Computational social processes

Lirong Xia

Rensselaer

Fall, 2016
Example: Crowdsourcing

Turker 1

Turker 2

Turker n
The Condorcet Jury theorem

The Condorcet Jury theorem.

- Given
  - two alternatives \{O, M\}.
  - \(0.5 < p < 1\),

- Suppose
  - each agent’s preferences is generated i.i.d., such that
    - w/p \(p\), the same as the ground truth
    - w/p \(1-p\), different from the ground truth

- Then, as \(n \to \infty\), the majority of agents’ preferences converges in probability to the ground truth
Today's schedule

• Parametric ranking models
  – Distance-based models
    • Mallows
    • Condorcet
  – Random utility models
    • Plackett-Luce

• Decision making
  – MLE
  – Bayesian
Parametric ranking models

• A statistical model has three parts
  – A parameter space: $\Theta$
  – A sample space: $S = \text{Rankings}(A)^n$
    • $A$ = the set of alternatives, $n$=#voters
    • assuming votes are i.i.d.
  – A set of probability distributions over $S$:
    $\{\Pr_{\theta}(s) \text{ for each } s \in \text{Rankings}(A) \text{ and } \theta \in \Theta\}$
Example

- Condorcet’s model for two alternatives
- Parameter space \( \Theta = \{ \, \, \, \, \} \)
- Sample space \( S = \{ \, \, \, \, \}^n \)
- Probability distributions, i.i.d.

\[
\Pr(\, \, \, \, \, | \, \, \, \, \, ) = \Pr(\, \, \, \, \, | \, \, \, \, \, ) = p > 0.5
\]
Mallows' model [Mallows-1957]

- Fixed dispersion $\varphi < 1$
- Parameter space
  - all full rankings over candidates
- Sample space
  - i.i.d. generated full rankings
- Probabilities:

$$\Pr_W(V) \propto \varphi \text{ Kendall}(V,W)$$
Example: Mallows for

- Probabilities: $Z = 1 + 2\varphi + 2\varphi^2 + \varphi^3$
Condorcet’s model
[Condorcet-1785, Young-1988, ES UAI-14, APX NIPS-14]

• Fixed dispersion $\varphi < 1$

• Parameter space
  – all binary relations over candidates

• Sample space
  – i.i.d. generated binary relations

• Probabilities:

  $\Pr_W(V) \propto \varphi \text{ Kendall}(V, W)$
Continuous parameters: $\Theta=(\theta_1,\ldots, \theta_m)$
- $m$: number of alternatives
- Each alternative is modeled by a utility distribution $\mu_i$
- $\theta_i$: a vector that parameterizes $\mu_i$

An agent’s latent utility $U_i$ for alternative $c_i$ is generated independently according to $\mu_i(U_i)$

Agents rank alternatives according to their perceived utilities
- $\Pr(c_2>c_1>c_3|\theta_1, \theta_2, \theta_3) = \Pr_{U_i \sim \mu_i}(U_2>U_1>U_3)$
Generating a preference-profile

\[ \text{Pr}(\text{Data} \mid \theta_1, \theta_2, \theta_3) = \prod_{V \in \text{Data}} \text{Pr}(V \mid \theta_1, \theta_2, \theta_3) \]

Parameters

\[ \theta_3 \quad \theta_2 \quad \theta_1 \]

Agent 1

\[ P_1 = c_2 \succ c_1 \succ c_3 \]

...  

Agent n

\[ P_n = c_1 \succ c_2 \succ c_3 \]
\textbf{Plackett-Luce model}

- $\mu_i$’s are Gumbel distributions
  - A.k.a. the Plackett-Luce (P-L) model [BM 60, Yellott 77]

- Alternative parameterization $\lambda_1, \ldots, \lambda_m$

\[
\Pr(c_1 \succ c_2 \succ \cdots \succ c_m \mid \lambda_1 \cdots \lambda_m) = \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_m} \times \frac{\lambda_2}{\lambda_2 + \cdots + \lambda_m} \times \cdots \times \frac{\lambda_{m-1}}{\lambda_{m-1} + \lambda_m}
\]

\begin{itemize}
  \item $c_2$ is the top preference to $\{c_1, \ldots, c_m\}$
\end{itemize}

\begin{itemize}
  \item Pros:
    \begin{itemize}
      \item Computationally tractable
        \begin{itemize}
          \item Analytical solution to the likelihood function
            \begin{itemize}
              \item The only RUM that was known to be tractable
            \end{itemize}
          \end{itemize}
        \end{itemize}
    \end{itemize}
  \end{itemize}

\begin{itemize}
  \item Cons: may not be the best model
\end{itemize}
Example

\[
\frac{1}{10} \times \frac{4}{9} \\
\]

\[
\frac{4}{10} \times \frac{1}{6} \\
\]

\[
\frac{1}{10} \times \frac{5}{9} \\
\]

\[
\frac{5}{10} \times \frac{1}{5} \\
\]

Truth

\[
\frac{4}{10} \times \frac{5}{6} \\
\]

\[
\frac{5}{10} \times \frac{4}{5} \\
\]
RUM with normal distributions

- $\mu_i$’s are normal distributions
  - Thurstone’s Case V [Thurstone 27]

**Pros:**
- Intuitive
- Flexible

**Cons:** believed to be computationally intractable
- No analytical solution for the likelihood function $\Pr(P \mid \Theta)$ is known

$$
\Pr(c_1 > \cdots > c_m \mid \Theta) = \int_{-\infty}^{\infty} \int_{U_m}^{\infty} \cdots \int_{U_2}^{\infty} \mu_m(U_m) \mu_{m-1}(U_{m-1}) \cdots \mu_1(U_1) dU_1 \cdots dU_{m-1} dU_m
$$

