Unsupervised Learning of Disease Progression Models

Xiang Wang
IBM Research
Yorktown Heights, NY
wangxi@us.ibm.com

David Sontag
New York University
New York, NY
dsontag@cs.nyu.edu

Fei Wang
IBM Research
Yorktown Heights, NY
fwang@us.ibm.com

ABSTRACT

Chronic diseases, such as Alzheimer's Disease, Diabetes, and Chronic Obstructive Pulmonary Disease, usually progress slowly over a long period of time, causing increasing burden to the patients, their families, and the healthcare system. A better understanding of their progression is instrumental in early diagnosis and personalized care. Modeling disease progression based on real-world evidence is a very challenging task due to the incompleteness and irregularity of the observations, as well as the heterogeneity of the patient conditions. In this paper, we propose a probabilistic disease progression model that address these challenges. As compared to existing disease progression models, the advantage of our model is three-fold: 1) it learns a continuous-time progression model from discrete-time observations with non-equal intervals; 2) it learns the full progression trajectory from a set of incomplete records that only cover short segments of the progression; 3) it learns a compact set of medical concepts as the bridge between the hidden progression process and the observed medical evidence, which are usually extremely sparse and noisy. We demonstrate the capabilities of our model by applying it to a real-world COPD patient cohort and deriving some interesting clinical insights.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications-Data mining

Keywords

Bayesian network, Markov jump process, disease progression modeling, medical informatics

1. INTRODUCTION

Chronic diseases usually progress slowly over time. For example, Chronic Obstructive Pulmonary Disease (COPD) can take well over 10 years to evolve from (according to the GOLD criteria [6]) Stage I (mild) to Stage IV (very severe). It may also take 10 years for Congestive Heart Failure (CHF) to progress from Stage I (mild) to Stage IV (severe). Late detection and intervention for such chronic diseases significantly increases the burden on both the patients and the healthcare system. Being able to detect the development of chronic diseases at an early stage is instrumental to preventive care and personalized medicine.

Disease Progression Modeling (DPM) [11], the modeling of the progression of a target disease with computational methods, is an important technique that can help with the early detection and management of chronic diseases. By characterizing the entire disease progression trajectory, DPM also facilitates disease prognosis improvement, drug development, and clinical trial design. There have been a few existing work on DPM that targets a specific domain. For example, Jackson et al. [9] presented a general Hidden Markov Model for simultaneously estimating the transition rates and the probabilities of stage misclassification. Ito et al. [8] conducted a meta analysis to model the longitudinal changes of patients with mild to moderate Alzheimer's disease. Zhou et al. [17] proposed a fused group lasso approach for disease progression modeling with known biomarkers. Exarchos et al. [5] developed a dynamic Bayesian network based technique to model the progression of Coronary Atherosclerosis. Some difficulties of applying these existing methods to general-purpose evidence-based disease progression modeling include:

- ***Multiple Covariates***. The progression of disease usually involves the evolution of many different types of hidden covariates. Identifying these covariates with limited or no supervision is a very challenging task.
- ***Progression Heterogeneity***. Different patients may have different progression trajectories. For instance, a COPD patient who had lung infection could progress rapidly, whereas another COPD patient who quit smoking could remain stable. There is no natural alignment between different records with varied progression rates.
- ***Incomplete Records***. The patient records are not complete, meaning that in most of the cases it only covers a segment of the entire progression trajectory.
- ***Discrete Observation***. Although the disease progression is a continuous-time process, the patient conditions are only observed at discrete timestamps with varied intervals.
- ***Irregular Visits***. The patients will only have medical records when they visit a medical facility. There-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
KDD'14, August 24–27, 2014, New York, NY, USA.
Copyright 2014 ACM 978-1-4503-2956-9/14/08 ...$15.00.
http://dx.doi.org/10.1145/2623330.2623754.

fore the time granularity of patient records vary significantly across different patients, and for the same patient over different time periods.

- ***Limited Supervision.*** For some diseases we have very limited yet crucial domain knowledge available, e.g. the known symptoms of a particular disease. Incorporating the available domain knowledge into the progression model is a nontrivial task.

To address these challenges, we propose an *unsupervised* disease progression model. As show in Figure 1, our model is composed of three layers: The top layer is a Markov Jump Process which captures the continuous-time diseases state transitions. The middle layer is a set of Markov chains capturing the relationship between the hidden state transitions and the onset pattern of a set of comorbid conditions (comorbidities). The third layer is a noisy-or network [14] (Figure 2) capturing the relationship between those comorbidities and the observed clinical evidence. Note that instead of linking the clinical observations directly to the disease progression states, we "group" them into comorbidities, which tend to evolve coherently with the progression of disease. This abstraction also makes the learned DPM more robust and interpretable. An Expectation-Maximization (EM) based algorithm is presented to estimate the parameters as well as the hidden variables. We apply our model to a real-world COPD patient cohort to demonstrate its capabilities.

It is worthwhile to highlight the following aspects of the proposed model:

- Our model is unsupervised. We learn the entire disease progression trajectory from the observed patient records without any training labels on the ground truth stages that a patient was in.
- Our model learns the continuous-time disease progression trajectory even though the medical records were only observed at discrete timestamps with irregular intervals.
- Our model can "stitch together" partial disease trajectories (i.e., incomplete records from individual patients) into a global path of disease progression.
- Our model can learn meaningful comorbidities associated with different disease stages. We allow the injection of *anchor findings*, which are clinical features that distinctly signifies the presence of a certain comorbidity, to improve the interpretability and medical validity of our model.

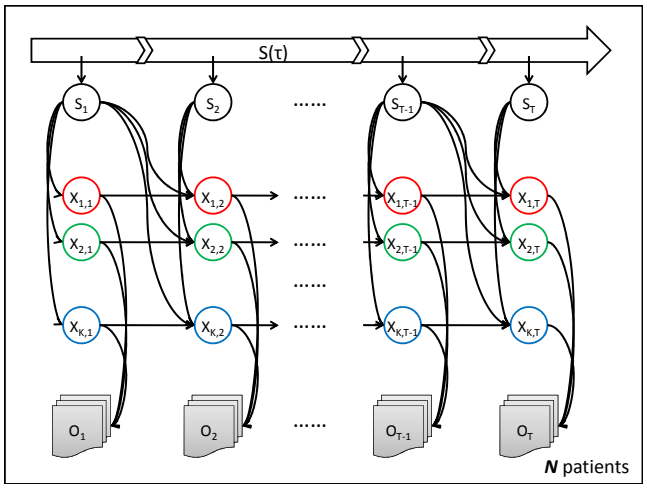


Figure 1: The outline of our model: S are progression state variables, X are comorbidity variables, and O are observed clinical findings.

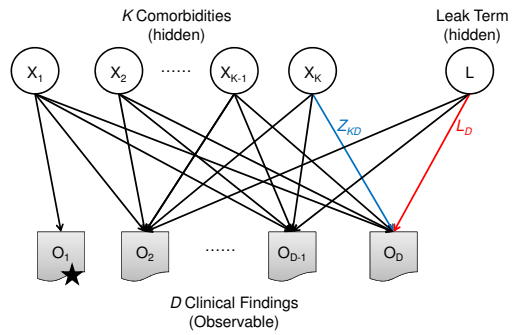


Figure 2: The noisy-or Bayesian network (also known as QMR-DT network). The clinical findings can be activated by a present comorbidity, or by the always-on leak term. The starred finding (O_1) is an anchor, which means it can only be activated by a specific comorbidity (X_1 in this case).

