns. However, the log-likelihood on the held-out data increases until we reach three daily activity patterns and three activity change patterns.

We decided to choose the model with $K_a = 3$ or $K_b = 3$ because a larger number of patterns does not bring significant improvement on the held-out data. The daily activity patterns are given on Fig. 4(a). All of them have two peaks, during the morning and during the evening. However, these patterns are characterized by different activity levels through the day. We discovered three activity change patterns representing "inactive", "moderately active" and "highly active" people. From Table 2 we can see that most of the people (90%) are either "inactive" or "moderately active". The activity change patterns are given in Fig. 4(b). The first pattern represents the people who decreased their activity levels through the day, mostly during the evening ("negative responders"). The second pattern represents the people who moderately increased their activities through the day, mostly during the morning ("moderate responders"). The third pattern represents people who greatly increased their activities through the day, both in the morning and in the evening ("strong responders").
Figure 4: The behavior of people before the intervention (daily activity patterns) and the behavior change of people after the intervention (activity change patterns). Daily activity patterns represent people who were "inactive", "moderately active" and "highly active" before the intervention. Activity change patterns represent people who were "negative responders", "moderate responders" and "strong responders" under the intervention.

Table 1: Prior and transition probabilities for the model with three daily activity patterns (DAP) and three activity change patterns (ACP) trained on the whole dataset. People who were less active before the intervention tend to respond more positively to the intervention.

<table>
<thead>
<tr>
<th>transition probability (τ)</th>
<th>prior (π)</th>
<th>ACP 1</th>
<th>ACP 2</th>
<th>ACP 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAP 1</td>
<td>0.4667</td>
<td>0.3333</td>
<td>0.5714</td>
<td>0.0952</td>
</tr>
<tr>
<td>DAP 2</td>
<td>0.4444</td>
<td>0.4000</td>
<td>0.4500</td>
<td>0.1500</td>
</tr>
<tr>
<td>DAP 3</td>
<td>0.0889</td>
<td>0.7500</td>
<td>0.0000</td>
<td>0.2500</td>
</tr>
</tbody>
</table>

Most people are either "negative responders" or "moderate responders" (85%). Both of these groups increased their activities during the morning, so this explains the difference between the pre-intervention and the post-intervention data in this period on Fig. 2. From the transition probabilities presented in Table 1 we conclude that it is more probable that the "inactive" people will become more active after the intervention than the "moderately active" and "highly active" people (67% of the inactive people, 60% of moderately active people and 25% of highly active people responded to the intervention). These people tend to become more active in the morning. By contrast, the "moderately active" and the "highly active" people who responded to the intervention tend to become more active in the evening. Although we don’t have enough data for a strong claim, it is more likely that the "strong responders" had greater activity levels before the intervention as well (25% of the "highly active" people, 15% of the "moderately active" people and 10% of the "inactive" people become "strong responders").
Table 2: Number of people that are associated to each pair of different daily activity patterns (DAP) and activity change patterns (ACP). The last column and the last row contain the number of people associated to each individual pattern. The most common transition is from "inactive" people to "moderate" responders.

<table>
<thead>
<tr>
<th># transitions</th>
<th>ACP 1</th>
<th>ACP 2</th>
<th>ACP 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAP 1</td>
<td>7</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>DAP 2</td>
<td>8</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>DAP 3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>

5. Conclusion

We presented a method to discover the patterns of behavior change for a given intervention. We call these patterns an interventions persuasion profiles. Our method is able to cluster people both according to their current habits and the patterns of behavior change after the intervention. It also discovers the conditional probability of a behavior change pattern given an observed existing pattern. We applied our method on the HealthyTogether dataset and have shown three persuasion profiles: "negative responders" decreased their activity levels, mainly due to impartible partners, "moderate responders" increased their activity levels, mostly during the morning, and "strong responders" increased their activity levels both during the morning and the evening. Most people (85%) fall into the first two profiles. Further, it is more probable that people with lower activity levels before the intervention will improve their activity levels after the intervention (67% of the inactive people, 60% of moderately active people and 25% of highly active people responded to the intervention). However, pairing active users with their sedentary counterparts can have a negative effect on the latter potentially due to an implicit fear structure.

Our method discovers the persuasive power of a given intervention. It can be used to design effective recommendations for behavior change intended for people with different habits as captured by fitness trackers. Furthermore, the system is able to predict post-intervention behavior change without actually administering the intervention, thus saving users from potential injuries. The most novel contribution is the discovery of windows of opportunities for persuasion (such as the morning time) due to the fine-grained approach. We believe this work can lead to the design of a novel behavior recommender system.

References


Cosma Shalizi. *Advanced data analysis from an elementary point of view*. Citeseer, 2013.
