Computing Choice

Devavrat Shah

Laboratory for Information and Decision Systems
Department of EECS
Massachusetts Institute of Technology
http://web.mit.edu/devavrat/www
(list of relevant references in the last set of slides)
Purpose

- Ideally
  - Graphical models
    - Belief propagation
  - Connections to Probability, Statistics, EE, CS, ...

- In reality
  - A set of very exciting (to me, may be others) questions at the interface of all of the above and more
  - Seemingly unrelated to graphical model
  - However, provide fertile ground to understand everything about graphical models (algorithms, analysis)
Scenarios

- Recommendations
  - What movie to watch
  - Which restaurant to eat
  - ...

- Precisely,
  - Suggest what you may like
  - Given what others have liked
    - By finding others like you and what they had liked
Scenarios

- Ranking
  - Players and/or Teams
    - Based on outcome of games
  - Papers at a competitive conference
    - Using reviews
  - Graduate admissions
    - From feedback of professors

- Precisely,
  - Global ranking of objects from partial preferences
Revealed preferences

- Partial preferences are revealed in different forms
  - Sports: Win and Loss
  - Social: Starred rating
  - Conferences: Scores

- All can be viewed as *pair-wise comparisons*
  - IND beats AUS: IND > AUS
  - Clio ***** vs No 9 Park ****: Clio > No 9 Park
  - Ranking Paper 9/10 vs Other Paper 5/10: Ranking > Other
Data and Decision

- Revealed preferences lead to
  - Bag of pair-wise comparisons

- Question of interest
  - Recommendations
    - Suggest what you may like given what others have liked
  - Ranking
    - Global ranking of objects given outcome of games/…

- This requires understanding (computing) choice model
  - What people like/dislike from pair-wise comparisons
Modeling people’s choice

- **Rational view:** Axiom of revealed preferences [Samuelson ‘37]
  - There is one ordering over all objects consistent across population
    - Unlikely (lack of transitivity in people’s preferences)

- **Meaningful view** – “discrete choice model”
  - Distribution over orderings of objects
    - Consistent with population’s revealed preferences

Data → Choice Model → Decision
Scenarios

- Object tracking (cf. Huang, Guestrin, Guibas ‘08)
  - Noisy observations of locations
  - Feasible to maintain partial information only
    - $Q=[Q_{ij}]$ – first-order information

![Diagram showing objects and locations with Q11 = P1→P1, Q12, Q13 connections between objects and locations]
Object tracking

- Noisy observations of locations
- Feasible to maintain partial information only
  - $Q = [Q_{ij}]$ – first-order information

**Scenarios**

Data $\rightarrow$ Choice Model $\rightarrow$ Decision
More generally

- Recommendation
- Ranking
- Object tracking
- Policy making
- Business operations (assortment optimization)
- Display advertising
- Polling,

- Canonical question
  - Decision using choice model learnt from partial preference data
Part I: Object tracking
Q. Given weighted bipartite graph $G = (V, E, Q)$
   - Find matching of objects/positions
   - That is ‘most likely’
Mode = Maximum Weight Matching

- Answer: maximum weight matching
  - Weight of a matching equals
    - summation of Q-entries of edges participating in the matching
Part II: Ranking & Recommendation
Q1. Given weighted comparison graph \( G=(V, E, A) \)
- Find ranking of/scores associated with objects

Q2. When possible (e.g. Conference/Crowd-Sourcing), choose \( G \) so as to
- Minimize the number of comparisons required to find ranking/scores
Random walk on comparison graph $G=(V,E,A)$
- $d = \max$ (undirected) vertex degree of $G$
- For each edge $(i,j)$:
  - $P_{ij} = \frac{A_{ji} + 1}{A_{ij} + A_{ji} + 2} \times \frac{1}{d}$
- For each node $i$:
  - $P_{ii} = 1 - \sum_{j \neq i} P_{ij}$

Let $G$ be connected
- Let $s$ be the unique stationary distribution of RW $P$
  - $s^T = s^T P$

Ranking:
- Use $s$ as scores of objects
Rank centrality + Graph choice

- Random walk on comparison graph \( G=(V,E,A) \)
  - \( d = \text{max (undirected) vertex degree of } G \)
  - For each edge \((i,j):\)
    - \( P_{ij} = (A_{ji} + 1)/(A_{ij} + A_{ij} + 2) \times 1/d \)
  - For each node \(i:\)
    - \( P_{ii} = 1 - \sum_{j \neq i} P_{ij} \)

- Ranking: use \( s \) as scores of objects, where
  - \( s \) be the unique stationary distribution of RW \( P \)
    \[ s^T = s^T P \]

- Choice of graph \( G \)
  - Subject to constraints, choose \( G \) so that
  - Spectral gap of natural RW on \( G \) is maximized
  - SDP [Boyd, Diaconis, Xiao ’04]
Rest of the Tutorial