2. RELATED WORK

Disease progression modeling is an important topic in medical informatics [11]. Existing work on disease progression models have been proved effective for drug development and early intervention. For example, Post et al. [12] proposed a family of models to describe the progression of degenerative diseases (such as type 2 diabetes and Parkinson's disease) as a function of disease process and treatment effects. De Winter et al. [3] developed a mechanism based technique for modeling the progression of diabetes mellitus by tracking the interaction between several key indicators. Ito et al. [8] presented a model based on literature meta-analysis to describe the longitudinal changes of patients with mild to moderate Alzheimer's disease. These papers from the medical field tend to be specific to a single target disease. They require substantial domain knowledge on the progression, mechanism, and key indicators/measurements for the target disease. This is not the case for our model because we aim to learn a general-purpose model for any chronic disease based on a general input data type: Electronic Health Records. We do not assume prior knowledge of either the ground truth progression stages or the key indicators that signify the stage transitions.

Another line of efforts, to which our approach belongs, model the progression of disease using machine learning and statistical techniques based on observational data, also referred to as evidence based modeling. For example, Jackson et al. [9] developed a multistage Hidden Markov Model and applied it to an aneurysm screening study. Sukkar et al. [15] applied Hidden Markov Model to Alzheimer's disease. Cohen et al. [2] performed hierarchical clustering of 45 physiological, clinical, and treatment variables collected

every minute in an intensive care unit to identify patient states. They used these clusters to visualize individual patient states over time, with each state obtained by assigning the corresponding data point to a cluster. Zhou et al. [17] proposed a fused group lasso formulation for disease progression modeling with known biomarkers. Zhou et al. modeled disease progression using a multi-task learning framework [18]. As pointed out by Yang et al. [16], if we model patient records as time sequences, disease progression stages can also be captured by joint time sequence segmentation. The difference between these existing techniques and our approach includes: 1) Unlike the discrete-time HMM based models, our model is continuous-time, which is a more natural way to handle the irregular time intervals between patient visits; 2) Some existing work were dealing with synchronous time sequences, i.e. there are explicit or implicit correspondence between the timestamps of different patients. This assumption is not true for our model where a patient record could start and end at any point of the progression path.

3. MODEL

We model the progression of a target disease based on the longitudinal clinical findings of a cohort of patients who have developed, or are at risk developing, such disease. First of all, we use a Markov Jump Process to model the transition of disease stages/states, which implies 1) the progression is continuous-time; 2) the transition probability to the future state only relies the current state and the time span.

Second, we use the onset pattern of comorbidities to drive the transitions of the Markov Jump Process. Generally speaking, a comorbidity is a disease or syndrome that co-occurs with the target disease. For example, hypertension is a common comorbidity of diabetes and osteoporosis is a common comorbidity of COPD. Since the onset of a new comorbidity often signifies the exacerbation of the target disease, we use the onset pattern of multiple comorbidities to collectively capture the state transitions of the target disease. We assume the comorbidities are conditionally independent given the state of the target disease. This is a mild assumption in our case because medically meaningful comorbidities are by definition mutually independent diseases.

Finally, in order to infer the presence of the comorbidities from the observed clinical findings, we use a bipartite noisy-or Bayesian network [7, 14] (Figure 2). Simply speaking, given a set of comorbidities and a set of clinical findings, we assume an observed clinical finding was "activated" by the presence of any of the comorbidities with a certain *activation probability*; it is also possible that none of the comorbidities is present and the finding was activated by an always-on hidden cause with a *leak probability*. Such structure is especially well suited to our setting due to its flexibility in modeling sparse and noisy observations.

Our overall model is illustrated in Figure 1. Some notations that are used throughout the rest of the paper are listed in Table 1.

3.1 Variables

We assume that there are K different comorbidities underpinning the progression of the target disease, D different clinical findings, and M different progression states. We have N different patients and each patient n has T_n visits, with timestamps $\tau_1, \ldots, \tau_{T_n}$. We introduce the following variables for our model:

Table 1: Notations and Meanings

N	Number of patients
M	Number of progression states
K	Number of comorbidities
D	Number of clinical findings
t/τ	Discrete/continuous timestamp
$\mathbf{S}$	Disease states
$\mathbf{X}$	Cormorbidities
$\mathbf{O}$	Clinical findings
π	Initial state distribution
Q	Transition generator matrix
$A(\Delta)$	Transition probability matrix over time span Δ
B	Onset probability for comorbidities
Z	Activation probability for clinical findings
L	Leak probability for clinical findings

$$\begin{aligned} \text{Disease States (hidden):} \quad & S_{n,t} \in \{1, \ldots, M\} \\ \text{Comorbidities (hidden):} \quad & X_{k,n,t} \in \{0, 1\} \\ \text{Clinical Findings (observable):} \quad & O_{d,n,t} \in \{0, 1\} \end{aligned}$$

The underlying patient state is assumed to evolve according to a continuous-time Markov process. However, we only observe evidence about the patient at specific times $\tau_1, \ldots, \tau_{T_n}$ when the patient interacts with the healthcare system (e.g. visits a doctor or fills a prescription). The random variables $S_{n,t}$ and $X_{k,n,t}$ for $t = 1, \ldots, T_n$ denote the patient's disease state and comorbidities, respectively, at these discrete times. $X_{k,n,t} = 1$ means patient n has comorbidity k at visit t, and is 0 otherwise. In our generative model, the presence or absence of comorbidity k at time t, i.e. $X_{k,n,t}$, depends on whether the patient had the comorbidity in the previous time step, $X_{k,n,t-1}$ as well as the patient's progression state. Associated with each visit is a set of observed clinical findings, $O_{d,n,t}$, which we assume to be conditionally independent given the disease's current state of progression and the comorbidities. We model these using a noisy-or bipartite network, shown in Figure 2 and discussed further in Section 3.4.

3.2 Markov Model of Disease Progression

The continuous-time Markov process is parameterized by an $M \times M$ **transition generator matrix** Q that drives the transition between M states. The transition probability from state i to j with a time span Δ is defined as:

$$\begin{aligned} A_{ij}(\Delta) &\triangleq P(S_t = j | S_{t-1} = i, \tau_t - \tau_{t-1} = \Delta; Q) \\ &= \text{expm}(\Delta Q)_{ij}, \end{aligned} \tag{1}$$

where expm$(\cdot)$ is the matrix exponential.

