Interacting Dynamic Processes for Social Network Data

Katherine Heller
Center for Cognitive Neuroscience
Department of Statistical Science

Department of Computer Science
Department of Neurobiology
Department of Neurology
My Research Areas

- Machine Learning for Healthcare
- Social Network Modeling
- Jointly Modeling Diverse Data Sets
  - Neuroscience
  - Chronic Kidney Disease
- Bayesian time series models

- Cognitive Science Applications with Bayesian Nonparametrics
- Bayesian Clustering
- Topic/language modeling
Social Interactions

- Fundamental to understanding human behavior.
- Very complex.
- Ideal for interdisciplinary research.
Outline

- Hawkes Processes and the Bayesian Echo Chamber
- Modeling the Spread of Disease with Graph-coupled HMMs
- Other Machine Learning for Healthcare Work
People organize into social groups.

We’d like to know the dynamics of these groups.

But these groups are not explicitly defined.

- Can use declared relationships (e.g. fb friends) to infer groups.
- However not always available, truthful, or useful.

We infer groups from real interactions.

- Data is a sequence of many events - actions from a sender to a receiver.
- Leverage reciprocity patterns to discover social groups.
Reciprocating Relationships
The Infinite Relational Model (IRM) is a method for clustering entities based on graphs of declared relationships. Kemp et al. developed this model, which incorporates probabilistic graphical models to represent the relationships between entities.
Poisson Process IRM

Often interaction data contains many interactions between the same pair of individuals.

Cannot be modeled with a vanilla IRM.

Can modify to use Gamma-Poisson observation model.

But then we cannot predict events into the future.

Therefore we consider a Poisson process.

\[
\pi \sim \text{CRP}(\alpha) \\
\lambda_{pq} \sim \text{Gamma}(\delta, \beta) \\
N_{uv}(\cdot) \sim \text{PoissonProcess}(\lambda_{\pi(u)\pi(v)})
\]
Hawkes Processes

**Drawbacks** of using the Poisson Process:
- Rate of events between pairs of clusters independent of all other pairs.
- Times of events are uniformly distributed.

Instead use a mutually-exciting Hawkes process:

\[ \lambda_{pq}(t) = \gamma_{pq} + \sum_{i: t_{i}^{qp} < t} g_{pq}(t - t_{i}^{qp}), \quad g_{pq}(\delta) = \beta_{pq} e^{-\frac{\delta}{\tau_{pq}}} \]
Multivariante Hawkes Process

Diagram showing the interaction between multiple processes, indicated by arrows and lines.
Inference

- Posterior inference is performed using MCMC.

- Unlike previous IRM models there is no conjugate prior for the likelihood.
  - Must sample parameters instead of integrating out.

- We infer the partition of entities, CRP concentration parameter, and Hawkes process parameters using Metropolis within Gibbs and Slice sampling.
  - Basically Neal’s algorithm 5 with some additional slice sampling.
**Experiments: Correlates of War**

- Militarized Interstate Disputes data set which captures correlates of war.
- Data: Years 1993-2001, all MID incidents and countries involved.
- Incidents vary from diplomatic threats to deployment of force.

- 3 main conflicts
  - Russia and Afghanistan
  - Taiwan and China
  - USA, Iraq and Kuwait

Blundell et al, NIPS 2012
Experiments:

- Log predictive probabilities of events falling in the last 10% of time.
- Also Enron data – 5 longest threads.

