Applications

Boi Faltings

AI Laboratory, EPFL
boi.faltings@epfl.ch
http://liawww.epfl.ch/

2016

Applications

- Self-monitoring
- Community Sensing
- Prediction Polls
- Human Computation
- Peer Grading
- Product Reviews

Self-Monitoring

- ISP, mobile phone service, cloud computing services provided to many users under Service Level Agreements (SLA).
- Agreement will stipulate penalties for insufficient quality.
- Monitoring and proving insufficient quality is costly...
- ...and best done by users themselves.

⇒ self-monitoring with incentives for truthfulness.

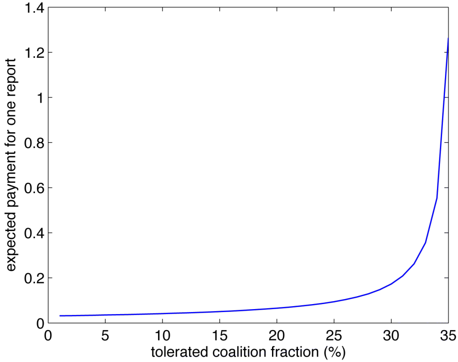
R. Jurca et al.: *Reliable QoS Monitoring Based on Client Feedback*, Proceedings of the 16th International World Wide Web Conference (WWW07), pp. 1003-1011, 2007

Example: Web Service Provider

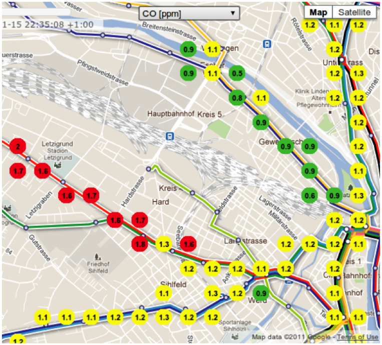
- Service (e.g. weather forecast) provided to a homogeneous population of users.
- 2 quality parameters:
  - Q_1: response before deadline $(0/1)$
  - Q_2: provided information is correct $(0/1)$
- cost of service $= 1$
- benefit for misreport $= 0.01$
- cost of misreporting (any number of reports): 10

Model for Peer Prediction

- Prior probability of $Q_{1/2} = 0.9$.
- Posterior probability changes 20%:
  $p(1|1) = 0.92, p(1|0) = 0.88$ and
  $p(0|0) = 0.12, p(0|1) = 0.08$.



Community Sensing



- Pollution is a distributed phenomenon - needs many sensors to form a map.
- Government sensors not considered trustworthy.

$\Rightarrow$ collect data from a community of sensors.

Example: Air Quality Egg

- Open source design developed in a kickstarter project
- Sold over 1000 times at $185.
- Measurements uploaded to a center controlled by manufacturer.
- Sensor quality insufficient for meaningful measurements.

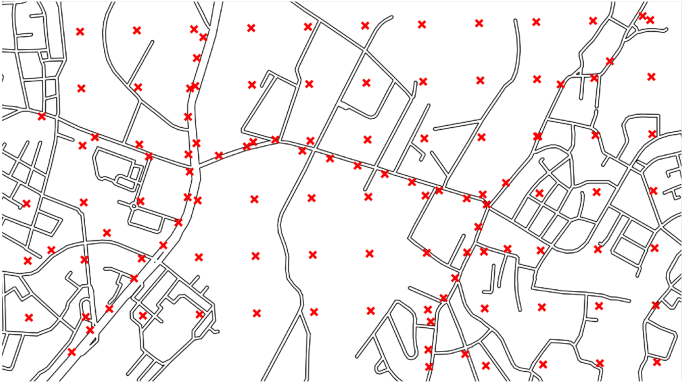
Peer Truth Serum applied to Community Sensing

- Mechanism requires peer reports, but no 2 sensors measure at the same location.
- Use prediction of Gaussian process model based on reports close by.
- Only reports in the same time interval are compared; previous aggregate is used as prior distribution R.

