Market Approaches to Aggregating Predictions and Data

Bo Waggoner
U. Colorado, Boulder

Makerere University
July 2019
Goal: **acquire** and **aggregate** information
Goal: **acquire** and **aggregate** information

- **beliefs** about future events or relationships
  *e.g. forecasting rainfall, crop growth, sales*

- **data** about individuals or processes
  *e.g. farming data, sales data*
Challenges:

- **acquiring** accurate and useful information
  incentives!

- **aggregating** the information accurately
  consider polls or surveys ... systematic bias, etc.
Outline:

1. Prediction markets - overview
2. Collaborative machine learning
3. Markets for data
Outline:

1. Prediction markets - overview
2. Collaborative machine learning
3. Markets for data
Prediction markets: goal

Predict a future event
- Political election
- Sporting event
- Weather
- Economics
- ...

1  Designer chooses initial prediction $p^0$
Prediction markets: mechanism

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
Prediction markets: mechanism

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
3. Second participant updates it to $p^2$
Prediction markets: mechanism

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
3. Second participant updates it to $p^2$
4. ...
Prediction markets: mechanism

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
3. Second participant updates it to $p^2$
4. ...
5. **Event occurs**
Prediction markets: mechanism

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
3. Second participant updates it to $p^2$
4. ...
5. **Event occurs**
6. Designer pays participants
Prediction markets: mechanism

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
3. Second participant updates it to $p^2$
4. ...
5. Event occurs
6. Designer pays participants
   How?
Building block: proper scoring rules

First step: incentivize single forecaster

1. Forecaster predicts $p$
Building block: proper scoring rules

First step: incentivize **single forecaster**

1. Forecaster predicts $p$
2. Event $y$ occurs

Payoff $S(p, y)$

$S$ is proper if truthfulness maximizes expected score

Examples:

$S(p, y) = \log p(y)$,

$S(p, y) = \| p - \delta y \|_2^2$

$\delta y$ = indicator vector for $y$, i.e. $(0, ..., 1, ..., 0)$. 
Building block: proper scoring rules

First step: incentivize single forecaster

1. Forecaster predicts $p$
2. Event $y$ occurs
3. Payoff $S(p, y)$

$S(p, y) = \log p(y)$, $S(p, y) = \|p - \delta y\|_2^2$

$\delta y$ = indicator vector for $y$, i.e. $(0, \ldots, 1, \ldots, 0)$.
Building block: proper scoring rules

First step: incentivize **single forecaster**

1. Forecaster predicts $p$
2. **Event** $y$ **occurs**
3. Payoff $S(p, y)$
4. $S$ is **proper** if truthfulness maximizes expected score
Building block: proper scoring rules

First step: incentivize **single forecaster**

1. Forecaster predicts $p$
2. **Event** $y$ **occurs**
3. Payoff $S(p, y)$
4. $S$ is **proper** if truthfulness maximizes expected score

Examples: $S(p, y) = \log p(y)$
First step: incentivize single forecaster

1. Forecaster predicts $p$
2. Event $y$ occurs
3. Payoff $S(p, y)$
4. $S$ is proper if truthfulness maximizes expected score

Examples: $S(p, y) = \log p(y)$, $S(p, y) = \|p - \delta_y\|_2^2$

$\delta_y = \text{indicator vector for } y$, i.e. $(0, \ldots, 1, \ldots, 0)$. 
Scoring rule based market$^1$

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
3. Second participant updates it to $p^2$
4. ...
5. **Event $y$ occurs**

$^1$[Hanson 2003]
Scoring rule based market$^1$

1. Designer chooses initial prediction $p^0$
2. First participant updates it to $p^1$
3. Second participant updates it to $p^2$
4. ...
5. **Event** $y$ occurs
6. Participant $t$ receives $S(p^t, y) - S(p^{t-1}, y)$

$^1$[Hanson 2003]
Some incentive properties

- Each person only participates once $\implies$ **truthful**
  otherwise, complicated ... e.g. [Chen, W. 2016]
Some incentive properties

- Each person only participates once $\implies$ truthful
  otherwise, complicated ... e.g. [Chen, W. 2016]

