A COLLABORATIVE MECHANISM FOR CROWDSOURCING PREDICTION PROBLEMS

CHRISTOPHER LEE

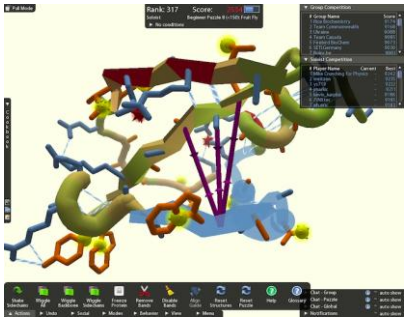
STEWART RICHARDSON

CS286R

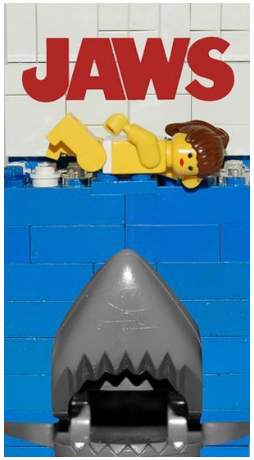
NOVEMBER 19, 2012

WHAT ARE SOME COMPETITIONS THAT DRAW FROM CROWD EXPERTISE?

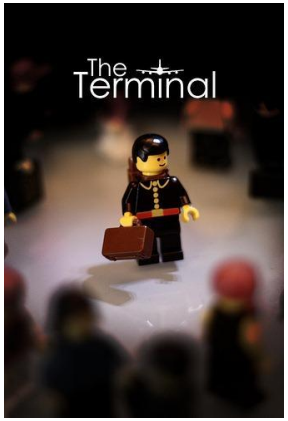
- NASA Space Antenna Design (EE/AI)
- FoldIT – University of Washington Protein Project (BIO/AI)
- DARPA Urban Grand Challenge (ENG/AI)
- Amazon Studios (ART)
- Netflix Prize (ML)

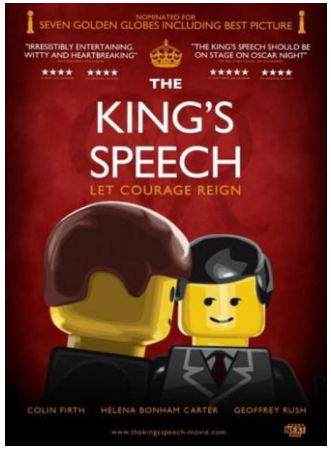


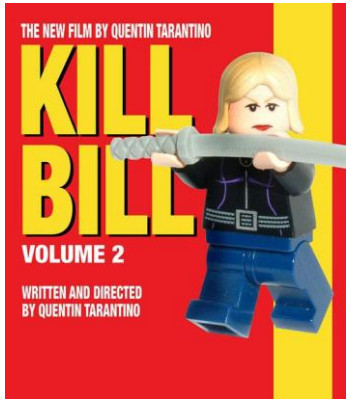

NETFLIX


JAWS


THE WOLFPACK IS BACK
THE HANGOVER PART II
BANGKOK HAS THEM NOW
MEMORIAL DAY


The Terminal


SEVEN GOLDEN GLOBES INCLUDING BEST PICTURE
THE KING'S SPEECH
LET COURAGE REIGN
COLIN FIRTH
HELENA BONHAM CARTER
GEOFFREY RUSH


THE NEW FILM BY QUENTIN TARANTINO
KILL BILL
VOLUME 2
WRITTEN AND DIRECTED BY QUENTIN TARANTINO

NETFLIX PRIZE

- **Netflix offered $1,000,000 Grand Prize for 10% improvement collaborative filtering algorithm to predict ratings for films, and $50,000 Progress Prize for 1% improvement**

- **Prediction competitions proven successful:**
  - Leverage the abilities and knowledge of the public at large, commonly known as crowdsourcing
  - Provide an incentivized mechanism for an individual or team to apply their own knowledge and techniques

- **Factoid:** AT&T Bell Labs's Pragmatic Chaos won the first prize, out of 50k contestants

WEAKNESSES OF NETFLIX PRIZE MARKET DESIGN?

1. It is anti-collaborative.
2. The incentives can be skewed and misaligned.
3. The winner-take-all prize structure means that 2nd place is as good as not competing at all.
4. Anything else?

CROWDSOURCED LEARNING MECHANISM IN LAYMAN'S TERMS?

What?

To use the framework of a prediction market as a tool for ~~information~~ “more complex knowledge" aggregation.

How?

Your payoff is directly correlated to your contribution measured at the end.

Why?

To serve the purpose of “aggregating” a hypothesis for a given learning problem.