The $M \times 1$ **initial state probability** π is defined as:

$$\pi_i \triangleq p(S_0 = i), \; i = 1, \ldots, M. \tag{2}$$

3.3 Model of Comorbidities

The disease progression is manifested through a number of comorbidities, which we model as binary random variables. The state of each comorbidity variable at time t, $X_{k,n,t}$, is decided by the current disease state $S_{n,t}$ as well as the state of the same comorbidity variable at the previous time step, $X_{k,n,t-1}$. We further constrain that a comorbidity onset can only happen when there is a state transition from $S_{n,t-1}$ to

$S_{n,t}$ (which means the overall condition of the patient has changed). This conditional distribution is parameterized by a $K \times M \times 2$ **comorbidity onset probability** B:

$$B_{k,m,a} \triangleq p(X_{k,n,t} = 1 | S_{n,t} = m, S_{n,t-1} \neq m, X_{k,n,t-1} = a), \tag{3}$$

$a \in \{0, 1\}$. We assume that B is homogeneous with time t and independent of the particular patient n. For what follows, we define the shorthand:

$$\begin{aligned} &\Psi(X_{k,n,t} | S_{n,t}, X_{k,n,t-1}) \\ &\triangleq B_{k,S_{n,t},X_{k,n,t-1}}^{X_{k,n,t}} (1 - B_{k,S_{n,t},X_{k,n,t-1}})^{1-X_{k,n,t}}. \end{aligned} \tag{4}$$

At $t = 0$, the initial onset probability B^0 is defined as:

$$B_{k,m}^0 \triangleq p(X_{k,n,0} = 1 | S_{n,0} = m). \tag{5}$$

3.4 Model of Clinical Findings

What remains is to describe how we generate the observed findings (e.g., the diagnosis codes) given the disease and comorbidity states. For this part of the model we use a bipartite noisy-or Bayesian network. Such networks have been previously used for medical diagnosis, a prominent example being the Quick Medical Reference (QMR-DT) [14]. However, rather than having a single diagnosis model per patient, we use a new one for each visit, tying them together by having the prior distribution for the comorbidities (the top-level in Figure 2) depend on the state of the comorbidities in the previous visit and on the overall disease state.

We have one parameter $Z_{k,d}$ for the **activation probability** of comorbidity k and finding d. The larger $Z_{k,d}$ is, the more likely observation $O_{d,n,t}$ is to be 1 whenever $X_{k,n,t}$ is 1. If none of the comorbidity variables are active, an observation $O_{d,n,t}$ still has the chance to be activated by an unknown cause, assumed to always be on, with the **leak probability** $L_d \triangleq p(O_{d,n,t} = 1 | \sum_{k=1}^K X_{k,n,t} = 0)$. Overall, $O_{d,n,t}$ follows a noisy-or distribution. We define $\Phi(O_{d,n,t}|\mathbf{X})$ as a shorthand for:

$$\Phi(O_{d,n,t}|\mathbf{X}) \triangleq \left(1 - (1 - L_d) \prod_{k=1}^K (1 - X_{k,n,t} Z_{k,d})\right)^{O_{d,n,t}} \left((1 - L_d) \prod_{k=1}^K (1 - X_{k,n,t} Z_{k,d})\right)^{1-O_{d,n,t}}$$

Only $\mathbf{O}$ is observable. $\mathbf{X}$ and $\mathbf{S}$ are hidden, and π,Q,B,L,Z are parameters to estimate.

4. INFERENCE

We give the progression model parameters B, L and Z Beta priors and perform marginal inference over both the latent variables and the model parameters using Gibbs sampling. Estimation of the Markov Jump Process is challenging in our setting due to incomplete observations nonequidistant in time [10], and for computational reasons it is preferable to perform Maximum Likelihood Estimation of the parameters of the continuous-time Markov chain, i.e. $\max_{\pi,Q} p(\mathbf{O}; \pi, Q)$. We use Expectation-Maximization to find a local optimum of the likelihood. The overall learning algorithm is summarized in Algorithm 1.

We omit details of the Gibbs sampling and EM algorithm that are standard, instead focusing our description on estimation of the continuous-time Markov model and on other aspects that are special to our setting.

Algorithm 1: Our Algorithm

Input: Clinical findings $\mathbf{O}$
Output: Transition generator matrix Q, initial state probability π, activation probability Z, leak probability L, onset probability B

```
Initialize S, X, Q, π, B, Z, L;
repeat
    // E-step
    repeat
        Gibbs sampling from p(S, X, B, L, Z|O; π, Q);
    until Convergence;
    Use samples to estimate p(S_{n,0} = i|O; π, Q) and
    p(S_{n,t-1} = i, S_{n,t} = j|O; π, Q);
    Compute C_ij(Δ) = Σ_{n=1}^{N} Σ_{t=1}^{T_n} p(S_{n,t-1} = i, S_{n,t} =
    j|O; π, Q) 1_{τ_t − τ_{t-1} = Δ}, ∀Δ;
    // M-step
    Update π_i ← Σ_{n=1}^{N} p(S_{n,0}=i|O;π,Q) / Σ_{n=1}^{N} Σ_{i=1}^{M} p(S_{n,0}=i|O;π,Q);
    repeat
        Compute E[N_ij(Δ)|k, l, Q] and E[R_i(Δ)|k, l, Q]
        (see Section 4);
        Update Q ← Σ_Δ Σ_{k,l∈[M]} E[N_ij(Δ)|k,l,Q] C_{k,l}(Δ) / Σ_Δ Σ_{k,l∈[M]} E[R_i(Δ)|k,l,Q] C_{k,l}(Δ);
    until Convergence;
until Convergence;
```

4.1 E-Step

The complete log-likelihood is $\log p(\mathbf{O}, \mathbf{S}, S(\tau); \pi, Q)$. Both the discrete states $\mathbf{S}$ and the continuous process $S(\tau)$ are hidden, so in the E-step we take their expectation with respect to the posterior distribution over $\mathbf{S}$ and $S(\tau)$ using the current parameter setting π', Q':

$$\begin{aligned} &\mathbb{E}_{p(\mathbf{S}, S(\tau)|\mathbf{O}; \pi', Q')}[\log p(\mathbf{O}, \mathbf{S}, S(\tau); \pi, Q)] \\ &= \mathbb{E}_{p(\mathbf{S}|\mathbf{O}; \pi', Q')}[\log \pi + \log p(\mathbf{O}|\mathbf{S})] \\ &\quad + \mathbb{E}_{p(\mathbf{S}, S(\tau)|\mathbf{O}; \pi', Q')}[\log p(\mathbf{S}, S(\tau); \pi, Q)]. \end{aligned} \tag{6}$$

The second term in Eq. (6) can be written as (see [10] for detailed derivation):

$$\mathbb{E}_{p(\mathbf{S}, S(\tau)|\mathbf{O}; \pi', Q')}[\log p(\mathbf{S}, S(\tau); \pi, Q)] = \sum_{\Delta} \sum_{i,j \in [M]} C_{ij}(\Delta) \left(\sum_{k,l \in [M], k \neq l} (\log Q_{kl}) \mathbb{E}[\mathcal{N}_{kl}(\Delta)|\mathbf{S}; Q'] - Q_{kl} \mathbb{E}[\mathcal{R}_k(\Delta)|\mathbf{S}; Q'] \right)$$

where we define

$$C_{ij}(\Delta) \triangleq \sum_{n=1}^N \sum_{t=1}^{T_n} p(\mathbf{O}, S_{n,t-1} = i, S_{n,t} = j; \pi', Q') \mathbf{1}_{\tau_t - \tau_{t-1} = \Delta}$$

and for which we need to compute

$$p(\mathbf{O}, S_{n,t-1} = i, S_{n,t} = j; \pi', Q'), \forall n, t, i, j. \tag{7}$$

$\mathcal{N}_{kl}(\Delta)$ is defined as the number of transitions from state k to l during the time interval Δ. $\mathcal{R}_k(\Delta)$ is the amount of time spent in state k during the time interval Δ. Note that these are auxiliary latent variables that we introduce to help us learn the continuous-time Markov process. We will later describe how to estimate their expectations.