- Maximum Weight Matching
  - How to compute it?
    - Belief propagation
  - Why does it make sense?
    - Max-likelihood estimation w.r.t. “exponential family”

- Rank centrality
  - How to compute it?
    - Power-iteration
  - Why does it make sense?
    - Mode for Bradley-Terry-Luce (or MNL) model
Justification: Maximum Weight Matching
Maximum weight matching
(all of below explained using class-board)

- Computation
  - Belief propagation
    - Algorithm
    - Why it works

- Model
  - Maximum entropy (max-ent) consistent distribution
    - Maximum Likelihood in exponential family
    - Maximum weight matching
      - “First-order” approximation of mode of this distribution
  - Exact computation of max-ent
    - Via dual gradient
    - Belief propagation/MCMC at rescue
Justification: Rank Centrality

A > B
B > C
C > A
C > A
Model for choice

- Choice model (distribution over permutations)
  [Bradley-Terry-Luce (BTL) or MNL (cf. McFadden) Model]
  - Each object $i$ has an associated weight $w_i > 0$
  - When objects $i$ and $j$ are compared
    - $P(i > j) = w_i / (w_i + w_j)$

- Sampling model
  - Edges $E$ of graph $G$ are selected
  - For each $(i,j) \in E$, sample $k$ pair-wise comparisons
Rank centrality: simulation

- Error(s) = \[ \frac{1}{\|w\|} \left( \sum_{i>j} (w_i-w_j)^2 I\{ (s(i)-s(j))(w_i-w_j)<0 \} \right)^{1/2} \]
- G: Erdos-Renyi graph with edge prob. d/n
Theorem 1.

Let $R = \max_{i,j} \frac{w_i}{w_j}$.

Let $G$ be Erdos-Renyi graph.

Under Rank centrality, with $d = \Omega(\log n)$

$$\frac{\|s-w\|}{\|w\|} \leq C \sqrt{\frac{R^5 \log n}{kd}}$$

That is, sufficient to have $O(R^5 n \log n)$ samples

- Optimal dependence on $n$ for ER graph
- Dependence on $R$?
Theorem 1.

- Let $R = \max_{i,j} \frac{w_i}{w_j}$.
- Let $G$ be Erdos-Renyi graph.
  - Under Rank centrality, with $d = \Omega(\log n)$
    
    $$\frac{\|s-w\|}{\|w\|} \leq C \sqrt{\frac{R^5 \log n}{kd}}$$

- Information theoretic lower-bound: for any algorithm

  $$\frac{\|s-w\|}{\|w\|} \geq C' \sqrt{\frac{1}{kd}}$$
Theorem 2.

Let \( R = (\max_{ij} w_i/w_j) \).

Let \( G \) be any connected graph:
- \( L = D^{-1} E \) be natural RW transition matrix
- \( \Delta = 1 - \lambda_{\text{max}}(L) \)
- \( \kappa = d_{\text{max}}/d_{\text{min}} \)

Under Rank centrality, with \( kd = \Omega(\log n) \)

\[
\frac{\|s-w\|}{\|w\|} \leq \frac{C}{\Delta} \kappa \sqrt{\frac{R^5 \log n}{kd}}
\]

That is, number of samples required \( O(R^5 \kappa^2 n \log n \times \Delta^{-2}) \)
- Graph structure plays role through it’s Laplacian.
Rank centrality and Graph choice

- **Theorem 2.**
  - Under Rank centrality, with $kd = \Omega(\log n)$
    
    $$\frac{\|s-w\|}{\|w\|} \leq \frac{C}{\Delta} \kappa \sqrt{\frac{R^5 \log n}{kd}}$$
  - That is, number of samples required $O(R^5 \kappa^2 n \log n \times \Delta^{-2})$

- Choice of graph $G$
  - Subject to constraints, choose $G$ so that
  - Spectral gap $\Delta$ is maximized
    
    SDP [Boyd, Diaconis, Xiao ‘04]
Some remarks on proof

Spectral analysis of reversible Markov chains

Markov chain \((P, \pi)\) is reversible iff \(\pi(i)P_{ij} = \pi(j)P_{ji}\)

\[
P_{ij} = \frac{1}{d_{\text{max}}} \frac{A_{ji}}{A_{ij} + A_{ji}}
\]

is not reversible, but the expectation \(\tilde{\pi}(i)\tilde{P}_{ij} = \tilde{\pi}(j)\tilde{P}_{ji}\)

\[
\tilde{P}_{ij} = \frac{1}{d_{\text{max}}} \frac{w_j}{w_i + w_j}
\]