<table>
<thead>
<tr>
<th></th>
<th>Hawkes IRM</th>
<th>Poisson IRM</th>
<th>Hawkes</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic</td>
<td>23.49±0.03</td>
<td>20.45±0.04</td>
<td>17.73±0.01</td>
<td>17.76±0.00</td>
</tr>
<tr>
<td>Small MID</td>
<td>10.21±0.26</td>
<td>-1.81±0.11</td>
<td>0.18±0.08</td>
<td>-2.30±0.07</td>
</tr>
<tr>
<td>Full MID</td>
<td>-127.45±0.45</td>
<td>-132.71±0.09</td>
<td>-188.29±0.02</td>
<td>-188.44±0.01</td>
</tr>
<tr>
<td>Enron 0</td>
<td>220.16±0.04</td>
<td>162.44±0.06</td>
<td>194.34±0.02</td>
<td>133.02±0.07</td>
</tr>
<tr>
<td>Enron 1</td>
<td>335.88±0.04</td>
<td>304.89±0.11</td>
<td>284.71±0.05</td>
<td>256.62±0.11</td>
</tr>
<tr>
<td>Enron 2</td>
<td>108.21±0.02</td>
<td>104.79±0.05</td>
<td>87.49±0.02</td>
<td>85.57±0.05</td>
</tr>
<tr>
<td>Enron 3</td>
<td>101.32±0.04</td>
<td>98.90±0.16</td>
<td>82.31±0.04</td>
<td>81.62±0.05</td>
</tr>
<tr>
<td>Enron 4</td>
<td>121.07±0.07</td>
<td>113.25±0.08</td>
<td>100.16±0.04</td>
<td>91.70±0.08</td>
</tr>
<tr>
<td>SB conv 23</td>
<td>127.91±0.28</td>
<td>47.39±0.10</td>
<td>26.00±0.01</td>
<td>19.63±0.02</td>
</tr>
<tr>
<td>SB conv 26</td>
<td>-6.19±0.04</td>
<td>-12.38±0.03</td>
<td>-7.15±0.03</td>
<td>-10.47±0.03</td>
</tr>
<tr>
<td>SB conv 12</td>
<td>53.23±0.38</td>
<td>45.67±0.12</td>
<td>17.44±0.06</td>
<td>-8.69±0.12</td>
</tr>
<tr>
<td>SB conv 49</td>
<td>199.27±0.30</td>
<td>128.05±0.13</td>
<td>125.30±0.02</td>
<td>122.44±0.03</td>
</tr>
<tr>
<td>SB conv 33</td>
<td>31.86±0.27</td>
<td>2.39±0.08</td>
<td>21.75±0.01</td>
<td>21.63±0.01</td>
</tr>
</tbody>
</table>
Bayesian Echo Chamber

Look at how people influence each other.
Temporal Influence

\[ N^p(a, b] \sim \text{Pois} \left( \int_a^b \lambda_0^p + \sum_{q \neq p} \sum_{t_n^q < t} \nu^q \exp \left( \frac{-(t - t_n^q)}{\tau^p} \right) dt \right) \]
Linguistic Accommodation

“During communication encounters, people will try to accommodate or adjust their style of speaking to others. [...] Convergence occurs when there is a strong need for social approval, frequently from powerless individuals.”

— West & Turner, 2010
Mutually Exciting Language Models

[Diagram showing the interaction between sloths and another term, possibly related to parkour.]
For each person...

\[ m_t^p \propto \mu^p + \sum_{q \neq p} \rho^{qp} \sum_{t_n^q < t} N^{t_n^q} \cdot \exp\left( \frac{-(t - t_n^q)}{\tau^p} \right) \]

- **language profile**
- **pairwise influence**
- sum over all other people
- word counts
Linguistic Influence

\[ m_t^p \propto \mu^p + \sum_{q \neq p} \sum_{t_n^q < t} \rho^{qp} N_{t_n^q} \cdot \exp \left( \frac{-(t - t_n^q)}{\tau^p} \right) \]
DC v Heller: Temporal Influence

Guo et al, AISTATS 2015
DC v Heller: Linguistic Influence

Guo et al, AISTATS 2015
12 Angry Men
Federal Reserve Board Meetings

Guo et al, AISTATS 2015
Relating Time and Content

Allow the amount of excitation to be dependent on content in an event based Hawkes process with IRM.

\[ \beta_{uv}(s) = e^{r_u(x_{vu}(s)) + s_v(x_{vu}(s))} \]

where

\[ r_u(\cdot) \sim GP(0, k_r) \]
\[ s_v(\cdot) \sim GP(0, k_s) \]

Results in improved log likelihood on held out test data (enron, santa barbara conversation corpus, citation), and more intuitive clusters.