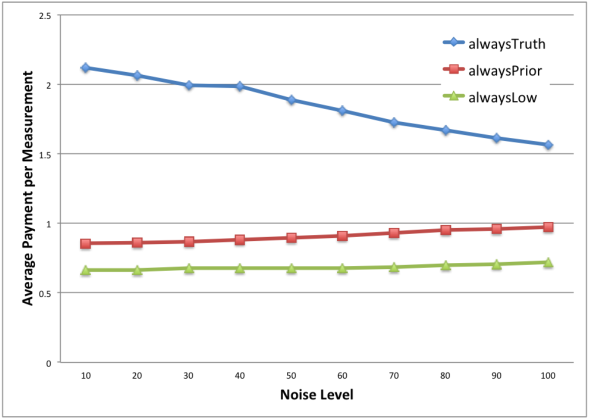
B. Faltings et al.: *Incentive Mechanisms for Community Sensing*, IEEE Transaction on Computers, 63(1), 115-128, 2014

Evaluation

- Simulation on an air pollution model of the city of Strasbourg (France):

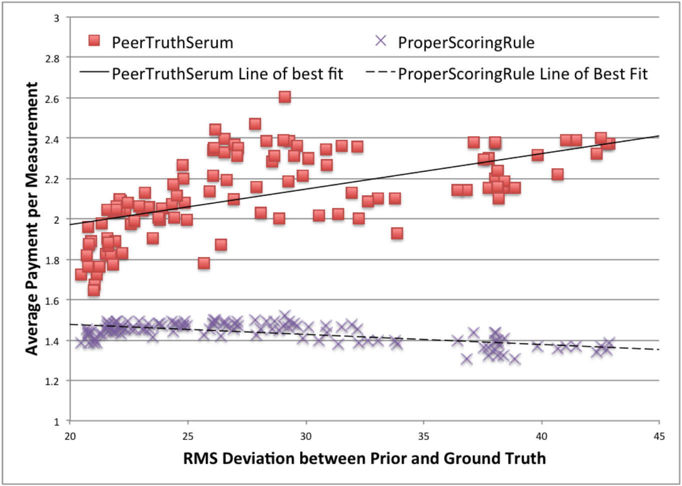


Rewards are Resistant to Noise



- Truthtelling remains the best strategy even with 100% measurement noise.

Rewarding Valuable Data

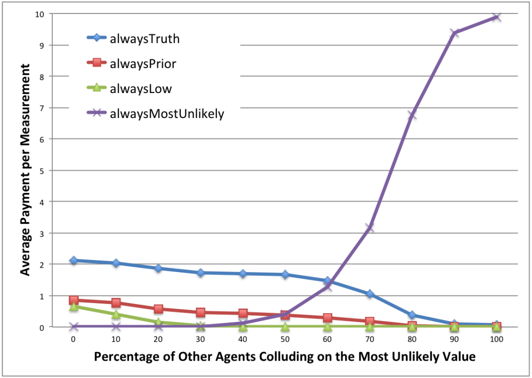


- PTS rewards data collection in uncertain areas, unlike proper scoring rules.

Collusion

- Incentives are vulnerable to collusion:
  - all agents report the same value x.
  - coordinate on x with smallest $R[x]$.
- Complicated by aggregation into Gaussian model.
- Collusion by many agents will be hard to implement.

Resistance to Collusion



- Collusion unprofitable unless $> 60\%$ of agents participate.
- Least likely value is the only promising strategy, but carries no reward unless at least half the agents participate.

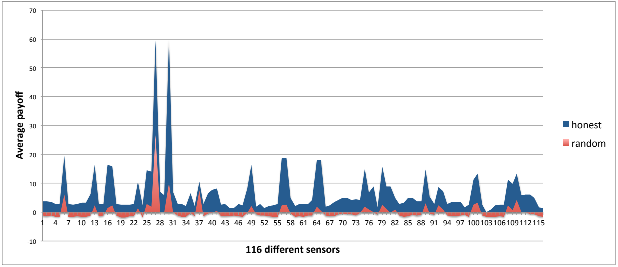
Peer Truth Serum for Crowdsourcing

- Same simulation model (Strasbourg).
- Use measurement distribution in entire area for R.
- Use correlated neighbouring sensors as peers.
- Consider NO_2 pollution, separated into 4 discrete levels.

