Truth-tracking via Approval Votng: Size Matters

Meeting in Deauville

Tahar Allouche, Jérôme Lang, Florian Yger

November 29, 2021

LAMSADE - PSL University

Table of contents

Introduction

Epistemic Social Choice - Truth Tracking

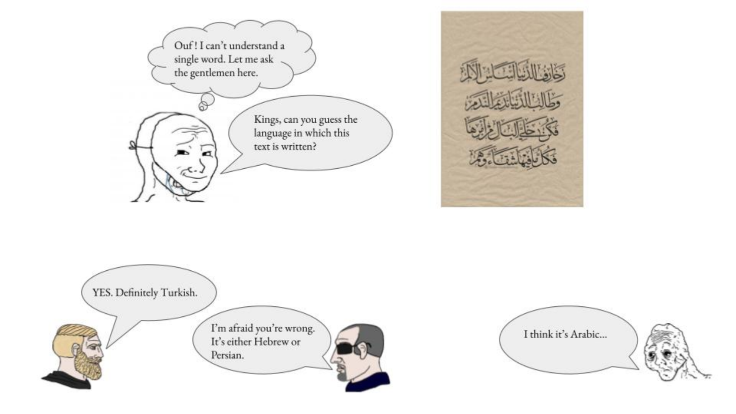


Figure 1: Real World Application of Epistemic Social Choice

Formally, consider:

- A set of m alternatives X.

Formally, consider:

- A set of m alternatives X.
- A set of n voters N.

Formally, consider:

- A set of m alternatives X.
- A set of n voters N.
- A hidden (unknown) **ground truth** alternative $a^* \in X$.

Formally, consider:

- A set of m alternatives X.
- A set of n voters N.
- A hidden (unknown) **ground truth** alternative $a^* \in X$.
- A profile of n approval ballots $A_i \subseteq X$.

Formally, consider:

- A set of m alternatives X.
- A set of n voters N.
- A hidden (unknown) **ground truth** alternative $a^* \in X$.
- A profile of n approval ballots $A_i \subseteq X$.

(+) Noise model: probability distribution over the set of possible approval ballots.

Formally, consider:

- A set of m alternatives X.
- A set of n voters N.
- A hidden (unknown) **ground truth** alternative $a^* \in X$.
- A profile of n approval ballots $A_i \subseteq X$.

(+) Noise model: probability distribution over the set of possible approval ballots.

$\implies$: Estimate the ground truth given the approval ballots by Maximum Likelihood Estimation.

Truth-tracking and Ballots' Sizes

When voters answer truthfully:

When voters answer truthfully:

- A voter who knows the correct answer would select a single alternative.

Main Idea

When voters answer truthfully:

- A voter who knows the correct answer would select a single alternative.
- A voter who selects all the alternatives has no idea of the correct answer.

When voters answer truthfully:

- A voter who knows the correct answer would select a single alternative.
- A voter who selects all the alternatives has no idea of the correct answer.

$\Longrightarrow$: Smaller ballots are of a better quality (are more accurate).

$\Longrightarrow$: Assign more weight to smaller ballots: *size-decreasing* rules.

Approval Mallows Noise

We consider the family of noises where:

Approval Mallows Noise

We consider the family of noises where:

- Voters' ballots are independent.

We consider the family of noises where:

- Voters' ballots are independent.
- There exist n parameters $\phi_i \in]0, +\infty[$ (reliability) and a function $d : X \times \mathcal{P}(X) \mapsto \mathbb{R}$ (distance to the ground truth) such that:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}, \forall a \in X$$

We consider the family of noises where:

- Voters' ballots are independent.
- There exist n parameters $\phi_i \in]0, +\infty[$ (reliability) and a function $d : X \times \mathcal{P}(X) \mapsto \mathbb{R}$ (distance to the ground truth) such that:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i} {\phi_i}^{d(a^*,A_i)}, \forall a \in X$$

We seek to estimate the ground truth by *Maximum Likelihood Estimation*:

$$\hat{a} = \arg\max_{a \in X} P_d(A_1, \ldots, A_n | a^* = a)$$

Given the noise expression:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}, \forall a \in X$$

Approval Mallows Noise

Given the noise expression:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}, \forall a \in X$$

Remark

The noise is neutral (invariant by permutation of alternatives) if and only if the distance $d(a^*, A)$ *only depends on:*