Gibbs Sampling

The posterior distribution for B, Z, and L can be obtained in closed-form since the Beta distribution is conjugate for the Bernoulli. To perform Gibbs sampling of these continuous parameters (with range [0,1]), we use a fixed set of discrete values. For instance, we can sample from 0.01 to 0.99 with 0.01 increment; finer sampling granularity can be chosen at the cost of longer runtime (the complexity is linear to the number of values).

We perform block sampling of both $\mathbf{X}$ and $\mathbf{S}$, substantially improving the mixing time of Gibbs sampling at the cost of only double the running time per iteration. For fixed n and k, we sample $X_{k,n,t}, t = 1, \ldots, T_n$ from the joint distribution:

$$\begin{aligned} &p(X_{k,n,0}, \ldots, X_{k,n,T_n}|\Theta, \mathbf{O}) \\ &= p(X_{k,n,0}|\Theta, \mathbf{O}) \prod_{t=1}^{T_n} p(X_{k,n,t}|\Theta, \mathbf{O}, X_{k,n,t-1}) \\ &= p(X_{k,n,0}|\Theta, \mathbf{O}) \prod_{t=1}^{T_n} \frac{p(X_{k,n,t}, X_{k,n,t-1}|\Theta, \mathbf{O})}{p(X_{k,n,t-1}|\Theta, \mathbf{O})}, \end{aligned} \tag{8}$$

in which $p(X_{k,n,t} = i|\Theta, \mathbf{O})$ and $p(X_{k,n,t} = j, X_{k,n,t-1} = i|\Theta, \mathbf{O}), i, j \in \{0, 1\}, \forall t$ need to be estimated. Here $\Theta = (\mathbf{X}_{-k}, \mathbf{S}, B, Z, L)$.

Define $\alpha_i(t) \triangleq p(O_{d,n,0}, \ldots, O_{d,n,t}, X_{k,n,t} = i|\Theta)$. It can be updated sequentially as

$$\begin{aligned} &\alpha_i(0) \leftarrow \Psi(X_{k,n,0} = i|S_{n,0}) \\ &\quad \prod_{d=1}^{D} \Phi(O_{d,n,0}|\{X_{k',n,0}\}_{k' \neq k}, X_{k,n,0} = i) \\ &\alpha_i(t+1) \leftarrow \\ &\left(\sum_{j=0,1} \alpha_j(t) \Psi(X_{k,n,t+1} = i|S_{n,t+1}, X_{k,n,t} = j) \right) \\ &\quad \prod_{d=1}^{D} \Phi(O_{d,n,t+1}|\{X_{k',n,t+1}\}_{k' \neq k}, X_{k,n,t+1} = i) \end{aligned} \tag{9}$$

Also define $\beta_i(t) \triangleq p(O_{d,n,t+1}, \ldots, O_{d,n,T_n}|X_{k,n,t} = i, \Theta)$. It can be updated sequentially as $\beta_i(T_n) \leftarrow 1$ and

$$\begin{aligned} \beta_i(t-1) \leftarrow &\sum_{j=0,1} (\beta_j(t) \Psi(X_{k,n,t} = j|S_{n,t}, X_{k,n,t-1} = i) \\ &\prod_{d=1}^{D} \Phi(O_{d,n,t}|\{X_{k',n,t}\}_{k' \neq k}, X_{k,n,t} = j)) \end{aligned}$$

Then it is easy to show [13] that:

$$\begin{aligned} &p(X_{k,n,t} = i|\Theta, \mathbf{O}) \propto \alpha_i(t)\beta_i(t), \\ &p(X_{k,n,t+1} = j, X_{k,n,t} = i|\Theta, \mathbf{O}) \\ &\propto \alpha_i(t)\beta_j(t+1)\Psi(X_{k,n,t+1} = j|S_{n,t+1}, X_{k,n,t} = i) \\ &\quad \prod_{d=1}^{D} \Phi(O_{d,n,t+1}|\{X_{k',n,t+1}\}_{k' \neq k}, X_{k,n,t+1} = j)). \end{aligned}$$

Therefore the forward sampling probability for $X_{k,n,t}$ is:

$$\begin{aligned} &p(X_{k,n,t+1} = j|X_{k,n,t} = i, \Theta, \mathbf{O}) \\ &\propto \frac{\beta_j(t+1)}{\beta_i(t)} \Psi(X_{k,n,t+1} = j|S_{n,t+1}, X_{k,n,t} = i) \\ &\quad \prod_{d=1}^{D} \Phi(O_{d,n,t+1}|\{X_{k',n,t+1}\}_{k' \neq k}, X_{k,n,t+1} = j)). \end{aligned} \tag{10}$$

The α's cancel out in Eq. (10), so we do not actually need to compute them (other than $\alpha_i(0)$). A similar forward sampling procedure can be derived for $S_{n,t}$, $t = 0, \ldots, T_n$. The details are omitted due to space limitations.

4.2 M-Step

In the M-step we update two parameters: π and Q. Updating π_i is straightforward:

$$\pi_i \leftarrow \frac{\sum_{n=1}^{N} p(\mathbf{O}, S_{n,0} = i; \pi', Q')}{\sum_{n=1}^{N} \sum_{i=1}^{M} p(\mathbf{O}, S_{n,0} = i; \pi', Q')}. \tag{11}$$

We update Q using the closed-form solution based on eigendecomposition introduced in [10]:

$$\begin{aligned} &Q_{ij} \leftarrow \frac{\sum_{\Delta} \sum_{k,l \in [M]} \mathbb{E}[\mathcal{N}_{ij}(\Delta)|S(\Delta) = l, S(0) = k; Q'] C_{k,l}(\Delta)}{\sum_{\Delta} \sum_{k,l \in [M]} \mathbb{E}[\mathcal{R}_i(\Delta)|S(\Delta) = l, S(0) = k; Q'] C_{k,l}(\Delta)} \\ &U \Lambda U^{-1} \leftarrow Q \text{ (eigendecomposition)} \\ &\chi_{pq}(\Delta) \leftarrow \begin{cases} \Delta \exp(\Delta \Lambda_p) & \Lambda_p = \Lambda_q \\ \frac{\exp(\Delta \Lambda_p) - \exp(\Delta \Lambda_q)}{\Lambda_p - \Lambda_q} & \Lambda_p \neq \Lambda_q \end{cases} \\ &\mathbb{E}[\mathcal{R}_i(\Delta)|S(\Delta) = l, S(0) = k; Q'] \\ &\quad \leftarrow \frac{1}{A_{kl}(\Delta)} \sum_{p=1}^{M} U_{kp} U_{pi}^{-1} \sum_{q=1}^{M} U_{iq} U_{ql}^{-1} \chi_{pq}(\Delta) \\ &\mathbb{E}[\mathcal{N}_{ij}(\Delta)|S(\Delta) = l, S(0) = k; Q'] \\ &\quad \leftarrow \frac{Q_{ij}}{A_{kl}(\Delta)} \sum_{p=1}^{M} U_{kp} U_{pi}^{-1} \sum_{q=1}^{M} U_{jq} U_{ql}^{-1} \chi_{pq}(\Delta) \end{aligned}$$

Recall that $A_{kl}(\Delta)$ is the transition probability from state k to l given the time span Δ (Eq. 1).

