Project description

Filipe Rodrigues

Francisco Pereira

DTU Management Engineering
Department of Management Engineering

Outline

- Rules
- Important dates
- Common mistakes, misconceptions and advice
- Great projects from last year
- Take-away ideas

Rules

- Topic is **free** (creativity highly encouraged!)
- No constraints on the dataset
- If in doubt, talk with us!
- We also provide some suggestions for projects

Important dates

Important dates

- April 9 - Milestone 1
  - Dataset and research question(s)
  - First draft of model(s) that you plan to try out (PGM + generative process)
  - Initial notebook (descriptive stats, data preparation)

Important dates

- April 9 - Milestone 1
  - Dataset and research question(s)
  - First draft of model(s) that you plan to try out (PGM + generative process)
  - Initial notebook (descriptive stats, data preparation)
- May 14 - Final delivery (Part I)
  - Fully self-explanatory notebook
- May 17 - Final delivery (Part II)
  - 2-page report (excluding figures and tables)

Common mistakes, misconceptions and advice

- No point putting priors on observed variables without any parents, unless doing imputation, otherwise these variables are always given

Common mistakes, misconceptions and advice

- No point putting priors on observed variables without any parents, unless doing imputation, otherwise these variables are always given
- Makes no sense to include observed variables that block the information path in the PGM
  - Consider for example: Intelligence $\rightarrow$ Course grade $\rightarrow$ Recommendation letter
  - If course grade is always observed, then it blocks the path between intelligence and recommendation letter. Therefore, "Intelligence $\rightarrow$ Course grade" and "Course grade $\rightarrow$ Recommendation letter" become two independent models.
  - There is no point estimating them jointly because there is no flow of information between the two (unless there is an alternative path, of course...)

Common mistakes, misconceptions and advice

- No point putting priors on observed variables without any parents, unless doing imputation, otherwise these variables are always given
- Makes no sense to include observed variables that block the information path in the PGM
  - Consider for example: Intelligence $\rightarrow$ Course grade $\rightarrow$ Recommendation letter
  - If course grade is always observed, then it blocks the path between intelligence and recommendation letter. Therefore, "Intelligence $\rightarrow$ Course grade" and "Course grade $\rightarrow$ Recommendation letter" become two independent models.
  - There is no point estimating them jointly because there is no flow of information between the two (unless there is an alternative path, of course...)
- Don't just think about what variables depend on other variables, and their distribution types. Think also of how you should model those dependencies (e.g. how to condition the parameters of a Beta distribution on another variable?)

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations
- Try different types of inference algorithms: VI and MCMC, and make sure they are doing the right thing and you can trust the results

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations
- Try different types of inference algorithms: VI and MCMC, and make sure they are doing the right thing and you can trust the results
- It is a generally good idea to start with a simple model and incrementally make it more complex towards the idealised/conceived PGM. Plus, try to have some baseline models for comparison whenever possible

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations
- Try different types of inference algorithms: VI and MCMC, and make sure they are doing the right thing and you can trust the results
- It is a generally good idea to start with a simple model and incrementally make it more complex towards the idealised/conceived PGM. Plus, try to have some baseline models for comparison whenever possible
- Use ancestral sampling to generate artificial data, and run inference on the model using Pyro/STAN to see if it is able to recover the true values/parameters that were used to generate the data. This is a great way of guarantying that the model is correctly implemented, and that inference is working correctly

2019 proj.: Analysis group membership in Social Networks I

- Graph data of human interaction is called social networks. The interaction could for example being friends on Facebook or talking to each other.
- The project inferred groups in social networks based on Game Of Thrones. Take a look at the data here https://github.com/mathbeveridge/gameofthrones
- Tried different models. Results below is from a Mixed membership stochastic block model - MMSBM.

2019 proj.: Analysis group membership in Social Networks II

1. For $n = 1..N$
   1. draw $\boldsymbol{\pi}_n \sim \text{Dirichlet}(\boldsymbol{\alpha})$
2. For $k, k' = \{1, 2, \dots, K\} \times \{1, 2, \dots, K\}$ where $k \neq k'$
   1. draw $\lambda_{k,k'} \sim \text{Gamma}(a, b)$
3. For k in $1..K$
   1. draw $\lambda_{k,k} \sim \text{Gamma}(c, d)$
4. For all pairs $i, j \in \{1, 2, ...N\} \times \{1, 2, ...N\}$ where $i \neq j$
   1. $x_{i,j} \sim \text{Poisson}(\boldsymbol{\pi}_i^T \lambda \boldsymbol{\pi}_j)$;

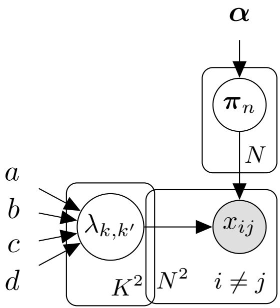


Figure: PGM for Mixed membership stochastic block model

2019 proj.: Analysis group membership in Social Networks III

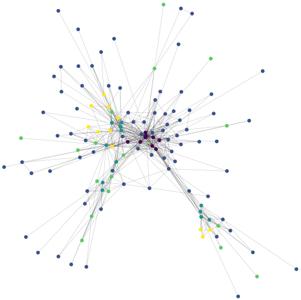


Figure: Each node is a GoT character from season 5. Colors indicate one of 5 inferred groups.