Approval Mallows Noise

Given the noise expression:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}, \forall a \in X$$

Remark

The noise is neutral (invariant by permutation of alternatives) if and only if the distance $d(a^*, A)$ *only depends on:*

- $|a^* \cap A| \in \{0, 1\}$*: The belonging or not of the ground truth to the ballot.*

Approval Mallows Noise

Given the noise expression:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}, \forall a \in X$$

Remark

The noise is neutral (invariant by permutation of alternatives) if and only if the distance $d(a^*, A)$ only depends on:

- $|a^* \cap A| \in \{0, 1\}$: *The belonging or not of the ground truth to the ballot.*
- $|A| \in \{0, \ldots, m\}$: *The size of the ballot.*

Given the noise expression:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}, \forall a \in X$$

Remark

The noise is neutral (invariant by permutation of alternatives) if and only if the distance $d(a^*, A)$ only depends on:

- $|a^* \cap A| \in \{0, 1\}$: *The belonging or not of the ground truth to the ballot.*
- $|A| \in \{0, \dots, m\}$: *The size of the ballot.*

Example

Hamming distance:

$$d_H(a, A) = |\overline{a} \cap A| + |a \cap \overline{A}| = 1 + |A| - 2|a \cap A| = \psi_d(|a \cap A|, |A|)$$

Suppose that voters share the same reliability: $\phi_i = \phi \in (0, 1)$:

Suppose that voters share the same reliability: $\phi_i = \phi \in (0, 1)$:

$$P_{\phi_i,d}(A_i|a^* = a) = P_d(A_i|a^* = a) = \frac{1}{\beta_i}\phi^{d(a^*,A_i)}, \forall a \in X$$

Suppose that voters share the same reliability: $\phi_i = \phi \in (0, 1)$:

$$P_{\phi_i,d}(A_i|a^* = a) = P_d(A_i|a^* = a) = \frac{1}{\beta_i}\phi^{d(a^*,A_i)}, \forall a \in X$$

Theorem

For $n \geq 3$, the maximum likelihood estimation rule ζ_d is a size-decreasing approval rule if and only if:

$$\Delta\psi_d : j \mapsto \psi_d(0, j) - \psi_d(1, j) \textit{ is decreasing}$$

Example

Consider the Jaccard distance:

$$d_J(a, A) = \frac{|\overline{a} \cap A| + |a \cap \overline{A}|}{|\overline{a} \cap A| + |a \cap \overline{A}| + |a \cap A|}$$

Example

Consider the Jaccard distance:

$$d_J(a, A) = \frac{|\overline{a} \cap A| + |a \cap \overline{A}|}{|\overline{a} \cap A| + |a \cap \overline{A}| + |a \cap A|}$$

$$\implies d_J(a, A) = \psi_{d_J}(|a \cap A|, |A|) = 1 - \frac{|a \cap A|}{|A| - |a \cap A| + 1}$$

Example

Consider the Jaccard distance:

$$d_J(a, A) = \frac{|\overline{a} \cap A| + |a \cap \overline{A}|}{|\overline{a} \cap A| + |a \cap \overline{A}| + |a \cap A|}$$

$$\implies d_J(a, A) = \psi_{d_J}(|a \cap A|, |A|) = 1 - \frac{|a \cap A|}{|A| - |a \cap A| + 1}$$

This yields:

$$\Delta\psi_{d_J}(j) = \psi_{d_J}(0, j) - \psi_{d_J}(1, j) = 1/j$$

ζ_{d_J} is a size-decreasing approval rule with weights $w_j = 1/j$.

Example

- The Dice distance:

$$d_D(a, A) = 1 - \frac{2|a \cap A|}{|\overline{a} \cap A| + |a \cap \overline{A}| + 2|a \cap A|}$$

yields a size-decreasing approval rule with weights $w_j = {}^{2}/_{j+1}$.

Example

- The Dice distance:

$$d_D(a,A) = 1 - \frac{2|a \cap A|}{|\overline{a} \cap A| + |a \cap \overline{A}| + 2|a \cap A|}$$

  yields a size-decreasing approval rule with weights $w_j = 2/j+1$.

- Any strictly concave transformation g of the Hamming distance:

$$d_g(a,A) = g(|\overline{a} \cap A| + |a \cap \overline{A}|)$$

  yields a size approval rule with weights $w_j = g(j+1) - g(j-1)$.