Tan et al, UAI 2016
Extensions

- Learning clusterings of neurons. Similar processes are used in neuroscience to model the activation and co-activation of neurons. These usually also involve inhibition, but don’t directly cluster.

- Scale up inference

- Deal with unknown recipients

- Online bullying
Infection in a Social Network

**Goal**: To model dynamical interactions between agents in a social network and apply to inferring the spread of infection.

Many traditional epidemics models work on a population level, treating each person the same way.

Contemporary data collection techniques allow us to model the spread of infection on an individual level.

Being able to make infection predictions on an individual level is enormously beneficial because it allows people to receive more personalized and relevant health advice.
Social Evolution Experiment

Data collected in the social evolution experiment allows us for the first time to closely track proximities and contagion in an entire community over a substantial period of time.

- Tracked “common cold” symptoms in an MIT residence hall from January to April 2009.
- Monitored over 80% of residents through their cell phones from October 2008 to May 2009, taking daily surveys and tracking their location, proximities and phone calls.
- Monthly surveys on social, health, and political issues taken. Locations taken by having cell phones scan nearby wifi access points and bluetooth devices.
In the Social Evolution experiment students were paid $1 a day to answer surveys about contracting infection.

The surveys asked about symptoms:
- Runny nose, nasal congestion, sneezing
- Nausea, vomiting, diarrhea
- Stress
- Sadness and depression
- Fever

64 of the 85 residents answered the surveys

Symptoms dependent on the social network. A student with a symptom had a 3-10x higher odds of having a friend with the same symptom.
We aim to leverage the social evolution data to predict the spread of infection to individuals in our social network using a model related to HMMs.
Graph-coupled HMMs

Associate each person in the dynamic interaction network with an HMM chain. Let interaction network structure determine the HMM couplings:

\[ X_{n,t} \sim \text{Categorical}(\phi_{n}, X_{e:\{n,\cdot\} \in G_{t}, t-1}) \]

\[ Y_{n,t} \sim F(\theta_{X_{n,t}}) \]

\[ \theta_{X_{n}} \sim \text{Conj}(\gamma) \]

\[ \phi_{n,X_{e:\{n,\cdot\} \in G_{t}}} \sim H(X_{e:\{n,\cdot\} \in G_{t}, \mu}) \]
GCHMM Inference

- Inference in the coupled HMM is very hard.
  - Typically ML estimation is done on few chains with few states or another approximation is made.

- In the worst case GCHMM inference is as difficult as CHMM inference.
  - When the graph is fully connected.

- Fortunately, since we’re dealing with social networks we can leverage a couple of properties:
  - Social networks are usually sparse
  - For many applications the influence of interactions can be modeled in a fairly simple way via a small number of parameters.
GCHMMs for Modeling Infection

In the case of the social evolution data the influence of other HMMs can be summarized by counts of interactions in the infectious state.

The GCHMM can provide an individual level version of the susceptible-infectious-susceptible (SIS) epidemiology model:

\[
\begin{align*}
\dot{S} &= -\beta \cdot SI + \gamma \cdot I \\
\dot{I} &= \beta \cdot SI - \gamma \cdot I
\end{align*}
\]

GCHMM for infection:

\[
\begin{align*}
X_{n,t} &\sim \text{Bernoulli}(\phi_{n, x_{e:\{n,\} \in G_t, t-1}}) \\
Y_{n,t,i} &\sim \text{Bernoulli}(\theta_{X_{n,t}}) \\
\theta_{X_n} &\sim \text{Beta}(\phi) \\
\alpha &\sim \text{Beta}(\alpha',\beta') \\
\beta &\sim \text{Beta}(\alpha'',\beta'') \\
\gamma &\sim \text{Beta}(\alpha''',\beta''')
\end{align*}
\]

\[
p(X_{n,t+1} = 0 | X_{n,t} = 1) = \gamma \\
p(X_{n,t+1} = 0 | X_{n,t} = 0, X_e:\{n, \} \in G_t) = (1 - \alpha)(1 - \beta)\sum_{e:\{n, \} \in G_t} X_{ne,t}
\]
Experimental Results