FORMAL DEFINITION OF CLM

1. **Mechanism that allows agents to modify hypothesis by wagering on a contribution; payoff maps to contribution improvement**

2. **Allows participants to collaboratively “learn” a hypothesis for a given prediction task**

3. **The approach draws heavily from the concept of a prediction market, but leveraging the cooperation of a crowd**

PRESENTATION OVERVIEW

1. **Review** of Scoring Rules, Prediction Markets, ML concepts that lead to…

2. **Contributions** of Abernethy and Frongillo's Crowdsourced Learning Mechanism (CLM) to fields of computer science and economics before…

3. **Ending** the Crowdsourcing Unit in regards to Crowdsourcing Quality (Karger, 2011), Crowdsourcing Contest (Cavallo, 2012), other readings, etc.

REMEMBER MONTHLY WEATHER REVIEW?

- *Monthly Weather Review* -> Verification scheme of weather forecast with multiple experts, then we measure the predictions once an outcome is known

- Many authors point to the paper of Brier (1950) as the earliest mention of what we now call a *proper scoring rule* in economics literature, or called a *proper loss* in ML literature.

- Led to the development of MSR, combinatorial outcome spaces, logarithmic scoring rule, design of prediction market, Bregman divergence, etc.

REMEMBER STRICTLY PROPER SCORING RULE? (REVIEW)

Proper Scoring Rule Review (Savage, 1971):

- Scoring rule rewards expert S(p, ω) where p is the prediction, and ω is the realized outcome
- This means that the score may not always be maximized when the true beliefs are submitted.

<u>Strictly</u> **Proper Scoring Rule Review (Gneiting and Raftery, 2007):**

- A scoring rule is strictly proper iff:

$$\mathbf{q} = \arg\max_{\mathbf{p} \in \Delta_n} \sum_{\omega} q_\omega S(\mathbf{p}, \omega)$$

- This means that the score <u>will</u> be maximized when true beliefs are submitted, thereby aligning incentives.

PREDICTION MARKETS (REVIEW)

- A **prediction market** is a financial mechanism whose purpose, given some uncertain future outcome, is to aggregate the subjective probability beliefs of this outcome from a large crowd of individuals (Hanson, 2003; Lecture 5)

- **Scoring Rules <> Design of Prediction Markets**
    1. Can use scoring rule not only to elicit correct forecasts from single individual, but also from multiple *agents*
    2. Central authority called *market maker* continue to publish joint forecast representing consensus
    3. Current consensus probability P_t is posted, and any trader can place bet by modifying probability P_{t+1}

$$\textbf{\textit{Payout}} = S(P_{t+1}, \omega) - S(P_t, \omega)$$

FROM SCORING RULES (ECON) TO LOSS FUNCTIONS (ML)

- A **scoring rule** (terminology drawn from economics literature) denotes the same concept as a **loss function (**terminology used in machine learning literature) but in which the goal is to maximize rather than minimize

- **Scoring Rule (max) –** Used to rewards expert to give the most accurate prediction closest to outcome. **Argmax S(p, ω).**
- **Loss Function (min) –** Used to incentivize “correctness” in a learning algorithm**.** Measures the performance of a hypothesis on a set of data. **Argmin L(w, X).**

SCORING RULE <> LOSS FUNCTION

- **Loss function (L) as incentives?**
- **L-functions as General Scoring Rules (GSR)?**
- **GSR vs. SR?**

CLM PRESENTATION OVERVIEW

1. **<u>Relate</u>** *General Scoring Rules* **(GSR)** to Loss Functions **(L)**

2. **<u>Describe</u>** CLM in detail, and discuss how to structure particular scoring function L, and given incentives to minimize L (Cost, Profit, Payout)

3. **<u>Explain</u>** how an *Automated prediction market maker* **(APMM)** in prediction markets is a special case of CLM!

LOSS FUNCTIONS AS INCENTIVES (RELATE)

- **Scoring rules give a basis to rate the performance of an input**
  - S : *P*x*O* -> **R**
- **Loss functions do the same, but allow for a broader type of inputs**
  - Hypothesis space *H*
  - L : *H*x*O* -> **R**
  - Smaller loss is better

LOSS FUNCTIONS AS GSR (RELATE)

- **Redefine the problem:**
  - **w** is the hypothesis suggested by the crowd
  - X is the data we use with the hypothesis to calculate the loss
  - The mechanism designer's goal is to minimize loss

GENERALIZED SCORING RULES (RELATE)

- **Any loss function L(w;X) which has a nonempty convex set**

$$W_L(P) := \operatorname*{argmin}_{\mathbf{w}\in\mathcal{H}} \mathbb{E}_{X\sim P}[L(\mathbf{w}; X)].$$

  for every P.