G.Radanovic et al.: *Incentives for Effort in Crowdsourcing using the Peer Truth Serum*, ACM TIST, 2016

Static Sensors

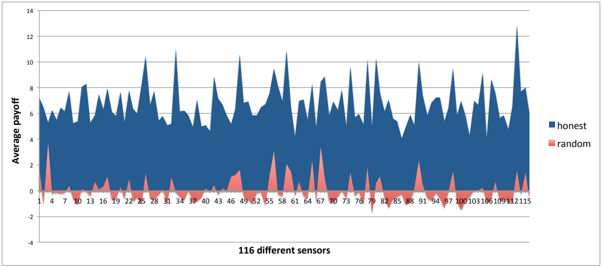
Strategy	mean	min	max	median
honest	6.779	-0.03	59.969	3.658
collude	2.323	-0.146	21.769	1.045
colludeLow	0	0	0	0
random	0.022	-1.974	26.779	-1.076
randomAll	0.071	-2.161	2.137	0.175



Locations with unusual values are more profitable.

Mobile Sensors

Strategy	mean	min	max	median
honest	6.779	4.064	12.941	6.456
collude	2.323	1.052	5.141	2.027
colludeLow	0	0	0	0
random	-0.008	-1.781	3.714	-0.294
randomAll	0.03	-1.446	1.792	-0.109



Mobility results in equal opportunities (note different scale).

Prediction Polls

Classical opinion polls no longer work:

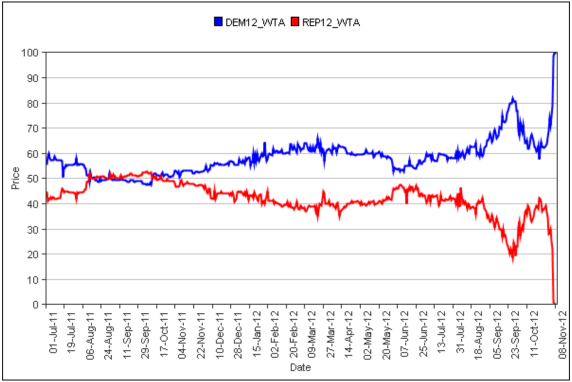
- few people answer surveys $\Rightarrow$ biased results.
- Internet polls have even more bias.

$\Rightarrow$ need to make polls attractive:

  - incentives
  - gamification
  - wide participation

Predicting Elections

- Iowa electronic market: operating since 1988.
- Mainly predicting elections in the US, with real money.
- Actual trading, no automated market maker.



Prediction Platforms

- Prediction can be a sport.

$\Rightarrow$ rewards = score $\Rightarrow$ position in leaderboard.

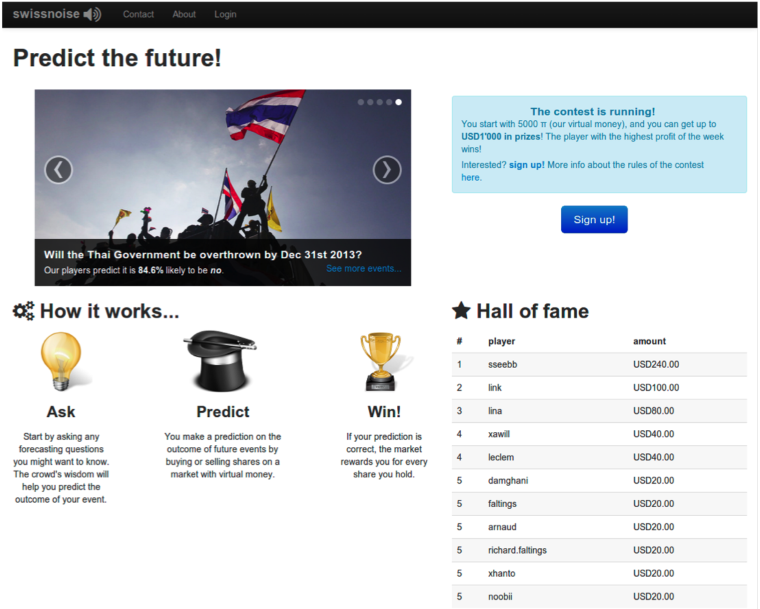
- Examples: Swissnoise, Scicast.
- Usually research projects with limited duration.