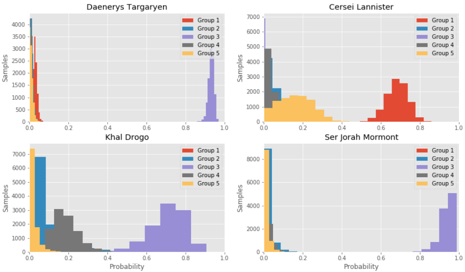


Figure: The histogram shows posterior probabilities of belonging to one of 5 groups for 4 main characters of GoT

2019 project: Bayesian semi-parametric soccer analysis

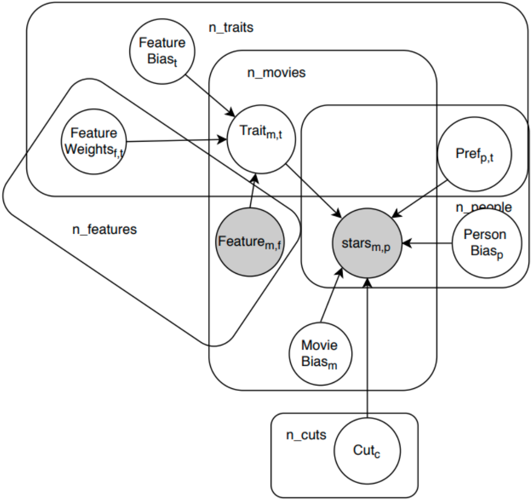
- Apply the time-to-event analysis methods to analyze the intensity with which goal scoring occurs in soccer matches
- Tried different time-to-event models from the class of Cox proportional hazards models and extend these to incorporate non-linear effects by use of the highly flexible Gaussian processes from the Bayesian non-parametrics toolbox
- The hazard in the i-th match at time interval j is

$$\lambda_{ij} = \lambda_j \exp(\beta_1 * skill\_gap + \beta_2(t) * x\_goal)$$

- Baseline of Cox model is a GP: $\log \lambda_0 \sim \mathcal{GP}(\mathbf{0}, K(t, t'))$
- Implemented everything in STAN and did a very good experimental evaluation

2019 project: Movie recommendations - MovieLens dataset

- Inspired by the example in Bishop's "Model-based Machine Learning" book, but with a good mix of creativity, lots of experimentation and insightful discussions



Take-away ideas

- Kaggle datasets
- Context-aware Bayesian Choice Models
- Bayesian Spatial Count Models

Kaggle datasets

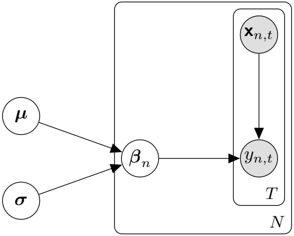
- Find a cool/crazy dataset on Kaggle - e.g. Avocado Prices (`https://www.kaggle.com/neuromusic/avocado-prices`)
- Formulate a research question(s) - e.g. how do the temporal dynamics of price change across regions
- Propose a PGM to model the data based on domain knowledge and assumptions of research question - e.g. hierarchical temporal model
- Fit model, extract results, revise model and data, propose model extensions, formulate new research questions... and iterate...

Bayesian Choice Models

- Choice models allow us to understand how people make choices and helps us come up with policy interventions (e.g. for promoting the use of bicycles)

Bayesian Choice Models

- Choice models allow us to understand how people make choices and helps us come up with policy interventions (e.g. for promoting the use of bicycles)
- Classical discrete choice modelling approach (mixed logit models):



N is the number of individuals; T is the number of choice situations

- Observed choices $y \sim \text{Categorical}(y|\text{Softmax}(u_1, \dots, u_J))$
- Utility function of choice alternative $j \in \{1, \dots, J\}$ is a linear parametric function of alternative attributes $\mathbf{x}_j$:

$$u_j = \text{ASC}_j + \beta_{\text{cost}} * \text{cost}_j + \beta_{\text{time}} * \text{travel\_time}_j + \dots$$

- Each individual n has his/her own preference parameters $\boldsymbol{\beta}_n$

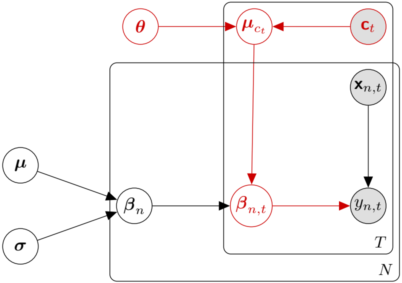
Context-aware Bayesian Choice Models

- Is it reasonable to assume that an individual's preference parameters $\boldsymbol{\beta}_n$ are independent of the context?
  - Maybe you don't care that going by bike takes longer than by car if the weather is good
  - So the value of β_{time} can vary depending on the weather (context)