- **Why do we require this?**

GSR VS SR (DETAIL)

- GSR doesn't have the same notion of properness as a traditional scoring rule
- GSR is used for a different purpose, from the mechanism designer's point of view
- GSR takes more abstract inputs

MSR BASED MARKETS (DETAIL)

- **How did they work?**
  - Myopic traders look to maximize score
  - Traders sequentially submit bids, pay cost of bid
  - Scores are calculated upon termination of market
  - Traders paid based on scores

GSR BASED: CLM (DETAIL)

- **Generalized scoring rules are implemented by a Crowdsourced Learning Mechanism**

Algorithm 1 Crowdsourced Learning Mechanism for $(\mathcal{H}, \mathcal{O}, \texttt{Cost}, \texttt{Payout})$

Mechanism sets initial hypothesis to some $\mathbf{w}_0 \in \mathcal{H}$
for rounds $t = 0, 1, 2, \ldots$ **do**
  Mechanism posts current hypothesis $\mathbf{w}_t \in \mathcal{H}$
  Some participant places a bid on the update $\mathbf{w}_t \mapsto \mathbf{w}'$
  Mechanism charges participant $\texttt{Cost}(\mathbf{w}_t, \mathbf{w}')$
  Mechanisms updates hypothesis $\mathbf{w}_{t+1} \leftarrow \mathbf{w}'$
end for
Market closes after T rounds and the outcome (test data) $X \in \mathcal{O}$ is revealed
for each t **do**
  Participant responsible for the update $\mathbf{w}_t \mapsto \mathbf{w}_{t+1}$ receives $\texttt{Payout}(\mathbf{w}_t, \mathbf{w}_{t+1}; X)$
end for

CLM (DETAIL)

- **How do they work?**
  - Myopic traders seek to minimize loss
  - Traders sequentially submit hypotheses, pay cost
  - Scores are calculated upon termination of market
  - Traders are paid based on scores

- **How did they work? (From Previous MSR Slide)**
  - Myopic traders look to maximize score
  - Traders sequentially submit bids, pay cost of bid
  - Scores are calculated upon termination of market
  - Traders paid based on scores

COST, PAYOUT, *PROFIT*

- **Each trader's profit is determined by their score:**

$$\texttt{Profit}(\mathrm{w}_1, \mathrm{w}_2; X) = L(\varphi(\mathrm{w}_1); X) - L(\varphi(\mathrm{w}_2); X).$$

$$\varphi : \mathcal{H} \to \mathcal{H}'$$

- **We could have** $\mathcal{H}' = \mathcal{H}$ **and** $\varphi = id_{\mathcal{H}}$

COST, *PAYOUT*, PROFIT

- **What is our Payout function?**
  - Should satisfy the *Escrow (ES)* property (why?):

$$\texttt{Payout}(\mathrm{w}, \mathrm{w}'; X) \geq 0 \text{ for all } X \in \mathcal{O}$$

  - Once we specify cost function, we say:

$$\texttt{Payout}(\mathrm{w}, \mathrm{w}'; X) = L(\mathrm{w}; X) - L(\mathrm{w}'; X) + \texttt{Cost}(\mathrm{w}, \mathrm{w}')$$

COST, PAYOUT, PROFIT

- **Cost function is not given any guidelines**
- **Mechanism designer may want to give vouchers to the first *m* traders**

MECHANISM DESIGNER'S COST

- **The mechanism designer has a simple cost:**

$$\sum_{t=0}^{T} \texttt{Profit}(\mathbf{w}_t, \mathbf{w}_{t+1}; X) = L(\mathbf{w}_0; X) - L(\mathbf{w}_T; X)$$

- **Desirable properties:**
  - Easy to compute
  - Can act as an insurance policy (how?)

DESIRED PROPERTIES OF CLM

- **Escrow**

- **Efficient Computation**
  - Cost and Payout are efficiently computable

- **Tractable Trading**
  - Traders may be on a budget
  - Must be able to efficiently compute an element of

$$\operatorname*{argmax}_{\mathbf{w}' \in \mathcal{H}} \left\{ \mathbb{E}_{X \sim P}\left[\texttt{Profit}(\mathbf{w}, \mathbf{w}', X)\right] : \texttt{Cost}(\mathbf{w}, \mathbf{w}') \leq B \right\}$$

EXAMPLE: DATA STREAM

- **A firm wants to encode in binary a mystery stream of *m* characters from an alphabet with *n* letters**
- **They have to pay for each contained in the encoded stream**
- **If letter *i* has frequency q(*i*), we use -log(q(*i*)) bits to encode it**

EXAMPLE: DATA STREAM

- **Loss function is**

$$L(\mathrm{q};\mathrm{p}) := \mathbb{E}_{i\sim p}\left[L(\mathrm{q};i)\right] = \mathrm{KL}(\mathrm{p};\mathrm{q}) + H(\mathrm{p})$$

- **Hypothesis is q**
- **X is a character sampled uniformly from the observed stream**

COST, PAYOUT

- **Cost function is efficiently computable:**

$$\mathtt{Cost}(\mathrm{q}, \mathrm{q}') := \max_{i \in [n]} \log(\mathrm{q}(i)/\mathrm{q}'(i))$$

