Ranking from Pairwise Comparisons

Sewoong Oh

Department of ISE
University of Illinois at Urbana-Champaign

Joint work with Sahand Negahban(MIT) and Devavrat Shah(MIT)
Rank Aggregation

Data

Algorithm

Decision

Score/Ranking

1 1.0
2 0.9
3 0.6
4 0.5
5 0.4
6 0.2
7 0.1

1
2
3
4
5
6
7
Example

Data  Algorithm  Decision

[kittenwar.com]

Click the cutest kitten picture!
Example

Data → Algorithm → Decision

.78
.77
.76
Example

Data → Algorithm → Decision

Data: Image of kittens with numerical values (0.78, 0.77, 0.76, 0.20, 0.21, 0.21)
Algorithm: Images of different cats
Decision: Images of different cats with numerical values (0.21, 0.21, 0.20)
Example

Data → Algorithm → Decision

What is our most important national priority?

- Legalize medical marijuana
- Eliminate poverty
- I can't decide

24392 votes on 348 ideas

Help your community
Add your own idea
### Example

![Diagram showing the process of Data -> Algorithm -> Decision]

#### What is our most important national priority?

<table>
<thead>
<tr>
<th>Priority</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving public education</td>
<td>82</td>
</tr>
<tr>
<td>Improve the quality of public education</td>
<td>81</td>
</tr>
<tr>
<td>Teach people how to think for themselves</td>
<td>80</td>
</tr>
<tr>
<td>Restore and grow the middle class, diminishing disparities in wealth transfer.</td>
<td>79</td>
</tr>
<tr>
<td>Health care</td>
<td>78</td>
</tr>
<tr>
<td>Education</td>
<td>78</td>
</tr>
<tr>
<td>Climate change</td>
<td>78</td>
</tr>
<tr>
<td>Education</td>
<td>78</td>
</tr>
<tr>
<td>Decreasing military funding and increasing funding for education</td>
<td>78</td>
</tr>
<tr>
<td>Education</td>
<td>77</td>
</tr>
</tbody>
</table>

See more...
Comparisons data

- Group decisions, recommendation and advertisement, sports and game
- Universal (ratings can be converted into comparisons)
- Consistent and reliable
- Natural (e.g. Sports and games, MSR’s TrueSkill)
- Aggregation is challenging
NP-hard approach: Kemeny optimal algorithm

Find the most consistent ranking

\[ \arg \min_{\sigma} \sum_{i \neq j} a_{ij} \mathbb{I}(\sigma(j) > \sigma(i)) \]

- NP-hard combinatorial optimization
- Optimal for a specific model

Strong

Weak

w.p. 1 - \( p \)

w.p. \( p \)
Traditional approach: $\ell_1$ Ranking

- Compute score:
  \[ s(i) = \frac{1}{|\partial i|} \sum_{j \in \partial i} \frac{a_{ji}}{a_{ij} + a_{ji}} \]

- $n$: number of items
- $m$: number of samples
- Compute in $O(m)$ time
- Works when everyone plays everyone else: $m = \Omega(n^2)$ [Ammar, Shah '12]
Traditional approach: $\ell_1$ Ranking

- Compute score:
  \[ s(i) = \frac{1}{|\partial i|} \sum_{j \in \partial i} \frac{a_{ji}}{a_{ij} + a_{ji}} \]

- $n$: number of items
- $m$: number of samples
- Compute in $O(m)$ time
- Works when everyone plays everyone else: $m = \Omega(n^2)$ [Ammar, Shah ’12]

Strong Weak

All wins are equally weighted
Traditional approach: $\ell_1$ Ranking

Compute score:

$$s(i) = \frac{1}{|\partial i|} \sum_{j \in \partial i} \frac{a_{ji}}{a_{ij} + a_{ji}}$$

- $n$: number of items
- $m$: number of samples
- Compute in $O(m)$ time
- Works when everyone plays everyone else: $m = \Omega(n^2)$ [Ammar, Shah ’12]

All wins are equally weighted
Traditional approach: $\ell_1$ Ranking

- Compute score:
  \[ s(i) = \frac{1}{|\partial i|} \sum_{j \in \partial i} \frac{a_{ji}}{a_{ij} + a_{ji}} s(j) \]

- $n$: number of items
- $m$: number of samples
- Compute in $O(m)$ time
- Works when everyone plays everyone else:
  \[ m = \Omega(n^2) \] [Ammar, Shah ’12]

All wins are equally weighted
Prior work

Data → Algorithm → Decision

- $\ell_1$ ranking
- $\ell_p$ ranking
- Kemeny optimal
- MNL
- Mixed MNL
- etc.

