Whom to Ask?

Jury Selection for Decision Making Tasks on Micro-blog Services

Caleb Chen CAO, Jieying SHE, Yongxin TONG, Lei CHEN
The Hong Kong University of Science and Technology
“Is Istanbul the capital of Turkey?”
Social Network/Media services

the *virtualization* and *digitalization* of
people’s social activities
• **Minor** as dressing for a banquet
• **Major** as prediction of macro economy trends

“two-option decision making tasks”
Wisdom of Crowd

“The basic argument there, drawing on a long history of intuition about markets, is that the aggregate behavior of many people, each with limited information, can produce very accurate beliefs.” —D. Easley, J. Kleinberg, “Networks, Crowds, and Markets”
Crowdsourcing-powered DB Systems

• Qurk, “Human powered Sorts and Joins”, VLDB’2012(MIT)
• Deco, “A System for Declarative Crowdsourcing”, VLDB’2012(Stanford)
• CrowdDB, “Answering Queries with Crowdsourcing”, SIGMOD’2011(Berkeley)
General Crowdsourcing Platforms

AMT

<table>
<thead>
<tr>
<th>Requester</th>
<th>HIT Expiration Date</th>
<th>Time Allotted</th>
<th>Reward</th>
<th>HITs Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sebastian Dar</td>
<td>Sep 19, 2012</td>
<td>30 minutes</td>
<td>$0.04</td>
<td>60499</td>
</tr>
<tr>
<td>Sebastian Dar</td>
<td>Sep 16, 2012</td>
<td>30 minutes</td>
<td>$0.06</td>
<td>28441</td>
</tr>
<tr>
<td>Richard Payne</td>
<td>Sep 10, 2012</td>
<td>10 minutes</td>
<td>$0.05</td>
<td>17518</td>
</tr>
<tr>
<td>Crowdsource</td>
<td>Aug 29, 2013</td>
<td>32 minutes</td>
<td>$0.10</td>
<td>14977</td>
</tr>
</tbody>
</table>
Can we extend the magic power of Crowdsourcing onto social network?
Microblog Users

• Simple
  – 140 characters
  – ‘RT’ + ‘@’

• But comprehensive
  – Large network
  – Various backgrounds of users
## Why Microblog Platform?

- **Twitter**
  - **Accessibility**: Highly convenient, on all kinds of mobile devices
  - **Incentive**: Altruistic or payment
  - **Supported tasks**: Simple task as decision making
  - **Communication Infrastructure**: ‘Tweet’ and ‘Reply’ are enough
  - **Worker Selection**: Active, Enabled by ‘@’

- **AMT**
  - **Accessibility**: Specific online platform
  - **Incentive**: Mostly monetary incentive
  - **Supported tasks**: Various types of tasks
  - **Communication Infrastructure**: Complex workflow control mechanism
  - **Worker Selection**: Passively, No exact selection
Outline

• Running Example
• Problem Definition
• Jury Selection Algorithms
• Evaluation
Motivation – Jury Selection Problem
Running Case(1)

• Given a decision making problem, with budget $1, whom should we ask?

Is Döner Kebab available in Hong Kong?
Motivation – Jury Selection Problem
Running Case(2)

• $\epsilon$: error rate of an individual
• $r$: requirement of an individual, can be virtual
• Majority Voting to achieve final answer

Is Döner Kebab available in Hong Kong?
Motivation – Jury Selection Problem
Running Case(2)

• Worker : Juror
• Crowds : Jury
• Data Quality : Jury Error Rate

Is Döner Kebab available in Hong Kong?
Motivation – Jury Selection Problem
Running Case(3)

• If (A, B, C) are chosen (Majority Voting)
  
  \[
  JER(A,B,C) = 0.1 \times 0.2 \times 0.2 + (1 - 0.1) \times 0.2 \times 0.2 + 0.1 \times (1 - 0.2) \times 0.2 + 0.1 \times 0.2 \times (1 - 0.2) = 0.072
  \]

  Better than A(0.1), B(0.2) or C(0.2) individually

Is Döner Kebab available in Hong Kong?

• If (A, B, C) are chosen (Majority Voting)
  
  \[
  JER(A,B,C) = 0.1 \times 0.2 \times 0.2 + (1 - 0.1) \times 0.2 \times 0.2 + 0.1 \times (1 - 0.2) \times 0.2 + 0.1 \times 0.2 \times (1 - 0.2) = 0.072
  \]

  Better than A(0.1), B(0.2) or C(0.2) individually
Motivation – Jury Selection Problem
Running Case(4)

• What if we enroll more
  – \( \text{JER}(A, B, C, D, E) = 0.0704 < \text{JER}(A, B, C) \)
  – The more the better?