- $U_m$: from $-\infty$ to $\infty$
- $U_{m-1}$: from $U_m$ to $\infty$
- $\cdots$
- $U_1$: from $U_2$ to $\infty$
Model selection

• Compare RUMs with Normal distributions and PL for
  – log-likelihood: $\log \Pr(D|\Theta)$
  – predictive log-likelihood: $\mathbb{E} \log \Pr(D_{\text{test}}|\Theta)$
  – Akaike information criterion (AIC): $2k - 2\log \Pr(D|\Theta)$
  – Bayesian information criterion (BIC): $k \log n - 2\log \Pr(D|\Theta)$

• Tested on an election dataset
  – 9 alternatives, randomly chosen 50 voters

<table>
<thead>
<tr>
<th>Value(Normal) - Value(PL)</th>
<th>LL</th>
<th>Pred. LL</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44.8(15.8)</td>
<td>87.4(30.5)</td>
<td>-79.6(31.6)</td>
<td>-50.5(31.6)</td>
</tr>
</tbody>
</table>

Red: statistically significant with 95% confidence

Project: model fitness for election data
Decision making
Maximum likelihood estimators (MLE)

Model: $\mathcal{M}_r$

“Ground truth” $\theta$

$V_1$ $V_2$ ... $V_n$

• For any profile $P=(V_1,\ldots,V_n)$,
  
  – The likelihood of $\theta$ is $L(\theta,P)=\Pr_{\theta}(P)=\prod_{V\in P}\Pr_{\theta}(V)$
  
  – The MLE mechanism
  
  $\text{MLE}(P)=\text{argmax}_\theta L(\theta,P)$
  
  – Decision space = Parameter space
Bayesian approach

- Given a profile $P=(V_1,\ldots,V_n)$, and a prior distribution $\pi$ over $\Theta$
- Step 1: calculate the posterior probability over $\Theta$ using Bayes’ rule
  - $\Pr(\theta|P) \propto \pi(\theta) \Pr(\theta|P)$
- Step 2: make a decision based on the posterior distribution
  - Maximum a posteriori (MAP) estimation
  - $\text{MAP}(P)=\text{argmax}_\theta \Pr(\theta|P)$
  - Technically equivalent to MLE when $\pi$ is uniform
Example

- $\Theta = \{\text{ }, \text{ }\}$
- $S = \{\text{ }, \text{ }\}^n$
- Probability distributions:
- Data $P = \{10@ + 8@\}$
- MLE
  - $L(O) = Pr_O(O)^6 \cdot Pr_O(M)^4 = 0.6^{10} \cdot 0.4^8$
  - $L(M) = Pr_M(O)^6 \cdot Pr_M(M)^4 = 0.4^{10} \cdot 0.6^8$
  - $L(O) > L(M)$, $O$ wins
- MAP: prior $O:0.2$, $M:0.8$
  - $Pr(O|P) \propto 0.2 \cdot L(O) = 0.2 \times 0.6^{10} \cdot 0.4^8$
  - $Pr(M|P) \propto 0.8 \cdot L(M) = 0.8 \times 0.4^{10} \cdot 0.6^8$
  - $Pr(M|P) > Pr(O|P)$, $M$ wins

$Pr(\text{ } | \text{ }) = Pr(\text{ } | \text{ }) = 0.6$
Decision making under uncertainty

• You have a biased coin: head w/p $p$
  – You observe 10 heads, 4 tails
  – Do you think the next two tosses will be two heads in a row?

• MLE-based approach
  – there is an unknown but fixed ground truth
  – $p = 10/14 = 0.714$
  – $\Pr(2\text{heads}|p=0.714) = (0.714)^2 = 0.51 > 0.5$
  – Yes!

• Bayesian
  – the ground truth is captured by a belief distribution
  – Compute $\Pr(p|\text{Data})$ assuming uniform prior
  – Compute $\Pr(2\text{heads}|\text{Data}) = 0.485 < 0.5$
  – No!

Credit: Panos Ipeirotis & Roy Radner
Statistical decision theory

- Given
  - statistical model: $\Theta$, $S$, $\Pr_\theta (s)$
  - decision space: $D$
  - loss function: $L(\theta, d) \in \mathbb{R}$

- Make a good decision based on data
  - decision function $f$: data $\rightarrow D$
  - Bayesian expected lost:
    - $\text{EL}_B(\text{data}, d) = \mathbb{E}_{\theta | \text{data}} L(\theta, d)$
  - Frequentist expected lost:
    - $\text{EL}_F(\theta, f) = \mathbb{E}_{\text{data} | \theta} L(\theta, f(\text{data}))$
  - Evaluated w.r.t. the objective ground truth
Top 250 movies

“Complex voter weighting system”

- Claimed to be accurate
  - a “true Bayesian estimate”
- Claimed to be fair

1. The Shawshank Redemption (1994) ★ 9.2
2. The Godfather (1972) ★ 9.2
4. The Dark Knight (2008) ★ 8.9
Different Voice

- Q: “This is unfair!”
  - “That film / show has received awards, great reviews, commendations and deserves a much higher vote!”
- IMDB: “…only votes cast by IMDb users are counted. We do not delete or alter individual votes”

IMDb Votes/Ratings Top Frequently Asked Questions
http://www.imdb.com/help/show_leaf?votestopfaq
Fairness of Bayesian estimators

- **Theorem**: Strict Condorcet
  
  No Bayesian estimator satisfies strict Condorcet criterion

- **Theorem**: Neutrality
  
  Neutral Bayesian estimators
  
  = Bayesian estimators of “neutral” models