\(\tilde{\pi}(i) \propto w_i\)
Some remarks on proof

Spectral analysis of reversible Markov chains
For any Markov chain \((P, \pi)\) and any reversible MC \((\tilde{P}, \tilde{\pi})\)

\[
\|\pi - \tilde{\pi}\| = \|P^T \pi - \tilde{P}^T \tilde{\pi}\|
\leq \|P^T \pi - P^T \tilde{\pi}\| + \|P^T \tilde{\pi} - \tilde{P}^T \tilde{\pi}\|
\leq \|\tilde{P}^T (\pi - \tilde{\pi})\| + \|\tilde{P}^T (\pi - \tilde{\pi})\| + \|P - \tilde{P}\| \|\tilde{\pi}\|
\leq (\lambda_2(\tilde{P}) + \|P - \tilde{P}\|) \|\pi - \tilde{\pi}\| + \|P - \tilde{P}\| \|\tilde{\pi}\|
\]

\[
\frac{\|\pi - \tilde{\pi}\|}{\|\tilde{\pi}\|} \leq \frac{\|P - \tilde{P}\|}{1 - \lambda_2(\tilde{P}) - \|P - \tilde{P}\|}
\]
Some remarks on proof

\[ \frac{\|\pi - \tilde{\pi}\|}{\|\tilde{\pi}\|} \leq \left(1 - \lambda_2(\tilde{P})\right) - \|P - \tilde{P}\| \]

- Bound on \(1 - \lambda_2(\tilde{P})\)
  - Use of comparison theorem [Diaconis-Saloff Coste ‘94]++

- Bound on \(\|P - \tilde{P}\|\)
  - Use of (modified) concentration of measure inequality for matrices

- Finally, use this to further bound Error(s)
Rank centrality: experiment
Washington Post: Allourideas

Who had the worst year in Washington?

The Working Poor
John Pistole

I can't decide
Add your own idea

Click on an idea to start voting.

This idea marketplace is powered by All Our Ideas
Rank centrality: experiment

- **Ground truth**: algorithm’s result with complete data
- **Error**: average position discrepancy

![Graph showing L1 ranking and Rank Centrality](image)
Rank aggregation: Background

- Input: complete preference (not comparisons)
  - Axiomatic impossibility [Arrow ’51]

- Some algorithms
  - Kemeny optimal: minimize disagreements
    - Extended Condorcet Criteria
    - NP-hard, 2-approx algorithm [Dwork et al ’01]

- Borda count: average position is score
  - Simple
  - Useful axiomatic properties [Young ’74]
Rank aggregation: Background

- Algorithms with partial data
  - Let pair-wise data available for all pairs
  - Kemeny distance depends on pair-wise marginal only

\[
\arg\min_{\sigma} \sum_{i,j} A_{ij} I(\sigma(i) < \sigma(j))
\]

- Data is consistent with a distribution on permutations
  - For example, obtained as the max-ent approximation
  - Kemeny optimal of this distribution is the same as above
    - NP-hard
  - 2-approx for this distribution acts as 2-approx for the above

[Ammar-Shah ’11 ’12]
Rank aggregation: Background

- Borda count
  - Average position
  - But, comparison do not have position information
  - Given pair-wise marginal $p_{ij}$ for all $i \neq j$

- For any distribution consistent with pair-wise marginal
  - Borda count is given as

$$c(i) \propto \sum_j p_{ij}$$

[Ammar-Shah ‘11 ’12]
Related work

- Finding winner and BTL choice model
  [Adler, Gemmell, Harchol-Balter, Karp, Kenyon ‘87]
  - $O(\log n)$ iteration adaptive algorithm, $O(n)$ total comparisons

- Noisy sorting and Mallow’s model [Braverman, Mossel ‘09]
  - $O(n \log n)$ samples in total (complete ordering)
  - Average position (Borda count) algorithm++
    - Polynomial($n$) time algorithm
Concluding remarks + Open question

- Choice model
  - A powerful model to tackle a range of questions
  - Many are in it’s infancy (e.g. recommendations)
  - Challenge being computation + statistics
  - Excellent “play ground” to resolve challenges of graphical models

- Two examples in this tutorial
  - Object tracking
    - Learning from first-order marginals
  - Ranking
    - Using pair-wise comparisons
Concluding remarks + Open question

- Open direction:
  - Learning graphical models efficiently
    - Computationally and statistically

- A concrete question:
  - Given pair-wise comparisons data
    - When can we learn the choice model efficiently?
  - For example, if exact pair-wise comparison marginals available
    - Then, can learn ‘sparse’ choice model up to $o(\log n)$ sparsity [Farias + Jagabathula + S ‘09 ‘12]
    - But what about noisy setting?
  - Or, max-ent learning?
References

- Part I:

Color coding

- covered in tutorial
- sparse choice model that is v. relevant
- others/ closely related/definitely worth a read

- definitely worth a read
References

- Part II:
References

- At large:
  - D. McFadden, "Disaggregate Behavioral Travel Demand’s RUM Side," A 30-year Retrospective, available online [http://emlab.berkeley.edu/pub/wp/mcfaddeno300.pdf](http://emlab.berkeley.edu/pub/wp/mcfaddeno300.pdf)

*Color coding*

- covered in tutorial
- sparse choice model that is v. relevant
- others/
close related/definitely worth a read