Is Döner Kebab available in Hong Kong?
Motivation – Jury Selection Problem

Running Case(5)

• What if we enroll even more?
  – \( \text{JER}(A,B,C,D,E,F,G) = 0.0805 > \text{JER}(A,B,C,D,E) \)
  – Hard to calculate JER
Motivation – Jury Selection Problem
Running Case(6)

• So just pick up the best combination?
  – JER(A,B,C,D,E)=0.0704
  – R(A,B,C,D,E) = $1.6 > budget($1.0)
Motivation – Jury Selection Problem
Running Case (7)

<table>
<thead>
<tr>
<th>Crowd</th>
<th>Individual Error-rate</th>
<th>Jury Error-rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>A</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>C,D,E</td>
<td>0.2, 0.2, 0.3</td>
<td>0.174</td>
</tr>
<tr>
<td>A,B,C</td>
<td>0.1, 0.2, 0.2</td>
<td>0.072</td>
</tr>
<tr>
<td>A,B,C,D,E</td>
<td>0.1, 0.2, 0.2, 0.3, 0.3</td>
<td>0.0703</td>
</tr>
<tr>
<td>A,B,C,D,E,F,G</td>
<td>0.1, 0.2, 0.2, 0.3, 0.3, 0.4, 0.4</td>
<td>0.0805</td>
</tr>
</tbody>
</table>

Worker selection for maximize the quality of a particular type of product: the reliability of voting.
Outline

• Motivation
• **Problem Definition**
• Jury Selection Algorithms
• Evaluation
Problem Definition

• Jury and Voting

**Definition 1  (Jury).** A jury $J_n = \{j_1, j_2, \cdots, j_n\} \subseteq S$ is a set of jurors with size $n$ that can form a voting.

<table>
<thead>
<tr>
<th>Jury $J_n$ = {j_1, j_2, j_3} with 3 jurors</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$j_1$</th>
<th>$j_2$</th>
<th>$j_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε(0.1)</td>
<td>ε(0.3)</td>
<td>ε(0.2)</td>
</tr>
<tr>
<td>r($0.3$)</td>
<td>r($0.4$)</td>
<td>r($0.2$)</td>
</tr>
</tbody>
</table>

**Definition 2  (Voting).** A voting $V_n$ is a valid instance of a jury $J_n$ with size $n$, which is a set of binary values.

<table>
<thead>
<tr>
<th>Voting $V_n$ = {1,0,1} from $J_n$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$j_1$</th>
<th>$j_2$</th>
<th>$j_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε(0.1)</td>
<td>ε(0.3)</td>
<td>ε(0.2)</td>
</tr>
<tr>
<td>r($0.3$)</td>
<td>r($0.4$)</td>
<td>r($0.2$)</td>
</tr>
</tbody>
</table>

| 1 | 0 | 1 |


**Problem Definition**

- **Voting Scheme**

**Definition 3** (Majority Voting - MV). Given a voting $V_n$ with size $n$, Majority Voting is defined as

$$MV(V_n) = \begin{cases} 1 & \text{if } \sum j_i \geq \frac{n+1}{2} \\ 0 & \text{if } \sum j_i \leq \frac{n-1}{2} \end{cases}$$

A Voting $V_n = \{1,0,1\}$ from $J_n$

$$MV(V_n) = 1, \quad (\sum j_i = 2 > 1)$$
Problem Definition

• Invididual Error-rate

**Definition 4** (Individual Error Rate - $\epsilon_i$). The individual error rate $\epsilon_i$ is the probability that a juror conducts a wrong voting. Specifically

$$\epsilon_i = \Pr(\text{vote otherwise} | \text{a task with ground truth A})$$

A Voting $V_n = \{1, 0, 1\}$ from $J_n$

**Definition 5** (Carelessness - $C$). The Carelessness $C$ is defined as the number of mistaken jurors in a jury $J_n$ during a voting, where $0 \leq C \leq n$. 
Problem Definition

Definition 6 (Jury Error Rate - JER(Jₙ)). The jury error rate is the probability that the Carelessness C is greater than \( \frac{n+1}{2} \) for a jury \( Jₙ \), namely

\[
JER(Jₙ) = \sum_{k=\frac{n+1}{2}}^{n} \sum_{A \in Fₖ} \prod_{i \in A} \epsilon_i \prod_{j \in A^c} (1 - \epsilon_j)
\]

\[
= \Pr(C \geq \frac{n + 1}{2} | Jₙ)
\]

where \( Fₖ \) is all the subsets of \( S \) with size \( k \) and \( \epsilon_i \) is the individual error rate of juror \( j_i \).

\[
JER(J₃) = 0.1 \times 0.3 \times 0.2 + (1-0.1) \times 0.3 \times 0.2 + 0.1 \times (1-0.3) \times 0.2 + 0.1 \times 0.3 \times (1-0.2)
\]

\[
= 0.029
\]
Problem Definition

• Crowdsourcing Models (model of candidate microblog users)

**Definition 7** (Altruisim Jurors Model - AltrM). While selecting a jury $J$ from all candidate jurors (choosing a subset $J \subseteq S$), any possible jury is allowed.