Score/Ranking:

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0.9</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>0.1</td>
<td>7</td>
</tr>
</tbody>
</table>
Challenges

- High-dimensional regime

\[ \# \text{ of samples} \propto n(\log n) \]

- Low computational complexity

\[ \# \text{ of operations} \propto \# \text{ of samples} \]

- Model independent

- Makes sense
Challenges

- High-dimensional regime
  \[
  \# \text{ of samples} \propto n(\log n)
  \]

- Low computational complexity
  \[
  \# \text{ of operations} \propto \# \text{ of samples}
  \]

- Model independent

- Makes sense

Under BTL model, Rank Centrality achieves optimal performance
Rank Centrality

- Define a random walk on $G$

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

\[
P_{ii} = 1 - \frac{1}{d_{\text{max}}} \sum_{j \neq i} \frac{a_{ij}}{a_{ij} + a_{ji}}
\]

- (unique) stationary distribution

\[
s^T P = s^T
\]

- Random walk spends more time on ‘stronger’ nodes
Rank Centrality

- Define a random walk on $G$

$$P_{ij} = \frac{1}{d_{\max}} \frac{a_{ij}}{a_{ij} + a_{ji}}$$

$$P_{ii} = 1 - \frac{1}{d_{\max}} \sum_{j \neq i} \frac{a_{ij}}{a_{ij} + a_{ji}}$$

- (unique) stationary distribution

$$s^T P = s^T$$

- Random walk spends more time on ‘stronger’ nodes
Define a random walk on $G$

$$P_{ij} = \frac{1}{d_{\text{max}}} \frac{a_{ij}}{a_{ij} + a_{ji}}$$

$$P_{ii} = 1 - \frac{1}{d_{\text{max}}} \sum_{j \neq i} \frac{a_{ij}}{a_{ij} + a_{ji}}$$

(Unique) stationary distribution

$$s^T P = s^T$$

Random walk spends more time on ‘stronger’ nodes
Rank Centrality

- Define a random walk on $G$

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

\[ P_{ii} = 1 - \frac{1}{d_{\text{max}}} \sum_{j \neq i} \frac{a_{ij}}{a_{ij} + a_{ji}} \]

- (unique) stationary distribution

\[ s^T P = s^T \]

- Random walk spends more time on ‘stronger’ nodes
Rank Centrality

- Define a random walk on $G$

$$P_{ij} = \frac{1}{d_{\max}} \frac{a_{ij}}{a_{ij} + a_{ji}}$$

$$P_{ii} = 1 - \frac{1}{d_{\max}} \sum_{j \neq i} \frac{a_{ij}}{a_{ij} + a_{ji}}$$

- (unique) stationary distribution

$$s^T P = s^T$$

- Random walk spends more time on ‘stronger’ nodes
**Rank Centrality**

- Define a **random walk** on $G$
  
  $$P_{ij} = \frac{1}{d_{\text{max}}} \frac{a_{ij}}{a_{ij} + a_{ji}}$$
  
  $$P_{ii} = 1 - \frac{1}{d_{\text{max}}} \sum_{j \neq i} \frac{a_{ij}}{a_{ij} + a_{ji}}$$

- (**unique**) stationary distribution
  
  $$s^T P = s^T$$

- Random walk spends more time on ‘stronger’ nodes
- Higher score for beating a ‘stronger’ node

$$s(i) = \left(1 - \frac{1}{d_{\text{max}}} \sum_{j \neq i} \frac{a_{ij}}{a_{ij} + a_{ji}} \right) s(i) + \sum_{j \neq i} P_{ji} s(j)$$

$$= \frac{1}{Z_i} \sum_{j \neq i} \frac{a_{ji}}{a_{ij} + a_{ji}} s(j)$$
Experiment: Polling public opinions

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.
Experiment: Polling

- **Ground truth**: what the algorithm produces with complete data
- **Error** = $\frac{1}{n} \sum_{i}(\sigma_i - \hat{\sigma}_i)$
Model

- Algorithm is model independent
- For experiments and analysis, need model to generate data
- Comparisons model
  - Bradley-Terry-Luce (BTL) model
    - there is a true ranking \( \{w_i\} \)
    - when a pair is compared, the noise is modelled by
      \[
P(i < j) = \frac{w_j}{w_i + w_j}
\]
Model

- Algorithm is model independent
- For experiments and analysis, need model to generate data
- Comparisons model
  - Bradley-Terry-Luce (BTL) model
    - there is a true ranking \(\{w_i\}\)
    - when a pair is compared, the noise is modelled by
      \[
P(i < j) = \frac{w_j}{w_i + w_j}\]
- Sampling model
  - Sample each pair with probability \(d/n\)
  - \(k\) comparisons for each pair
Experiment: BTL model

\[ \text{Error} = \frac{1}{\|w\|} \sum_{i>j} (w_i - w_j)^2 \mathbb{1}(\hat{\sigma}_i - \hat{\sigma}_j)(w_i - w_j) > 0 \]

![Graph showing the relationship between k and Error](image1)

![Graph showing the relationship between d/n and Error](image2)
Performance guarantee

- For $d = \Omega(\log n)$
- BTL model $\{w_i\}$ with $w_{\min} = \Theta(w_{\max})$

Theorem (Negahban, O., Shah, ’12)