Dong et al, UAI 2012
Aiello Group Data

- eX-FLU study at University of Michigan
  - 590 students from 6 dorms
  - Chain referral scheme

- A 103 student subset participated in iEpi
  - Smartphone based study where location is tracked and surveys taken

- Unlike MIT study confirmation of interaction was recorded on phones and flu testing was done on students who reported being ill.

- Also an isolation intervention was tested.
Hierarchical GCHMMs

Add a hierarchical level to where beta distributed infection parameters are learned:

Beta-exponential link

\[ \eta_{., \cdot} \sim N(\mu, \Sigma) \]
\[ \gamma_n \sim \text{Beta}(\exp(z_n^T \eta_{r,1}), \exp(z_n^T \eta_{r,2})) \]
\[ \alpha_n \sim \text{Beta}(\exp(z_n^T \eta_{a,1}), \exp(z_n^T \eta_{a,2})) \]
\[ \beta_n \sim \text{Beta}(\exp(z_n^T \eta_{b,1}), \exp(z_n^T \eta_{b,2})) \]

Sigmoid link

\[ \eta \sim N(\mu, \Sigma) \]
\[ \gamma_n = \sigma(z_n^T \eta_r), \quad \alpha_n = \sigma(z_n^T \eta_a), \quad \beta_n = \sigma(z_n^T \eta_b) \]

Inference Gibbs-EM algorithm but follow up papers e.g. stochastic VB
Results

Fan et al, KDD 2016

Table 2: Coefficients Estimation on exFlu Dataset

<table>
<thead>
<tr>
<th>Feature</th>
<th>Recovery $\eta_r$</th>
<th>Outside Infect $\eta_o$</th>
<th>Inside Infect $\eta_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default=1</td>
<td>-1.3022 ± 0.0146</td>
<td>-5.1517 ± 0.0024</td>
<td>-4.1619 ± 0.0281</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.1575 ± 0.0118</td>
<td>-0.2428 ± 0.0074</td>
<td>-0.1457 ± 0.0078</td>
</tr>
<tr>
<td>Age</td>
<td>0.0074 ± 0.0082</td>
<td>-0.2376 ± 0.0051</td>
<td>-0.0181 ± 0.0017</td>
</tr>
<tr>
<td>Alc_Day</td>
<td>0.1090 ± 0.0078</td>
<td>-0.1534 ± 0.0003</td>
<td>-0.0410 ± 0.0018</td>
</tr>
<tr>
<td>Vacc_Ever</td>
<td>-0.0698 ± 0.0104</td>
<td>0.1092 ± 0.0095</td>
<td>0.0382 ± 0.0085</td>
</tr>
<tr>
<td>Flushot_Yr</td>
<td>0.0769 ± 0.0092</td>
<td>-0.3209 ± 0.0073</td>
<td>0.0837 ± 0.0055</td>
</tr>
<tr>
<td>Smoker</td>
<td>-0.1080 ± 0.0029</td>
<td>-0.0536 ± 0.0008</td>
<td>0.0773 ± 0.0021</td>
</tr>
<tr>
<td>Drinker</td>
<td>-0.1335 ± 0.0092</td>
<td>0.0628 ± 0.0030</td>
<td>0.1408 ± 0.0029</td>
</tr>
<tr>
<td>Act_Days</td>
<td>0.0356 ± 0.0099</td>
<td>0.0054 ± 0.0063</td>
<td>-0.0622 ± 0.0078</td>
</tr>
<tr>
<td>Sleep_Qual</td>
<td>0.0225 ± 0.0069</td>
<td>-0.3686 ± 0.0051</td>
<td>-0.0162 ± 0.0077</td>
</tr>
<tr>
<td>Wash_Opt</td>
<td>0.0024 ± 0.0103</td>
<td>0.0816 ± 0.0132</td>
<td>-0.0714 ± 0.0048</td>
</tr>
<tr>
<td>High_Risk</td>
<td>-0.1274 ± 0.0116</td>
<td>-0.1252 ± 0.0058</td>
<td>-0.0727 ± 0.0007</td>
</tr>
</tbody>
</table>
Extensions