Swissnoise

- Public prediction platform operated at EPFL from spring 2013 to summer 2015.
- Users can suggest questions to put up on the platform.
- Initially run as prediction market (with logarithmic scoring rule market maker), later added PTS as alternative.
- Users randomly distributed among prediction market/PTS to compare accuracy and behavior.

F. Garcin and B. Faltings: *Swissnoise: Online Polls with Game-Theoretic Incentives*. Proceedings of the 26th Conference on Innovative Applications of AI, 2972-2977, 2014

Swissnoise


swissnoise
Contact
About
Login
Predict the future!
Will the Thai Government be overthrown by Dec 31st 2013?
Our players predict it is 84.6% likely to be no.
See more events...
The contest is running!
You start with 5000 π (our virtual money), and you can get up to USD1'000 in prizes! The player with the highest profit of the week wins!
Interested? sign up! More info about the rules of the contest here.
Sign up!
How it works...
Ask
Start by asking any forecasting questions you might want to know. The crowd's wisdom will help you predict the outcome of your event.
Predict
You make a prediction on the outcome of future events by buying or selling shares on a market with virtual money.
Win!
If your prediction is correct, the market rewards you for every share you hold.
Hall of fame
player amount
1 sseebb USD240.00
2 link USD100.00
3 lina USD80.00
4 xawill USD40.00
4 leclem USD40.00
5 damghani USD20.00
5 faltings USD20.00
5 arnaud USD20.00
5 richard.faltings USD20.00
5 xhanto USD20.00
5 noobii USD20.00

Swissnoise Statistics

- 250+ users
- avg. 15 unique users/day
- 230+ questions
- 19'700+ operations
- CHF 20 gift card to user with highest profit of the week.

Example Questions

swissnoise | My Portfolio | Events | Leaderboard | Help! | Contact | About | Hello, fgarcin ! logout?

Who will win the 2014 FIFA world cup?

Started on	Jul 7, 2014 13:35:00
Expires on	Jul 13, 2014 00:00:00
Submitted by	admin

- › Information
- › Trade
- › Tweets

› back to list of events

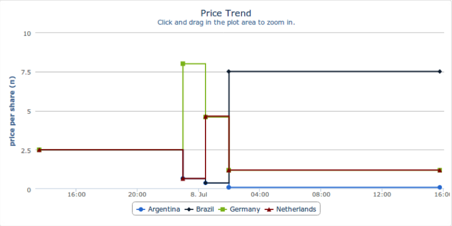
Background

The **2014 FIFA World Cup** will be the 20th FIFA World Cup, an international men's football tournament, that is scheduled to take place in Brazil from 12 June to 13 July 2014. The national teams of 32 countries will take part in the finals tournament.

4 teams are still in the competition:
- Argentina
- Brazil
- Germany
- Netherlands

Start trading !

Trend



Placing a Bet

swissnoise | My Portfolio | Events | Leaderboard | Help! | Contact | About | Hello, fgarcin ! logout?

Who will win the 2014 FIFA world cup?

Started on	Jul 7, 2014 13:35:00
Expires on	Jul 13, 2014 00:00:00
Submitted by	admin

- Information
- Trade
- Tweets

- back to list of events

Buy outcomes

Buy shares of the outcome you think will occur. Win π10.0 per share if the outcome happens.

Argentina
1.0% likely
π0.10 per share

Brazil
75.1% likely
π7.51 per share

Germany
11.9% likely
π1.19 per share

Netherlands
12.0% likely
π1.20 per share

Sell outcomes

Germany
249 shares
current value: π115.76
current profit: π-1210.07

© swissnoise 2013
Terms, Conditions and Privacy

Understanding the Market Maker

- Market maker computes price changes.
- Slider lets user see how many shares are obtained.