- In equilibrium, converges* to optimal accuracy
  * e.g. [Ostrovsky 2013]
Some incentive properties

- Each person only participates once $\implies$ **truthful**
  otherwise, complicated \ldots e.g. [Chen, W. 2016]

- In *equilibrium*, converges* to **optimal accuracy**
  * e.g. [Ostrovsky 2013]

- Can be rephrased as a **financial market**
  [Hanson 2003, \ldots, Abernethy, Chen, Wortman-Vaughan 2013]
Some incentive properties

- Each person only participates once $\Rightarrow$ **truthful** otherwise, complicated ... e.g. [Chen, W. 2016]
- In *equilibrium*, converges* to **optimal accuracy**
  * e.g. [Ostrovsky 2013]
- Can be rephrased as a **financial market**
  [Hanson 2003, ..., Abernethy, Chen, Wortman-Vaughan 2013]
- Extends to **expectations of random variables**...
  e.g. [ACW13]
Some incentive properties

- Each person only participates once $\implies$ **truthful** otherwise, complicated \ldots e.g. [Chen, W. 2016]
- In *equilibrium*, converges* to **optimal accuracy** * e.g. [Ostrovsky 2013]
- Can be rephrased as a **financial market** [Hanson 2003, \ldots, Abernethy, Chen, Wortman-Vaughan 2013]
- Extends to **expectations of random variables**\ldots e.g. [ACW13]
- \ldots and beyond!? 

*Coming up: machine learning connection*
Recap so far

**Scoring-rule based markets (SRMs) for predicting future events**

- Collaboratively maintain a single estimate/prediction
- Participants propose updates
- Reward is **improvement in score**
- Better predictions $\implies$ higher rewards
Outline:

1 Prediction markets - overview
2 Collaborative machine learning
3 Markets for data
Next step: expectations

**Goal:** predict expectation of random variable
Next step: expectations

**Goal:** predict expectation of random variable

**Need:** “proper scoring rule” for the mean
Next step: expectations

**Goal:** predict expectation of random variable

**Need:** “proper scoring rule” for the mean

**Question:** minimizing **which loss** gives the mean?

\[
\arg \min_r \mathbb{E}_{y \sim p} \ell(r, y)
\]
Next step: expectations

Goal: predict expectation of random variable

Need: “proper scoring rule” for the mean

Question: minimizing which loss gives the mean?

\[
\arg \min_r \mathbb{E}_{y \sim p} \ell(r, y)
\]

Example: Squared loss, \( \ell(r, y) = (r - y)^2 \)

For vectors: \( \|r - y\|_2^2 \); there are others
Prediction market for expectations

Example: expected cm of rain next month

1. Designer chooses initial estimate $r^0$
2. First participant updates it to $r^1$
3. Second participant updates it to $r^2$
4. ...
5. **Event $y$ occurs** e.g. total rainfall measured
6. Participant $t$ receives $\ell(r^{t-1}, y) - \ell(r^t, y)$
   where $\ell(r, y) = (r - y)^2$
Other kinds of predictions

Can extend to any *elicitable* statistic...

[Lambert, Pennock, Shoham 2008; Abernethy, Frongillo 2011]

- Median
- Mode
- ...
Other kinds of predictions

Can extend to any *elicitable* statistic...

[Lambert, Pennock, Shoham 2008; Abernethy, Frongillo 2011]

- Median $|r - y|$
- Mode $\mathbb{1}[r = y]$
- ...

...though financial market properties may not extend

[Frongillo, W. 2018]
Collaborative machine learning

Key idea from [Abernethy, Frongillo 2011]:
use a test dataset instead of the future event!
**Example:** classifier to predict sun or rain based on data

1. Designer chooses initial **classifier** $h^0$
2. First participant updates it to $h^1$
3. ...
Collaborative machine learning

**Example:** classifier to predict sun or rain based on data

1. Designer chooses initial **classifier** $h^0$
2. First participant updates it to $h^1$
3. ...
4. **Designer picks test dataset**
   
   e.g. *random historical days*
**Collaborative machine learning**

**Example:** classifier to predict sun or rain based on data

1. Designer chooses initial **classifier** $h^0$
2. First participant updates it to $h^1$
3. ...
4. **Designer picks test dataset**
   
   *e.g. random historical days*
5. Participant $t$ receives $\ell(h^{t-1}; D) - \ell(h^t; D)$

   where $\ell(h; D)$ is average loss on dataset
Implications

Structured as kaggle-like contest, but...
Implications

Structured as kaggle-like contest, but...