In our implementation, instead of returning to Gibbs sampling over $\mathbf{S}$ immediately after Q is updated, we first repeat the above procedure until Q converges (i.e., only recomputing expectations of $S(\tau)$ with respect to the new parameters), and then go back to the Gibbs sampling step. This is an approximation based on the assumption that a small change in Q will not lead to a significant change in the marginal distribution of $\mathbf{S}$.

4.3 Fast Evaluation of the Noisy-Or Network

In the Gibbs sampling step, we need to repeatedly evaluate the likelihood of the noisy-or network given the current parameter and variable assignments:

$$\prod_{d=1}^{D} \Phi(O_{d,n,t}|\{X_{k',n,t}\}_{k' \neq k}, X_{k,n,t} = i)),$$

$i \in \{0, 1\}$, for a given n and t.

To compute this likelihood in its original form, the complexity is proportional to D, the total number of findings, which is typically very large (tens of thousands for EHR

data). However, notice that we do not need to evaluate this over all possible findings:

$$\begin{aligned}
&\prod_{d=1}^{D} \Phi(O_{d,n,t}|\{X_{k',n,t}\}_{k'\neq k}, X_{k,n,t}=i)) \\
&= \prod_{d=1}^{D} \left(1-(1-L_d)\prod_{k=1}^{K}(1-X_{k,n,t}Z_{k,d})\right)^{O_{d,n,t}} \\
&\quad \left((1-L_d)\prod_{k=1}^{K}(1-X_{k,n,t}Z_{k,d})\right)^{1-O_{d,n,t}} \qquad (12) \\
&= \prod_{d:O_{d,n,t}=1} \left(1-(1-L_d)\prod_{k=1}^{K}(1-X_{k,n,t}Z_{k,d})\right) \\
&\quad \prod_{d:O_{d,n,t}=0} \left((1-L_d)\prod_{k=1}^{K}(1-X_{k,n,t}Z_{k,d})\right)
\end{aligned}$$

When we sample a specific $X_{k,n,t}$, most terms in the last line of the above equation, corresponding to the negative findings, cancel out. Consequently the ratio between the two likelihoods can be simplified to:

$$\begin{aligned}
&\frac{\prod_{d=1}^{D} \Phi(O_{d,n,t}|\{X_{k',n,t}\}_{k'\neq k}, X_{k,n,t}=1))}{\prod_{d=1}^{D} \Phi(O_{d,n,t}|\{X_{k',n,t}\}_{k'\neq k}, X_{k,n,t}=0))} \\
&= \prod_{d:O_{d,n,t}=0} (1-Z_{k,d}) \\
&\frac{\prod_{d:O_{d,n,t}=1}\left(1-(1-L_d)(1-Z_{k,d})\prod_{k'\neq k}(1-X_{k',n,t}Z_{k',d})\right)}{\prod_{d:O_{d,n,t}=1}\left(1-(1-L_d)\prod_{k'\neq k}(1-X_{k',n,t}Z_{k',d})\right)} \qquad (13)
\end{aligned}$$

With this improvement, the complexity of evaluating Φ is only proportional to the number of **positive** findings, which is far smaller than the total number of findings. More specifically, let N be the number of patients, M the number of progression states, T the number of visits (assuming each patient has the same number of visits), K the number of Comorbidities, and D the number of clinical findings. The cost of sampling $\mathbf{X}$ is reduced from $\mathcal{O}(NTK^2D)$ to $\mathcal{O}(NTK^2D^+)$, where $D^+ \ll D$ is the number of positive findings for an individual patient. The cost of sampling L, which benefits from the same optimization, is reduced from $\mathcal{O}(NTKD+NTMD)$ to $\mathcal{O}(NTKD^++NTMD)$; the cost of sampling Z is reduced from $\mathcal{O}(K^2D^2NT)$ to $\mathcal{O}(K^2D^+DNT)$. The cost of sampling B remains $\mathcal{O}(KMNT)$.

5. EMPIRICAL STUDY

In this section we apply our model to COPD, one of most prevalent chronic diseases worldwide. It is a particularly interesting target for a disease progression study because it has a prolonged progression path and is associated with a range of well-studied comorbidities [1,4]. Understanding the progression trajectory of COPD is crucial for early intervention and personalized care plan management.

We tested our model on a real COPD cohort to demonstrate that our model can automatically discover medically meaningful comorbidities among COPD patients. We present the COPD progression stages as identified by our model, which help explain the medical characteristics of COPD at different stages and the key medical events that trigger exacerbation. We showcase a typical application of the learned model, which is to infer the progression trajectories of individual patients based on their clinical observations. We also demonstrate the convergence behavior of our model and the importance of incorporating anchor findings.

Note that there are some existing clinical evaluation criteria for the progression/severity of COPD, such as GOLD and BODE index [6]. The progression trajectories defined by these criteria, which are mainly focused on the deterioration of respiratory function, are fundamentally different from the progression trajectories defined by our model, which are the onset patterns of various comorbidities. In practice our model is complementary to these existing standards in offering the medical practitioners a comprehensive characterization of the patients' condition.

5.1 Data Description

Our data came from a real-world longitudinal Electronic Medical Record (EMR) database of over 300,000 patients over the course of 4 years. For each patient encounter, a set of International Classification of Diseases - Version 9 (ICD-9) codes were recorded to indicate what medical conditions that patient had at that time point. Other information, such as drug prescription, lab test results, were also recorded. Note that in our database we did not have the lab measurements needed by either the GOLD or BODE criteria.

We first consulted our medical experts to identify the patients who were confirmed of having COPD at any time point during the record. The criteria we developed were the occurrence of at least one ICD-9 code that is related to COPD *and* the prescription of at least one COPD-related medication. Based on this selection criteria, we were able to identify 3,705 confirmed COPD patients from the database. For these patients, we extract their records in terms of ICD-9 codes. In total we had 5,263 unique ICD-9 codes associated with these patients. We removed infrequent ICD-9 codes that appeared for fewer than 200 patients, leaving us with 264 distinct codes (our findings). Considering the relative slow progression of COPD, we further decrease the time granularity of the original record. We segmented the time dimension into disjoint 90-day windows and combined all the observations within each window, viewing it as one encounter. In the end we had 34,976 encounters with 189,815 positive observations. Notice that the observations were very sparse: on average each patient had 10 encounters and 5.4 positive observations (ICD-9 code assignment) per encounter. The average record span of a patient was 816 days and the largest span was 1,094 days. Notice that this time span is significantly shorter than the span of regular COPD progression, which makes our setting very challenging.