Explaining Peer Prediction

Prediction Market

trading stocks

Peer Prediction

lottery tickets

Implementing the Lottery

- Lottery compares ticket with randomly chosen peer ticket.
- Reward paid of both agree.
- Lottery run once per day at 3am with all unprocessed tickets.
- After lottery, predictions are integrated.
- Users can buy tickets for multiple consecutive days.

Betting with Lottery

swissnoise | My Portfolio | Events | Leaderboard | Help! | Contact | About | Hello, fgarcin ! logout?

Who will win the 2014 FIFA world cup?

Started on	Jul 7, 2014 13:33:47
Expires on	Jul 13, 2014 00:00:00
Submitted by	admin

- Information
- Trade
- Tweets

- back to list of events

This event allows you to earn up to π200.0 per day to trade on market events.
You can vote on an outcome by buying lottery tickets. Every day at 3am, we randomly choose another player's lottery ticket and if your opinions match, you are rewarded. The reward is inversely proportional to the current likelyhood of an outcome. Surprising answers are less likely to be matched but bring a higher reward.

Current Tickets

4

Current Vote

Select the outcome you think will occur. ***You can update your vote at any time!***

Argentina	Brazil	Germany	Netherlands
8.3% likely π200.00 if matched	30.6% likely π114.69 if matched	52.8% likely π78.07 if matched (selected)	8.3% likely π200.00 if matched

Risk Aversion

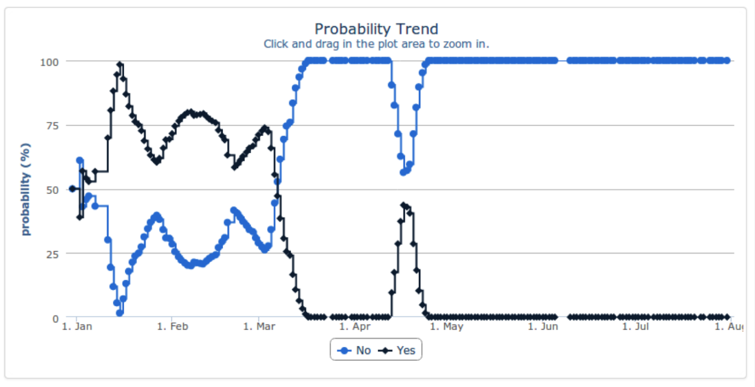
- Risk aversion: unlikely answers need to carry a higher reward.
- Scale according to Weber-Fechner's law: sensitivity to variations should be scale-invariant.

$\Rightarrow$ would mean increasing rewards *exponentially* with $1/R[x]$.

- This proved too extreme: users gambled on least likely values.
- Solved by averaging between $1/R$ and $e^{1/R}$.

Will Scotland be independent (Pred. Market)

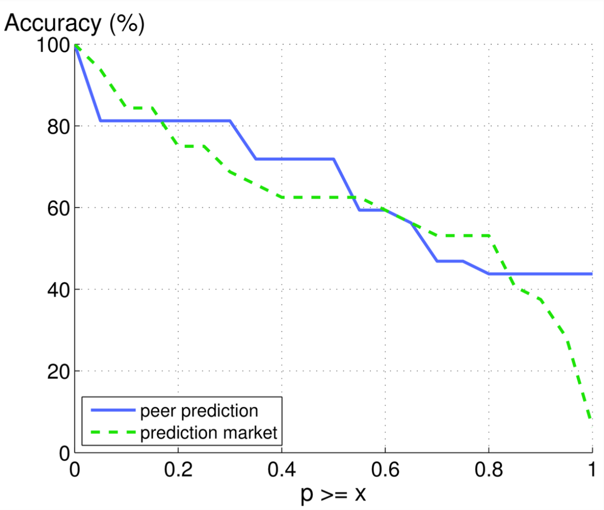
Will Scotland be independent (Peer Prediction)



Observations

- Problem with prediction market: cashing rewards changes predictions.
- Very tricky to adjust liquidity parameter (how much price influenced by share demand).
- PTS is more stable: no perturbation from participants who cash in their winnings.
- PTS more "fun": rewards can be gained every day.
- Accuracy is about the same.
- But Peer Prediction does not require ground truth ⇒ more widely applicable.