- Other applications of the GCHMM to network data, like modeling the influence of opinions.
- Scale up to larger communities (Fan et AAAI 2016)
- Nonparametric Bayesian extensions
- Other factors that may influence contracting an infection
- Learn latent network structure
More ML for Healthcare Work
Kidney Function (eGFR)

Age

eGFR

1\textsuperscript{st} Nephrology Visit
Nephrology Visit
PCP Visit
ED Visit
Hospital Admission
Acute MI
Death
Age 47

Untreated diabetes & high blood pressure.

Normal kidney function, but with evidence of kidney damage.

No regular medical care.
Age 49
Kidney function now 50%
Age 51
Referred to kidney specialist
Three months later presents to ER with kidney failure symptoms and “crash starts” dialysis.
Missed Opportunities:

To prevent or delay kidney failure

To prepare for kidney failure
<10% with moderate CKD
<50% with severe CKD
even aware of illness!
Model for a single trajectory

Conditional likelihood factorizes across $P$ labs:

$$p(y_i | z_i, b_i, c_i; x_i) = \prod_{p=1}^{P} p(y_{ip} | z_i, b_i, c_i; x_i)$$

$$y_{ip}(t) \sim N(\mu_{ip}(t), \sigma^2_p)$$

$$\mu_{ip}(t) \sim GP(\Lambda^{(p)} x_i + \Phi_z(t)^\top \beta_{zip} + \Phi_l(t)^\top b_{ip}, K_p)$$

Population effect

Latent subpopulation curve

Individual long-term deviations

Individual transient deviations (GP)

$$K_p(t, t') = a_p^2 \exp\{-l_p^{-1}|t - t'|\}$$

Futoma et al, MLHC 2016
Chronic Kidney Disease (CKD)

Heart disease

Diabetes
Proposed Joint Model

- Goal: jointly model risks of future loss of kidney function, cardiac events
  - Heart attacks (AMI), Stroke (CVA)
- Hierarchical latent variable model: captures dependencies between disease trajectory and event risk
  - Submodels for longitudinal, event data with shared latent variables
- $\tilde{y}_i$: eGFRs at times $\tilde{t}_i$; $\tilde{u}_i$: event times (may be none); covariates
- Conditional independence in joint likelihood:
  $$p(\tilde{y}_i, \tilde{u}_i | z_i, b_i, f_i, v_i; x_i) = p(\tilde{y}_i | z_i, b_i, f_i; x_i)p(\tilde{u}_i | z_i, b_i, f_i, v_i; x_i)$$

Futoma et al, UAI 2016
surgical complications

approximately 15 out of every 100 surgical procedures performed results in a complication
**Prediction**: learning relationship between predictors & outcomes

<table>
<thead>
<tr>
<th>OUTCOME</th>
<th>most predictive variables</th>
<th>% risk increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-day Mortality</td>
<td>ASA Class 5</td>
<td>55%</td>
</tr>
<tr>
<td></td>
<td>Totally dependent functional status</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>Preoperative septic shock</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>DNR status</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>Preoperative ventilator dependence</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>Liver disease (varices or ascites)</td>
<td>9%</td>
</tr>
<tr>
<td>30-day Any Morbidity</td>
<td>Dx-Esophageal cancer</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>Totally dependent functional status</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>Preoperative septic shock</td>
<td>24%</td>
</tr>
<tr>
<td></td>
<td>Dx-Nutritional deficiency</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>Dx-Injury</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>ASA Class 4</td>
<td>19%</td>
</tr>
</tbody>
</table>
CONTINUOUS LEARNING