**Definition 8** (Pay-as-you-go Model - PayM). While selecting a jury $J$ from all candidate jurors (choosing a subset $J \subseteq S$), each candidate juror $j_i$ is associated with a payment requirement $r_i$ where $r_i \geq 0$, the possible jury $J$ is allowed when the total payment of $J$ is no more than a given budget $B$, namely $\sum_{j_i \in J} r_i \leq B.$
Problem Definition

• Jury Selection Problem (JSP)

**Definition 9** (Jury Selection Problem - JSP). Given a candidate juror set $S$ with size $|S| = N$, a budget $B \geq 0$, a crowdsourcing model (AltrM or PayM), the Jury Selection Problem (JSP) is to select a jury $J_n \subseteq S$ with size $1 \leq n \leq N$, that $J_n$ is allowed according to crowdsourcing model and $JER(J_n)$ is minimized.

We hope to form a Jury $J_n$, allowed by the budget, and with lowest JER
Outline

• Motivation
• Problem Definition
• Jury Selection Algorithms
• Evaluation
Computation of Jury Error Rate

• The number of careless jurors (*Carelessness*-C) is a random variable following Poisson Binomial Distribution

\[
JER(J_n) = \sum_{k=\frac{n+1}{2}}^{n} \sum_{A \in F_k} \prod_{i \in A} \epsilon_i \prod_{j \in A^c} (1 - \epsilon_j)
\]

\[
= \Pr(C \geq \frac{n + 1}{2} | J_n)
\]

• The naïve computation of JER is exponentially increasing
Computation of Jury Error Rate (2)

- Alg1: Dynamic Programming to compute JER in $O(n^2)$

**Lemma 1.** The calculation of JER of Jury with size $n$ can be split into smaller ones:

$$
\Pr(C \geq L|J_n) = \Pr(C \geq L - 1|J_{n-1}) \cdot \epsilon_n + \Pr(C \geq L|J_{n-1}) \cdot (1 - \epsilon_n)
$$

where

$$
\Pr(C \geq 0|J_m) = 1 \quad \forall \quad 0 \leq m \leq n
$$

$$
\Pr(C \geq m|J_n) = 0 \quad \forall \quad m > n
$$
Computation of Jury Error Rate (3)

- **Alg2**: Convolution-based to compute JER in $O(n \log^2 n)$
  - Treat probability distribution as coefficients of polynomials
  - Divide larger jury in two smaller juries
  - Merge by polynomial multiplication
    - Can be speeded up by using FFT
Computation of Jury Error Rate(4)

- **Alg2**: Convolution-based to compute JER in $O(n\log^2 n)$

```
Algorithm 1 Convolution-based Algorithm (CBA)

<p>| Input: | A jury $J_n$ |</p>
<table>
<thead>
<tr>
<th>Output:</th>
<th>the vector of distribution of $C$, $D_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:</td>
<td>if $n = 1$ then</td>
</tr>
<tr>
<td>2:</td>
<td>$D_C[0] = 1 - \epsilon_1$ ;</td>
</tr>
<tr>
<td>3:</td>
<td>$D_C[1] = \epsilon_1$ ;</td>
</tr>
<tr>
<td>4:</td>
<td>return $D_C$ ;</td>
</tr>
<tr>
<td>5:</td>
<td>else</td>
</tr>
<tr>
<td>6:</td>
<td>Dividing $J_n$ into two parts: $J_{n1}$ and $J_{n2}$, where $</td>
</tr>
<tr>
<td>7:</td>
<td>$D_{C1} = CBA(J_{n1})$ ;</td>
</tr>
<tr>
<td>8:</td>
<td>$D_{C2} = CBA(J_{n2})$ ;</td>
</tr>
<tr>
<td>9:</td>
<td>$D_C =$ convolution of $D_{C1}$ and $D_{C2}$ ;</td>
</tr>
<tr>
<td>10:</td>
<td>end if</td>
</tr>
<tr>
<td>11:</td>
<td>return $D_C$ ;</td>
</tr>
</tbody>
</table>
```

- Divide into two smaller juries
- Merge, using FFT to speed up convolution
Computation of Jury Error Rate (5)

- Alg3: lower bound of JER in $O(n)$ time
  - Paley-Zygmund inequality

**Lemma 3** (Lower Bound-based Pruning). Given a jury with size $n$, the lower bound of $JER(J_n)$ is shown as follows,

$$JER(J_n) \geq \frac{(1-\gamma)^2 \mu^2}{(1-\gamma)^2 \mu^2 + \sigma^2}$$

where $\mu = \sum_{i=1}^{n} \epsilon_i, \sigma^2 = \sum_{i=1}^{n} (1-\epsilon_i)\epsilon_i$, and $\gamma = (\frac{n+1}{2} / \mu) \in (0, 1)$. 