- **Rank Centrality** achieves

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

- Information-theoretic lower bound:

$$\inf_s \sup_{w \in \mathcal{W}} \frac{\|w - s\|}{\|w\|} \geq C' \sqrt{\frac{1}{kd}}$$
Performance guarantee

- For \( d = \Omega(\log n) \)
- BTL model \( \{w_i\} \) with \( w_{\min} = \Theta(w_{\max}) \)

**Theorem (Negahban, O., Shah, '12)**

- **Rank Centrality** achieves
  \[
  \frac{\|w - s\|}{\|w\|} \leq C \sqrt{\frac{\log n}{kd}}
  \]

- Information-theoretic lower bound:
  \[
  \inf_s \sup_{w \in \mathcal{W}} \frac{\|w - s\|}{\|w\|} \geq C'' \sqrt{\frac{1}{kd}}
  \]

\# of samples = \( O(n \log n) \) suffices to achieve arbitrary small error
Performance guarantee for general graphs

- Oftentimes we do not control how data is collected
- Let $G$ denote the (undirected) graph of given data
- Compute spectral gap of $G$ (cf. mixing time of natural random walk)

$$
\xi \equiv 1 - \frac{\lambda_1(G)}{\lambda_2(G)}
$$

**Theorem (Negahban, O., Shah, '12)**

**Rank Centrality** achieves

$$\frac{\|w - s\|}{\|w\|} \leq C \frac{d_{\text{max}}}{\xi d_{\text{min}}} \sqrt{\frac{\log n}{k d_{\text{max}}}}$$
Performance guarantee for general graphs

- Oftentimes we do not control how data is collected
- Let $G$ denote the (undirected) graph of given data
- Compute spectral gap of $G$ (cf. mixing time of natural random walk)

$$\xi \equiv 1 - \frac{\lambda_1(G)}{\lambda_2(G)}$$

Theorem (Negahban, O., Shah, ’12)

**Rank Centrality** achieves

$$\frac{\|w - s\|}{\|w\|} \leq C \frac{d_{\text{max}}}{\xi d_{\text{min}}} \sqrt{\frac{\log n}{k d_{\text{max}}}}$$

# of samples = $O(n \log n)$ suffices to achieve arbitrary small error
1. Spectral analysis of reversible Markov chains
   Markov chain \((P, \pi)\) is reversible iff \(\pi(i)P_{ij} = \pi(j)P_{ji}\)

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

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

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

\(\tilde{\pi}(i) \propto w_i\)
Proof technique

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

\[
\|\pi - \tilde{\pi}\| = \|P^T \pi - \tilde{P}^T \tilde{\pi}\| \\
\leq \|P^T \pi - P^T \tilde{\pi}\| + \|P^T \tilde{\pi} - \tilde{P}^T \tilde{\pi}\| \\
\leq \|\tilde{P}^T (\pi - \tilde{\pi})\| + \|(P - \tilde{P})^T (\pi - \tilde{\pi})\| + \|P - \tilde{P}\|_2 \|\tilde{\pi}\| \\
\leq (\lambda_2(\tilde{P}) + \|P - \tilde{P}\|_2) \|\pi - \tilde{\pi}\| + \|P - \tilde{P}\|_2 \|\tilde{\pi}\|
\]

\[
\frac{\|\pi - \tilde{\pi}\|}{\|\tilde{\pi}\|} \leq \frac{\|P - \tilde{P}\|_2}{1 - \lambda_2(\tilde{P}) - \|P - \tilde{P}\|_2}
\]
Proof technique

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

\[
\|\pi - \tilde{\pi}\| = \|P^T \pi - \tilde{P}^T \tilde{\pi}\|
\leq \|P^T \pi - P^T \tilde{\pi}\| + \|P^T \tilde{\pi} - \tilde{P}^T \tilde{\pi}\|
\leq \|\tilde{P}^T (\pi - \tilde{\pi})\| + \|(P - \tilde{P})^T (\pi - \tilde{\pi})\| + \|P - \tilde{P}\|_2 \|\tilde{\pi}\|
\leq (\lambda_2(\tilde{P}) + \|P - \tilde{P}\|_2) \|\pi - \tilde{\pi}\| + \|P - \tilde{P}\|_2 \|\tilde{\pi}\|
\]

\[
\frac{\|\pi - \tilde{\pi}\|}{\|\tilde{\pi}\|} \leq \frac{\|P - \tilde{P}\|_2}{1 - \lambda_2(\tilde{P}) - \|P - \tilde{P}\|_2}
\]

2. Bound \(1 - \lambda_2(\tilde{P})\) by comparisons theorem
3. Bound \(\|P - \tilde{P}\|_2\) by concentration of measure inequality for matrices
Conclusion

- Rank aggregation from comparisons
- High-dimensional regime \( nk \sim n \log n \)
- Low complexity \( O(nkd \log n) \)
- Model independent
- Optimal under BTL up to \( \log n \)
- Other Markov chain approaches e.g. "Rank Aggregation Revisited"
  C. Dwork, R. Kumar, M. Naor, and D. Sivakumar