<table>
<thead>
<tr>
<th>Outcomes (2 of 8)</th>
<th>AUC - NoTransferLearning</th>
<th>AUC - TransferLearning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumonia</td>
<td>0.832</td>
<td>0.848</td>
</tr>
<tr>
<td>Cardiac</td>
<td>0.909</td>
<td>0.920</td>
</tr>
</tbody>
</table>
MS Mosaic App

Passively collects HealthKit data

82 Different Data Types (>10 thought to affect MS)

Tagged with date, time, and provenance
Top 3 most severe symptoms:

**Weakness**
Avg. of 4.2 on severity scale
Learn more about weakness

**Fatigue**
Avg. of 3.8 on severity scale
Learn more about fatigue

**Cramps**
Avg. of 3.5 on severity scale
Learn more about cramps

Top changes in symptoms (since previous week):
Initial Analyses

Develop a sparse logistic regression model for predicting the likelihood of each symptom experience

Incorporate a hierarchical layer based on Gaussian processes for modeling time series data (e.g. sleep)

Discover hidden subpopulations within symptoms (using clustering methodology, such as Dirichlet Process mixture models)

Evaluate the efficacy of symptom interventions using longitudinal models and clinical trials
Future Work

- Further work in machine learning for health care
  - Sepsis
  - Congestive Heart Failure
  - Surgical Transplants

- Implement infectious disease work locally
  - Develop models for data from more diverse sensors
  - Causal Inference

- Additional sensors on other groups related to physiometrics
  - Does heart rate impact basketball players shot percentage

- Other social network applications – like online bullying

- Other joint modeling applications – measuring diverse data sets from e.g. different sensory modalities (and animals) to infer functional neural networks in the brain.
Future Work

- Role of other media in ability to make health predictions
  - Xbox
  - Cortana
  - Bing

- Incorporation of Microsoft Health devices into research
  - Microsoft Band
  - Development of machine learning for Microsoft apps
  - HealthVault Insights

- Incorporation of behavior and choice into patient feedback
  - Medication Adherence
  - Exercise
  - Health bot

- Joint modeling of Microsoft data with other data (e.g. NIH)
<table>
<thead>
<tr>
<th>Collaborators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charles Blundell</td>
</tr>
<tr>
<td>Kai Fan</td>
</tr>
<tr>
<td>Joseph Futoma</td>
</tr>
<tr>
<td>Hanna Wallach</td>
</tr>
<tr>
<td>Allison Aiello</td>
</tr>
<tr>
<td>Elizabeth Lorenzi</td>
</tr>
<tr>
<td>Richard Guo</td>
</tr>
<tr>
<td>Marissa Eisenberg</td>
</tr>
<tr>
<td>Blake Cameron</td>
</tr>
<tr>
<td>Alan Qi</td>
</tr>
<tr>
<td>Alison Walsh</td>
</tr>
<tr>
<td>Mark Sendak</td>
</tr>
<tr>
<td>Vinayak Rao</td>
</tr>
<tr>
<td>Sandy Pentland</td>
</tr>
<tr>
<td>Lee Hartsell</td>
</tr>
<tr>
<td>Xi Tan</td>
</tr>
<tr>
<td>Wen Dong</td>
</tr>
<tr>
<td>Erich Huang</td>
</tr>
<tr>
<td>Syed Naqvi</td>
</tr>
<tr>
<td>Jerry Zhu</td>
</tr>
<tr>
<td>Jeff Sun</td>
</tr>
<tr>
<td>Wei Zhang</td>
</tr>
<tr>
<td>Sanjay Hariharan</td>
</tr>
<tr>
<td>Fan Bu</td>
</tr>
</tbody>
</table>