5.2 Model Customization

Setting Anchors for Comorbidities: Our model infers the disease progression through the onsets of the comorbidities. Therefore the validity and relevance of the comorbidities have a great impact on the interpretability of our model. On the one hand, our model can be fully unsupervised and learn the comorbidities from the clinical findings. On the other hand, if we already have some domain knowledge on the important comorbidities, our model can incorporate such knowledge via setting anchor findings and produce comorbidities in alignment with the domain

Table 2: COPD comorbidities, associated medical conditions, and anchor ICD-9 diagnosis codes used.

Comorbidity	Representative Conditions (Anchor ICD-9 Codes)
COPD	Chronic Bronchitis (491), Emphysema (492, 518), Chronic Airway Obstruction (496)
Asthma	Asthma (493)
Cardiovascular	Hypertension (401), Congestive Heart Failure (428), Arrhythmia (427), Ischemic Heart Disease (414)
Lung Infection	Pneumonia (481, 485, 486)
Lung Cancer	Malignant Neoplasm of Upper/Lower Lobe, Bronchus or Lung (162)
Diabetes	Diabetes with Different Types and Complications (250)
Musculoskeletal	Spinal Disorders (724), Soft Tissue Disorders (729), Osteoporosis (733)
Kidney	Acute Kidney Failure (584), Chronic Kidney Disease (585), Renal Failure (586)
Psychological	Anxiety (300), Depression (296, 311)
Obesity	Morbid Obesity (278)

Table 3: Top-5 ICD-9 codes with highest activation probability of each comorbidity group as identified by our model. * means anchor findings.

COPD	**Asthma**
.45 Chronic Airway Obstruction*	.40 Asthma*
.05 Chronic Bronchitis with Exacerbation*	.07 Allergic Rhinitis
.04 Emphysema*	.05 Cough
.04 Chronic Bronchitis without Exacerbation*	.04 Acute Bronchitis
.03 Respiratory Abnormalities	.03 Acute Upper Respiratory Infections
Cardiovascular	**Lung Infection**
.30 Benign Essential Hypertension*	.30 Pneumonia*
.25 Unspecified Essential Hypertension*	.10 Shortness of Breath
.15 Atrial Fibrillation*	.10 Respiratory Abnormalities
.10 Congestive Heart Failure*	.10 Cough
.10 Hyperlipidemia	.06 Abnormal Findings of Lung
Lung Cancer	**Diabetes**
.60 Cancer of Bronchus and Lung, Unspecified*	.60 Type II Diabetes without Complication*
.15 Other Diseases of Lung	.15 Type II Diabetes without Complication, Uncontrolled*
.15 Cancer of Other Parts of Bronchus or Lung*	.10 Hyperlipidemia
.15 Cancer of Upper Lobe, Bronchus or Lung*	.06 Pure Hypercholesterolemia
.15 Swelling, Mass, or Lump in Chest	.06 Type II Diabetes with Renal Manifestations*
Musculoskeletal	**Kidney**
.15 Lumbago*	.20 Chronic Kidney Disease, Moderate*
.15 Pain in Limb*	.15 Anemia
.10 Osteoporosis*	.10 Chronic Kidney Disease, Unspecified*
.06 Myalgia and Myositis*	.08 Urinary Tract Infection
.04 Acquired Hypothyroidism	.08 Chronic Kidney Disease, Severe*
Psychological	**Obesity**
.15 Depression, Unspecified*	.25 Obesity, Unspecified*
.15 Anxiety*	.10 Morbid Obesity*
.04 Other Malaise and Fatigue	.03 Pure Hypercholesterolemia
.04 Major Depression, Recurrent Episode*	.02 Edema
.03 Major Depression, Single Episode*	.02 Sleep Apnea

Leak Terms: Hyperlipidemia .07, Need for Flu Vaccine .05, Acute Bronchitis .04, Routine Medical Examination .04, Acquired Hypothyroidism .04, Mammogram Screening .03, Actinic Keratosis .03, Urinary Tract Infection .03, Cough .03, Pure Hypercholesterolemia .03

knowledge. The study of COPD belongs to the latter case since COPD has some well-studied comorbidities, and discovering the progression of these designated comorbidities in terms of the progression of COPD is of great practical significance. We reviewed the literature with our medical experts and identified nine important COPD comorbidities (or comorbidity groups): asthma, cardiovascular diseases, lung infection, lung cancer, diabetes, musculoskeletal disorders, kidney diseases, psychological disorders, and obesity. In Table 2 we list these comorbidities along with the ICD-9 codes we used to anchor them. We also created an additional comorbidity to track the target disease itself, which was anchored by the ICD-9 codes we used to confirm COPD. Following the notions introduced in [7], if feature d is anchor of comorbidity k, we set $Z_{k',d}$ to $1e-6$ for any $k' \neq k$ (we chose this small value instead of 0 to avoid numerical underflow).

Constraining Progression Trajectory: For the results in the next section, we add a constraint to the transition generator matrix Q enforcing that we learn a linear pro-

Table 4: An example comorbidity identified without anchors. Although important conditions were identified, they were mixed together with no coherent medical interpretation (comparing to Table 3).

.35 Benign Essential Hypertension
.30 Type II Diabetes without Complication
.25 Mixed Hyperlipidemia
.20 Coronary Atherosclerosis Of Native Coronary Artery
.20 Pure Hypercholesterolemia
.15 Type II Diabetes without Complication, Uncontrolled
.08 Coronary Atherosclerosis Of Unspecified Type
.08 Chest Pain
.06 Anemia
.06 Chronic Ischemic Heart Disease, Unspecified

gression trajectory: we disallowed transitions from a later state to an earlier state by setting $Q_{ij} = 0$, $\forall j < i$. This constraint could be relaxed in many different ways, allowing us to learn a non-linear progression trajectory and/or multiple progression trajectories simultaneously. Such extensions will be discussed at the end of the paper.

Setting the Parameters: In order to learn a more sparse noisy-or network (thus better interpretability), we set the priors for Z to Beta$(0.1, 1)$. Thus our model will suppress the influence of findings that are loosely connected to the underlying comorbidities. We did not impose any such preference on L and B and set their priors to Beta$(1, 1)$. We sample Z from 0.01 to 0.1 with 0.01 increment and from 0.1 to 0.95 with 0.05 increment; we sample B and L from 0.01 to 0.99 with 0.01 increment. Since most of our model's comorbidities are chronic and their presence is expected to be persistent, we fix $B_{k,m,1} \triangleq p(X_{k,n,t} = 1 | S_{n,t} = m, X_{k,n,t-1} = 1) = 1$. This implies that once a comorbidity is turned on, it cannot be turned off in the future. We also introduce a monotonicity constraint on B^0 enforcing that the initial probability of comorbidity onset is higher for patients who at $t = 0$ are at a later progression stage than those at an earlier progression stage. After introducing this constraint, we update B^0 in the outer EM loop (together with π and Q) using a standard convex optimization procedure to maximize the likelihood subject to the monotonicity constraint. Finally, we set the number of states to $M = 6$ and the number of comorbidities to $K = 10$.