[1] Reference: `https://arxiv.org/abs/2004.05426`

Context-aware Bayesian Choice Models

- Is it reasonable to assume that an individual's preference parameters β_n are independent of the context?
  - Maybe you don't care that going by bike takes longer than by car if the weather is good
  - So the value of β_{time} can vary depending on the weather (context)
- **Research topic:** How to extend previous model[1] to have non-linear dependencies on context variables $\mathbf{c}$ without compromising linear relationship on $\mathbf{x}$?



- Context-specific bias term $\boldsymbol{\mu}_{c_t}$ shifts preference parameters $\boldsymbol{\beta}_n$

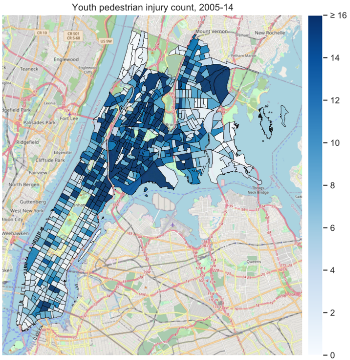
$$\boldsymbol{\beta}_{n,t} = \boldsymbol{\beta}_n + \boldsymbol{\mu}_{c_t}, \qquad \boldsymbol{\mu}_{c_t} \sim \mathcal{N}(\boldsymbol{\mu}_{c_t} \,|\, \text{NNet}_{\boldsymbol{\theta}}(\mathbf{c}_t), \boldsymbol{\Sigma})$$

[1] Reference: `https://arxiv.org/abs/2004.05426`

The Danish National Travel Survey (TU)

- Since 1992 about 350,000 Danish residents have been interviewed
- Detailed information on more than 750,000 trips
- TU is one of the largest surveys of its kind worldwide
- More details about TU:
  `http://www.modelcenter.transport.dtu.dk/english/tvu/what_is_tu`
- Goal: get creative! :-)
  - Model the way people choose transportation modes
  - Model the way people decide to have a car or not
  - Predict next trip based on previous trips
  - Cluster respondents based on their trip profiles
  - Etc.

Bayesian Spatial Count Models



- For each area i, the observed count y_i follows a Negative Binomial distribution

$$y_i \sim \mathsf{NB}(r, p_i), \qquad \text{(likelihood)}$$

$$p_i = \frac{\exp(\psi_i)}{1+\psi_i}, \qquad \text{(link function)}$$

$$\psi = \boldsymbol{\beta}^T \mathbf{X}_i + \phi_i, \qquad \text{(linear function of inputs } \mathbf{X}_i\text{)}$$

where $\mathbf{X}_i$ are the characteristics of the area and ϕ_i is a correlated error term

Bayesian Spatial Count Models

- For each area i, the observed count y_i follows a Negative Binomial distribution

$$
\begin{aligned}
y_i &\sim \mathsf{NB}(r, p_i), && \text{(likelihood)} \\
p_i &= \frac{\exp(\psi_i)}{1+\exp(\psi_i)}, && \text{(link function)} \\
\psi &= \boldsymbol{\beta}^T \mathbf{X}_i + \phi_i, && \text{(linear function of inputs } \mathbf{X}_i\text{)}
\end{aligned}
$$

  where $\mathbf{X}_i$ are the characteristics of the area and ϕ_i is a correlated error term

- **Research topic 1**: How to have non-linear dependencies on inputs $\mathbf{X}_i$?
  - Neural networks?
  - Hierarchical modelling?

[2]Reference: `https://arxiv.org/abs/2007.03681`

Bayesian Spatial Count Models

- For each area i, the observed count y_i follows a Negative Binomial distribution

$$y_i \sim \mathsf{NB}(r, p_i), \qquad \text{(likelihood)}$$
$$p_i = \frac{\exp(\psi_i)}{1+\exp(\psi_i)}, \qquad \text{(link function)}$$
$$\psi = \boldsymbol{\beta}^T \mathbf{X}_i + \phi_i, \qquad \text{(linear function of inputs } \mathbf{X}_i\text{)}$$

  where $\mathbf{X}_i$ are the characteristics of the area and ϕ_i is a correlated error term

- **Research topic 1**: How to have non-linear dependencies on inputs $\mathbf{X}_i$?
  - Neural networks?
  - Hierarchical modelling?
- **Research topic 2**: How to capture spatial correlations between areas?
  - Capture correlations through the noise term?[2]

$$\exp(\tau \mathbf{W})\boldsymbol{\phi} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$$

  - Neural networks? E.g., convolutions or graph convolution neural nets?

[2]Reference: `https://arxiv.org/abs/2007.03681`