Accuracy Comparison



- Both schemes have very similar accuracy.

Human Computation

- Outsource tasks to human computation platforms such as Amazon Mechnical Turk.
- Workers are tempted to cut corners and guess answers without solving the task.
- Many tasks have similar answers $\Rightarrow$ worker errors have bias.
- Bias is impossible to correct by increasing number of workers.
- Can we do better by incentivizing workers?

B. Faltings et al.: *Incentives to Counter Bias in Human Computation*
Proceedings of HCOMP 2014, pages 59-66. AAAI, 2014

Countering Bias

- Assume bias is known as prior answer distribution R.

$\Rightarrow$ derive payment using PTS (1/R) principle.

- Example task: counting binoculars in an image.

- Count imaging devices.
- Correct answer = 34.
- Priming to 34 and 60.
- No bonus, vague bonus, peer consistency and PTS.

Influence of Priming

Bonus	Priming	Average Error	t-test
no bonus	none	1.0667	
	60	5.6316	$p = 0.0266$
	34	2.9434	$p = 0.2092$
vague	none	2.2500	
	60	6.6563	$p = 0.0810$
	34	9.0984	$p = 0.0032$
peer conf.	none	0.3492	
	60	3.3429	$p = 0.0554$
	34	2.4194	$p = 0.1496$
peer conf.	none	0.3492	
PTS	60	0.8000	$p = 0.4036$
	34	2.1667	$p = 0.2145$

Comparison of Bonus Schemes

Bonus Scheme	Priming	Average Error	t-test
none	60	5.6316	
vague	60	6.6563	$p = 0.3782$
peer conf.	60	3.3429	$p = 0.1306$
PTS	60	0.8000	$p = 0.0088$
none	34	2.9434	
vague	34	9.0984	$p = 0.0110$
peer conf.	34	2.4194	$p = 0.4020$
PTS	34	2.1667	$p = 0.3731$

- PTS corrects bias.
- Little influence if priming is to correct value.

Peer Grading: AI course quizzes

- Each student grades solutions of 5 others.
- 2 types of questions:
  - fill in a piece of code.
  - correct incorrect code.
- Answers categorized into 4 classes (correct and 3 different kinds of wrong answers).
- Incentives = bonus points for the course.
- Ground truth = expert grader (Ph.D. student).

G.Radanovic et al.: *Incentives for Effort in Crowdsourcing using the Peer Truth Serum*, ACM TIST, 2016

Empirical Performance

Error rates (measured against expert grader):

Mechanism	Num. student	Error rate (%)
PTSC	16	6.88
peer consistency	16	10.48
constant	14	11.98

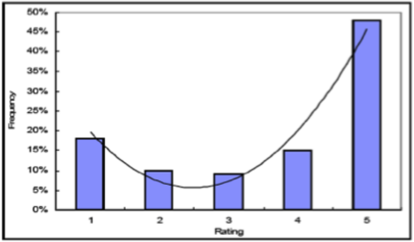
Decrease is significant (p-values):

Mechanism	PTSC	peer consistency	constant
PTSC	-	0.0255	0.0497
peer consistency	0.0255	-	0.5566
constant	0.0497	0.5566	-

Product Reviews

- Writing reviews is work $\Rightarrow$ needs a reason:
- Being extremely satisfied (or paid by the seller).
- Being extremely dissatisfied (or paid by competitor).
- Distribution of reviews can have a "U" shape.

Figure 3: Distribution of the Ratings on Amazon.com (fitted with a U-shaped curve) for a Music CD (Mr. A-Z)



Paying for Reviews

- Solution: Reviewers need to be paid for their data!
- Best criterion: reviews should be *suprising* and *confirmed*.

⇒ Peer Truth Serum is a good scheme.

- So far no practical experiment...

Applications to be tried

- Rewarding reviews: difficult to evaluate since ground truth is not known.
- Applying PTSC to crowdwork: can no longer do experiments on AMT.
- Actual crowdsensing implementation: better sensor platform in development, hope we can do this in 2017.