5.3 Results and Analysis

5.3.1 Comorbidities Identified

First we present the comorbidities as well as the leak term learned by our model (Table 3). For each comorbidity we list the top-5 ICD-9 codes in terms of activation/leak probability ($Z_{k,d}$ and L_d). All the comorbidities have clear and coherent clinical interpretations: the top findings either describe the main conditions corresponding to the comorbidity (e.g. Diabetes and Kidney), or are the common symptoms caused by those conditions (e.g., Lung Infection).

We contribute such interpretability to the anchor findings (starred entries in Table 3). As a comparison, we trained our model using the exact same settings but without the anchors. As a result, conditions from different comorbidities were mixed into one (see Table 4). This is natural from the model's perspective because these conditions frequently co-occur (thus the name comorbid) and no extra knowledge indicates they should be separated into different groups.

It is important to point out that our model was guided, but not fixed, by the anchor findings. For instance, there were anchor ICD-9 codes that did not appear in the top findings because the model did not consider them indicative enough; there were also some ICD-9 codes that were not anchored but ranked highly in relevant comorbidities (e.g. 'Lump in Chest' for Lung Cancer and 'Urinary Tract Infection' for Kidney). Rhinitis and Acute Bronchitis showed up in the Asthma group because they are among the top burdens of asthma patients.

5.3.2 Progression Trajectories Identified

Next we characterize the progression trajectories learned by our model from the COPD patient cohort. Given the generative nature of our model, we used it to generate 10,000 "virtual patients". We chose virtual patients over real patients because the patient records we had were only 4 years long and cannot cover the entire progression path of COPD. To generate a virtual patient, we assume the patient comes in at timestamp 0 in State 1. Then we used the learned model to generate the patient's subsequent states and the comorbidity onsets corresponding to those states until the patient reached State 6 (the last state). We averaged over the 10,000 records we generated and computed the average holding time for each state as well as the prevalence of different comorbidities at each state, as summarized in Figure 3.

Figure 3 gives us an intuitive view of how our trained model depicts the average progression trajectories of confirmed COPD patients. First, comparing different comorbidity groups across all progression stages, COPD has the highest prevalence (this is expected because we had a COPD case cohort); other highly prevalent comorbidities include cardiovascular diseases and musculoskeletal disorders; lung cancer is the rarest comorbidity. Next, we can observe that State 1 is the "healthy" state where the prevalence of all comorbidities are low. The transition from State 1 to 2 is driven by the onset of COPD, cardiovascular diseases, and musculoskeletal disorders. Starting from State 3, the prevalence of COPD is already 100% while the prevalence of other comorbidities rise steadily. The sudden increase in the onset rate of diabetes, kidney diseases, and lung infections at later stages of COPD conform well to previous studies in the medical literature. The onset rates of lung cancer, asthma, psychological disorders, and obesity are relatively stable across all progression stages.

5.3.3 Evidence Based Inference of Progression Trajectories

Next we showcase how to use our model to infer the progression trajectory of an individual patient based on the existing medical evidence. We picked two representative patients from our cohort and used MAP inference to infer their most likely progression stages and comorbidity onsets at each timestamp, respectively. Specifically, given the trained model, we repeatedly sample $\mathbf{X}$ and $\mathbf{S}$ conditioned on the observations from the query patient 1,000 times initialized by 10 different random seeds. We illustrated the results with maximum posterior probability in Figure 4.

We can see that patient (a) was correctly diagnosed by our model with musculoskeletal disorder and psychological disorder at the beginning and was assigned to the least se-

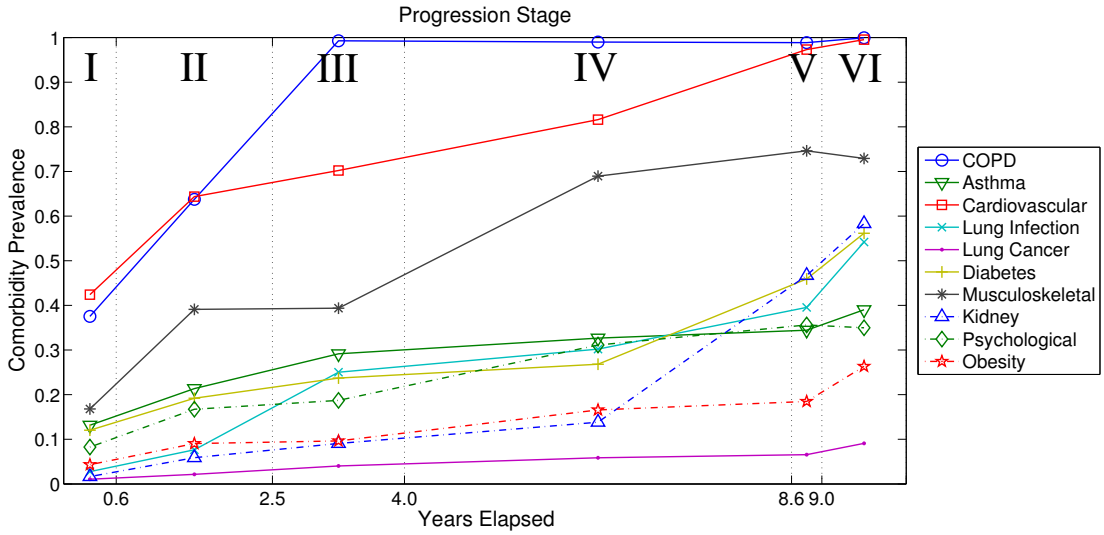
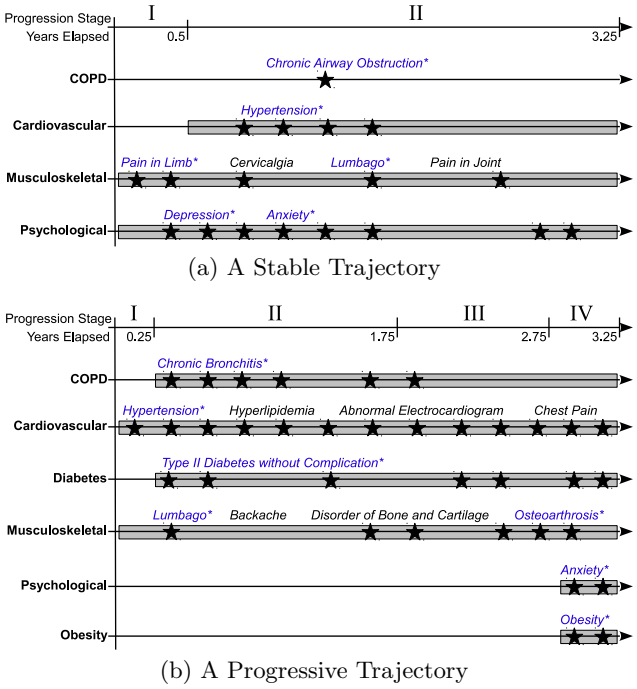


Figure 3: The change in comorbidity prevalence vs. the progression state transitions averaged over 10,000 virtual patients generated by our model.