- **Payout function is as well**

$$\mathtt{Payout}(\mathrm{q}, \mathrm{q}'; i) := \log(\mathrm{q}(i)/\mathrm{q}'(i)) + \mathtt{Cost}(\mathrm{q}, \mathrm{q}')$$

APMM

- **Characterized by:**
  - Outcome space $\mathcal{O}$
  - Share space $\mathcal{S}$
  - Cost function $C : \mathcal{S} \to \mathbb{R}$
  - Payout function $\rho : \mathcal{O} \to \mathbb{R}^n$

APMM TO CLM

- **Remember, a CLM is defined by**

$$(\mathcal{H}, \mathcal{O}, \texttt{Cost}, \texttt{Payout})$$

- **An APMM is defined by**

$$(\mathcal{S}, \mathcal{O}, \rho, C)$$

- **These are equivalent for**

$$\mathcal{H} = \mathcal{S}(= \mathbb{R}^n)$$
$$\texttt{Cost}(\mathrm{s}, \mathrm{s}') = C(\mathrm{s}') - C(\mathrm{s})$$
$$\texttt{Payout}(\mathrm{s}, \mathrm{s}'; X) = \rho(X) \cdot (\mathrm{s}' - \mathrm{s})$$

BREGMAN DIVERGENCE

- **Defines a distance**

$$D_R(\mathbf{x}, \mathbf{y})$$

- **Between members of a set**

$$\mathbf{x}, \mathbf{y} \in \mathrm{dom}(R)$$

- **Given a function on that set**

$$R : \mathbb{R}^d \to \mathbb{R}$$

BREGMAN DIVERGENCE

- **Formally:**

$$D_R(\mathbf{x}, \mathbf{y}) = R(\mathbf{x}) - R(\mathbf{y}) - \nabla R(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})$$

DIVERGENCE-BASED GSRS

Definition 3. *We say that a GSR* $L : \mathcal{H} \times \mathcal{O} \to \mathbb{R}$ *is* divergence-based *if there exists an alternative hypothesis space* $\mathcal{H}' \subset \mathbb{R}^m$*, for some* m*, where we can write*

$$L(\mathbf{w}; X) \equiv D_R(\rho(X), \psi(\mathbf{w})) + f(X) \tag{1}$$

for arbitrary maps $\rho : \mathcal{O} \to \mathcal{H}'$, $f : \mathcal{O} \to \mathbb{R}$*, and* $\psi : \mathcal{H} \to \mathcal{H}'$*, and any closed strictly convex* $R : \mathcal{H}' \to \mathbb{R}$ *whose convex conjugate* R^* *is finite on all of* $\mathbb{R}^m$.

- **This seems overly complex**
- **Don't worry about *f* and *psi***

BREGMAN DIVERGENCE TO APMM

Proposition 3. *An APMM* $(\mathcal{S}, \mathcal{O}, \boldsymbol{\rho}, C)$ *with a efficiently computable* C *satisfies the EC and TT properties.*

Theorem 2. *If APMM* $(\mathcal{S}, \mathcal{O}, \boldsymbol{\rho}, C)$ *implements a GSR* $L : \mathcal{H} \times \mathcal{O} \to \mathbb{R}$*, then* L *is divergence-based.*

EXAMPLES

1. DATA STREAM
2. LINEAR REGRESSION
3. NETFLIX

CROWDSOURCED LEARNING PROBLEMS

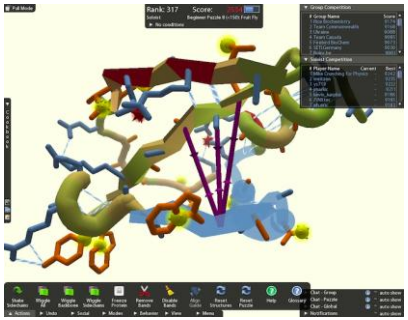
What are the strength and weaknesses of CLM structure?

(Crowd)

(Super Strengths & Weakness)

EXTENSIONS

- What are some other applications of prediction markets and/or crowdsourced learning mechanism?



MAIN POINTS

- **Contributions of Abernethy & Frongillo:**
  - Extend basic ideas from Prediction Markets to Crowdsourced Learning Mechanism

- **Grasp of intuitive technical points:**
  - The APMM is precisely a special case of a CLM!
  - How CLM works -> A = (H, O, Cost, Payout)

CLM IN CONTEXT OF THIS UNIT

- Human Computation
- Getting Crowds to Work
- Crowdsourcing Workflow Control
- Crowdsourcing Quality Control
- Crowdsourcing Contest
- Social Media and Social Influence
- Influence over Social Networks
- User-Generated Content
- **General Crowdsourcing Market**


HAPPY THANKSGIVING!!!