Counterfactual Prediction for Time Series

Junhao (Bear) Xiong

May 2018

1 Introduction

1.1 Setting

For a given patient, we are given some **irregularly** measured values of a signal (e.g. INR) on some **discrete** time points, as well as values of some contexts, including **treatments** (binary, e.g. whether the patient is given anticoagulant at those time points) and **chronic conditions** (binary, e.g. whether the patient has chronic kidney failure), at some time points.
We are interested in the future trajectory of a signal when we change the values of the contexts. As an example, we can ask questions like "how would the patient's INR progress after time t had we stop administering anticoagulant, while in reality we did administer it to him/her?"

1.2 Goal

We want to be able to predict counterfactually the trajectory of any signal, given the past trajectory and some contexts.

2 Definition

- Treatments
  $\boldsymbol{x}_t = (x_{t1}, ..., x_{tN})$ an N-dimensional vector, where x_{tn} is a binary value indicating whether treatment n is given at time t.
- Chronic condition
  $\boldsymbol{c} = (c_1, ..., c_M)$ an M-dimensional vector, where c_m is a binary value representing whether a patient has chronic condition m. Note $\boldsymbol{c}$ is the same for all time points.
- Signal
  y_t a scalar representing the value of the signal at time t.

3 Model Form

Our model is a dynamic linear model (DLM), which is a state-space model where the transition and emission functions are linear (represented by matrices) and the noises are Gaussian.

- Transition
$$z_t = z_{t-1} + \epsilon_t \tag{1}$$
where
  - z_t is a scalar representing the latent state value at time t. The interpretation of z_t is the underlying value of a marker with some natural fluctuations, without the interference of any treatments, chronic conditions, etc.
  - $\epsilon_t \sim \mathcal{N}(0, \sigma_1^2)$ are i.i.d

- Emission
$$y_t = z_t + \sum_{j=1}^{J} \boldsymbol{a}_j \cdot \boldsymbol{x}_{t-j} + \boldsymbol{b} \cdot \boldsymbol{c} + \delta_t \tag{2}$$
where
  - y_t is the observed value of a marker, determined by its underlying value, treatments received, chronic conditions, etc.
  - $\boldsymbol{a}_j$ is an I-dimensional coefficient vector. Note that $\boldsymbol{a}_j$ captures the effect of treatment at time $t-j$. For example, $\boldsymbol{a}_1$ always captures the effect of treatment given at the time point prior to the current one, so it's reasonable to assume $\boldsymbol{a}_1$ to be constant. We only consider the effect of the previous J treatments and we sum over J to capture the effects of past treatments.
  - $\boldsymbol{b}$ is a **time-invariant** vector of coefficients of dimensions M
  - $\delta_t \sim \mathcal{N}(0, \sigma_2^2)$ are i.i.d

4 Learning and Inference

- Inference
  Since the model is DLM, we can use Kalman Filter to perform exact inference. Murphy 18.3.1.6 includes derivation of the Kalman Filter updates for a model of similar form.

- Learning
  Parameters to be learned are $\boldsymbol{a_j}$, $\boldsymbol{b}$, σ_1^2, σ_2^2. We can use EM to learn these parameters. The derivations of EM for Kalman Filter are in Ghahramani and Hinton 1996b. We can modify the derivation for our model.

5 Prediction

After parameters are learned, we can then recursively compute the means and covariance of $P(y_{t+k}|y_1, ..., y_t)$ for k-step ahead prediction. Counterfactual prediction can be attained by setting $\boldsymbol{x}_t$ to different values at each predicted time point.
Examples prediction plots on real INR data are shown in the last section.

6 Missingness

The observational data is irregularly sampled and the sampling time points may be different among individuals, which motivates "binning" the data, i.e. putting the time series into the same time interval. After binning, there would inevitably be time intervals in which there are no observations. Kalman filter has a natural way to deal with missingness by essentially treating the emission matrix at the missing time points as zero and using the usual updates.

7 Potential Future Directions

- Including interaction terms between treatments and/or chronic conditions.
  The current model still has some issues identifying the correct directions of some coefficients. Explicitly modeling the interaction effects could perhaps help capture the true impact of treatments on the observed measurements.

- Better way to incorporate individual heterogeneity
  The current model assumes the same set of parameters generates all the data points. Experiments show that the resulting model has lower prediction mean square error compared to the model that assumes a different set of parameters for each individual. However, different individuals might still respond to treatments differently and there could be better way to incorporate this information into the model.

- Using more complex and more expressive model
  The current model is linear with identity transition and emission matrix and added effects. Incorporating non-linearity could potential improve the model.

- Modeling different signals
  Some analysis on the current data set shows that the treatments which supposed to increase the INR values (e.g. anticoagulant, aspirin, NSAID) only increase the INR values about half of the times when we consider treatments given within 2 days before an INR measurement is made. This could make it very difficult for the model to identify the actual treatment effects. Modeling other signals could test the hypothesis that the model would identify the coefficients better if the treatment effects are more clearly shown from the data.

8 Example Plots

These plots are generated on real INR data after some modeling choices are made.