(a) A Stable Trajectory

(b) A Progressive Trajectory

Figure 4: Inference of the progression trajectory and comorbidity onset of individual patients. Predicted stages are at the top of each plot; gray bars indicate predicted comorbidity onset; and stars denote the occurrence of relevant ICD-9 codes in the data (descriptions are on top of the timeline).

vere stage. After 6 months, our model diagnosed this patient with cardiovascular disease. The patient was subsequently moved into Stage 2 and stayed there throughout the record span. Note that the patient had a single occurrence of COPD related diagnosis code in the record. However,

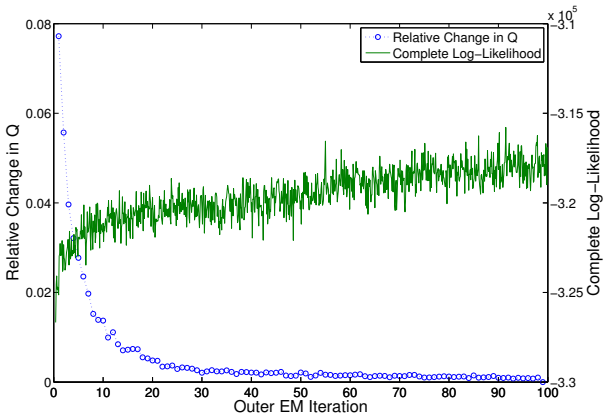


Figure 5: The convergence of our algorithm.

our model decided this was not enough evidence to turn on the COPD comorbidity; as a result, it was explained by the leak term (for all anchor findings we assigned a fixed leak probability of 10^{-6}). In contrast to patient (a), patient (b) exhibited a quick exacerbation of COPD. The patient started with cardiovascular disease and musculoskeletal disorder (Stage 1), followed by COPD and diabetes onset soon afterward (Stages 2 and 3), then eventually developed psychological disorder and obesity (Stage 4). We use these two representative patients to demonstrate the potential of incorporating our model into clinical decision support systems to provide medical practitioners with evidence based prediction and recommendation.

5.3.4 *Model Convergence and Runtime*

In Figure 5 we show the convergence behavior of the learning algorithm in terms of the relative change in Q as well as the complete data log-likelihood. We did 10 Gibbs sampling updates before first updating Q and π. We repeated the

outer EM iterations 100 times. Despite the small number of Gibbs sampling updates, our model converged quickly thanks to the use of anchors. We implemented our algorithm in Python. With parallelization and aforementioned optimizations, each full Gibbs sampling update took about 3 minutes on a 24-core 2.6 GHz machine.

6. DISCUSSION

The techniques presented in this paper open the door to many new applications of disease progression modeling. When paired with visualization tools, clinicians could use these models to better understand the evolution of chronic illnesses, optimize treatments, and design clinical guidelines.

Our empirical study focused on the problem of discovering a single disease progression model for COPD. Thus, we constrained the continuous-time Markov model to allow only forward transitions. However, for many diseases it is unlikely that there is a single disease trajectory that most patients follow, but rather a set of common trajectories. We can modify our algorithm to learn a *mixture* of disease trajectories simply by partitioning the state space and enforcing forward transitions within each partition.

Our proposed approach for learning disease progression models is flexible and can easily accommodate new sources of data and/or domain knowledge. For example, we could include lists of medications prescribed or procedures performed in the set of clinical findings to complement diagnosis codes. Because we use a generative model, it is trivial to marginalize over clinical findings that are unobserved on a specific visit. We could also introduce patient-specific global variables, either latent or observed. For example, the distribution over the initial disease state could be parameterized to be a function of age, gender, and family history. The rate matrix for the continuous-time Markov process could depend on factors such as whether the patient has stopped smoking, which might decrease the rate of progression. Going a step further, each specialist (e.g., endocrinologist versus a gynecologist) is likely to have a different distribution of diagnosis codes that they are likely to assign. Thus, if the specialist is provided in the data, the model could make the activation probabilities for clinical findings be different for each specialist.

An interesting question for follow-up work is whether it is possible to use non-case patients to learn disease progression models. In our experimental results we used only patients that were diagnosed with COPD. However, this limits the amount of training data. At a minimum, it may be possible to use the non-case patients to initialize the distribution of clinical findings for each comorbidity.

7. REFERENCES

[1] F. Baty, P. M. Putora, B. Isenring, T. Blum, and M. Brutsche. Comorbidities and burden of COPD: a population based case-control study. *PLoS ONE*, 8(5):e63285, 2013.

[2] M. Cohen, A. Grossman, D. Morabito, M. M. Knudson, A. Butte, and G. Manley. Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis. *Critical Care*, 14(1):R10, 2010.

[3] W. De Winter, J. DeJongh, T. Post, B. Ploeger, R. Urquhart, I. Moules, D. Eckland, and M. Danhof. A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying type 2 diabetes mellitus. *Journal of pharmacokinetics and pharmacodynamics*, 33(3):313–343, 2006.

[4] M. Decramer, W. Janssens, and M. Miravitlles. Chronic obstructive pulmonary disease. *Lancet*, 379(9823):1341–1351, Apr 2012.

[5] K. Exarchos, T. Exarchos, C. Bourantas, M. Papafaklis, K. Naka, L. Michalis, O. Parodi, and D. Fotiadis. Prediction of coronary atherosclerosis progression using dynamic bayesian networks. In *EMBC*, pages 3889–3892, 2013.

[6] G. I. for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of copd, January 2014.

[7] Y. Halpern and D. Sontag. Unsupervised learning of noisy-or bayesian networks. In *UAI*, pages 272–281, 2013.

[8] K. Ito, S. Ahadieh, B. Corrigan, J. French, T. Fullerton, and T. Tensfeldt. Disease progression meta-analysis model in alzheimer's disease. *Alzheimer's & Dementia*, 6(1):39–53, 2010.

[9] C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy, and E. Couto. Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(2):193–209, 2003.

[10] P. Metzner, I. Horenko, and C. Schütte. Generator estimation of markov jump processes based on incomplete observations nonequidistant in time. *Physical Review E*, 76(6):066702, 2007.

[11] D. Mould. Models for disease progression: new approaches and uses. *Clinical Pharmacology & Therapeutics*, 92(1):125–131, 2012.

[12] T. M. Post, J. I. Freijer, J. DeJongh, and M. Danhof. Disease system analysis: basic disease progression models in degenerative disease. *Pharmaceutical research*, 22(7):1038–1049, 2005.

[13] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[14] M. A. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. Probabilistic diagnosis using a reformulation of the internist-1/qmr knowledge base. *Meth. Inform. Med*, 30:241–255, 1991.

[15] R. Sukkar, E. Katz, Y. Zhang, D. Raunig, and B. T. Wyman. Disease progression modeling using hidden markov models. In *EMBC*, pages 2845–2848. IEEE, 2012.

[16] J. Yang, J. J. McAuley, J. Leskovec, P. LePendu, and N. Shah. Finding progression stages in time-evolving event sequences. In *WWW*, pages 783–794, 2014.

[17] J. Zhou, J. Liu, V. A. Narayan, and J. Ye. Modeling disease progression via fused sparse group lasso. In *KDD*, pages 1095–1103. ACM, 2012.

[18] J. Zhou, J. Liu, V. A. Narayan, and J. Ye. Modeling disease progression via multi-task learning. *NeuroImage*, 78(0):233 – 248, 2013.