33
JSP on AltrM(1)

• Monotonicity with given jury size on varying individual error-rate

**Lemma 4.** The lowest JER originates from the Jurors with lowest individual error-rate among the candidate jurors set $S$.

**Proof.** W.l.o.g, we pick one $j_i$ of the $n$ jurors in a given Jury $J_n$ with size $n$. Then $JER(J_n)$ can be transformed as below:

$$JER(J_n) = \Pr(C \geq \frac{n+1}{2}|J_n)$$

$$= \epsilon_i(\Pr(C \geq \frac{n+1}{2} - 1|J_{n-1})) + (1 - \epsilon_i)(\Pr(C \geq \frac{n+1}{2}|J_{n-1}))$$

$$= \epsilon_i(\Pr(C = \frac{n+1}{2} - 1|J_{n-1}) + (\Pr(C \geq \frac{n+1}{2}|J_{n-1}))$$

$$= \epsilon_i \cdot A + B$$

• In English: “best jury comes from best jurors”
• Decide the size
JSP on AltrM(2)

- Algorithm for JSP on AltrM

Alg_AltrM{
1. Sort according to error-rate;
2. Starting from 1 to n, increase the jury size by two;
   1. Compute JER;
   2. Update best current jury;
3. Output best jury;
}

//keep the size odd

Might be convex, future work
JSP on PayM(1)

- Budget is a constraint
- Objective function is JER
- NP-hardness
  - Reduce to an *nth-order 0-1 Knapsack Problem*

\[
\text{optimize } \sum_{i_1 \in n} \sum_{i_2 \in n} \ldots \sum_{i_n \in n} V[i_1, i_2, \ldots, i_n] \cdot x_1 x_2 \ldots x_n
\]

Given an instance of traditional KP, we can construct an *nOKP* instance by defining the profit *n*-dimensional vector as \(V[i, i, \ldots, i] = p_i\) and \(V[\text{otherwise}] = 0\) for all \(i\), where \(p_i\) is the profit in traditional KP. The weight vector and objective value remain the same. \(\Box\)
• Approximate Algorithm

Alg_PayM{
1. Sort according to \((\text{requirement} \times \text{error-rate})\);
2. Starting from 1 to n, increase the jury size by two;
   1. Keep track of pair;       //increment might be conducted by size of 1
   2. Check whether adding new juror will exceed budget;
   3. If so, compute and compare JER;
   4. Update best current jury;
3. Output best jury;
}

JSP on PayM(2)
Framework

A Decision Making Task

GIVEN
n users

Integrated Constraint
\( r_1, r_2, \ldots, r_n \)

Individual Error-rate
\( \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \)

User Activity
\( t_1, t_2, \ldots, t_n \)

User Experience
\( e_1, e_2, \ldots, e_n \)

Profiling of users

Knowledge diffusion graph based on “RT”

CHOSE
A subset to vote

NP-hardness

Poisson Binomial Distribution

Rank users with PageRank and HITS

Majority Voting

YES

OR

NO
Outline

• Motivation
• Problem Definition
• Jury Selection Algorithm
• Evaluation
Parameter Estimation

• How to estimate such parameter is itself a research topic

• Individual Error Rate ($\epsilon$) -- ‘RT’ graph
  – PageRank and HITS
  – The score in rank is normalized to be the individual error rate

• Integrated requirement ($r$) – account info
  – Account Age and Account Activity
Data Preparation

• We test our algorithms on both synthetic data and real Twitter data
• Varying
  – Size
  – Mean
  – Variance
• 3.4GHz Win7 PC, programmed in C++
• Mean = 0.5 is the turning point
• On the right side, “truth rests in the hands of a few.”
While the budget increases
• The total cost also increases
• The jury error rate decreases
• Green – Accurate Algorithm (test with N=20)
• Blue – approximation algorithm
  • $O(n \log n)$
  • Good approximation on JER

(e) APPX v.s. OPT on Total Cost
(f) APPX v.s. OPT on JER
Take-away and Future Work

• Take-away
  – Cultivate a pool of candidate jurors
  – JER deceases very fast according to the size of jury

• Future Work
  – Beyond direct payment
    • Prediction Market
  – Beyond decision making
    • Campaign Boosting
Thank You

• Q & A

Is Döner Kebab available in Hong Kong?