- **collaborative** rather than **competitive**
Implications

Structured as kaggle-like contest, but...

- **collaborative** rather than **competitive**
- **split rewards** rather than **winner-take-all**
Implications

Structured as kaggle-like contest, but...

- **collaborative** rather than **competitive**
- **split rewards** rather than **winner-take-all**
- **incentive-aligned**  

*does not encourage wild guesses*
Outline:

1. Prediction markets - overview
2. Collaborative machine learning
3. Markets for data
Markets for data

[Waggoner, Frongillo, Abernethy 2015]

**Idea:** instead of updating the model directly...
Markets for data

[Waggoner, Frongillo, Abernethy 2015]

Idea: instead of updating the model directly... people provide data, and we compute the updates!

![Diagram showing flow of data from people to a central node, with test data input and processing]
Markets for data

Key points:

- Reward for data = **improvement in loss**
- Incentive-aligned: better data = better payoff
- Fake data is ok!
Extensions

[Waggoner, Frongillo, Abernethy 2015]

If hypotheses lie in an RKHS (use kernels):

- Can provide **differential privacy** for data
- Can still phrase as a **market** with securities

*not generally true: [Frongillo, Waggoner 2018]*
Collaborative ML on Blockchain

[Harris, Waggoner, IEEE Blockchain 2019]
Collaborative ML on Blockchain

[Harris, Waggoner, IEEE Blockchain 2019]

1. Initialize ML model in a smart contract
Collaborative ML on Blockchain

[Harris, Waggoner, IEEE Blockchain 2019]

1. Initialize ML model in a smart contract
2. Participants arrive, provide data
Collaborative ML on Blockchain

[Harris, Waggoner, IEEE Blockchain 2019]

1. Initialize ML model in a smart contract
2. Participants arrive, provide data
3. Model automatically updates

Implementation on the Ethereum blockchain: https://github.com/microsoft/0xDeCA10B
Collaborative ML on Blockchain

[Harris, Waggoner, IEEE Blockchain 2019]

1. Initialize ML model in a smart contract
2. Participants arrive, provide data
3. Model automatically updates
4. Model is free and open for all to use

Implementation on the Ethereum blockchain: https://github.com/microsoft/0xDeCA10B
Collaborative ML on Blockchain

[Harris, Waggoner, IEEE Blockchain 2019]

1. Initialize ML model in a smart contract
2. Participants arrive, provide data
3. Model automatically updates
4. Model is **free and open for all to use**
5. Can use prediction-market reward structure

Implementation on the **Ethereum blockchain**:
https://github.com/microsoft/0xDeCA10B
Recap and applications

Using a prediction market structure:

- **incentivizes** providing good data or predictions
- **aggregates** into a single, collaborative ML model

Possible applications: farming, maps, personal assistants, recommendations, ...
Future work

- Implement and **deploy** these mechanisms!
  
  *work with domain experts*

- Decrease **risk**
  
  *currently: participants may lose money*

- Other reward mechanisms?

- Generally: marketplaces for data

---

Thanks to my collaborators: Raf Frongillo (U. Colorado), Yiling Chen (Harvard), Jake Abernethy (Georgia Tech), Justin Harris (Microsoft Research).
Future work

- Implement and **deploy** these mechanisms!
  *work with domain experts*

- Decrease **risk**
  *currently: participants may lose money*

- Other reward mechanisms?

- Generally: marketplaces for data

Thanks to my collaborators: Raf Frongillo (U. Colorado), Yiling Chen (Harvard), Jake Abernethy (Georgia Tech), Justin Harris (Microsoft Research).

Thank you!