Example

What about the Hamming distance itself???

Example

What about the Hamming distance itself???

$$d_H(a, A) = |\overline{a} \cap A| + |a \cap \overline{A}| \implies \psi_{d_H}(|a \cap A|, |A|) = 1 + |A| - 2|a \cap A|$$

Example

What about the Hamming distance itself???

$$d_H(a, A) = |\bar{a} \cap A| + |a \cap \bar{A}| \implies \psi_{d_H}(|a \cap A|, |A|) = 1 + |A| - 2|a \cap A|$$

This implies that $\Delta\psi_{d_H}(j) = 2$ is not decreasing !

Example

What about the Hamming distance itself???

$$d_H(a, A) = |\overline{a} \cap A| + |a \cap \overline{A}| \implies \psi_{d_H}(|a \cap A|, |A|) = 1 + |A| - 2|a \cap A|$$

This implies that $\Delta\psi_{d_H}(j) = 2$ is not decreasing !

More generally, any distance which is separable:

$$\psi_d(|a \cap A|, |A|) = f(|a \cap A|) + g(|A|)$$

does not yield a size-decreasing rule !

Now each voter has her own reliability parameter:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}$$

such that:

Heterogeneous Reliability

Now each voter has her own reliability parameter:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}$$

such that:

$$\begin{cases} \phi_i & \in (0, +\infty) \\ & \end{cases}$$

Heterogeneous Reliability

Now each voter has her own reliability parameter:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}$$

such that:

$$\begin{cases} \phi_i & \in (0, +\infty) \\ d(a, A) & = \psi_d(|a \cap A|, |A|) \quad = f(|a \cap A|) + g(|A|) \end{cases}$$

Heterogeneous Reliability

Now each voter has her own reliability parameter:

$$P_{\phi_i,d}(A_i|a^* = a) = \frac{1}{\beta_i}{\phi_i}^{d(a^*,A_i)}$$

such that:

$$\begin{cases} \phi_i & \in (0,+\infty) \\ d(a,A) & = \psi_d(|a \cap A|, |A|) \quad = f(|a \cap A|) + g(|A|) \end{cases}$$

Example

How does the Hamming distance decompose into two separable terms ?

$$d_H(a,A) = |\overline{a} \cap A| + |a \cap \overline{A}| \implies \psi_{d_H}(|a \cap A|, |A|) = \underbrace{1 + |A|}_{g(|A|)} \overbrace{-2|a \cap A|}^{f(|a \cap A|)}$$

Theorem

If for every $1 \leq t < k \leq m-1$ *we have that:*

$$g(k) - g(t) \geq \frac{k-t}{2}\left[f(1) - f(0)\right]$$

Then:

$$\frac{\partial \mathbb{E}_{\phi,d}[|A_i|]}{\partial \phi} \geq 0$$

Heterogeneous Reliability: Sufficient Condition

Theorem

If for every $1 \leq t < k \leq m-1$ *we have that:*

$$g(k) - g(t) \geq \frac{k-t}{2}\,[f(1) - f(0)]$$

Then:

$$\frac{\partial \mathbb{E}_{\phi,d}[|A_i|]}{\partial \phi} \geq 0$$

Good news: This sufficient condition holds for the Hamming distance (and for many other distances also). But we can still go further !

Heterogeneous Reliability: Condorcet Noise

The Mallows noise with Hamming is equivalent to the Condorcet noise:

$$P_{p_i}(a \in A_i | a = a^*) = P_{p_i}(a \notin A_i | a \neq a^*) = p_i = \frac{\phi_i}{1+\phi_i}, \forall a \in X$$

Heterogeneous Reliability: Condorcet Noise

The Mallows noise with Hamming is equivalent to the Condorcet noise:

$$P_{p_i}(a \in A_i | a = a^*) = P_{p_i}(a \notin A_i | a \neq a^*) = p_i = \frac{\phi_i}{1 + \phi_i}, \forall a \in X$$

Theorem

For $m \geq 2$, we have that:

$$\mathbb{E}_p[|A_i|] = (m - 1) - (m - 2)p$$

Heterogeneous Reliability: Condorcet Noise

The Mallows noise with Hamming is equivalent to the Condorcet noise:

$$P_{p_i}(a \in A_i | a = a^*) = P_{p_i}(a \notin A_i | a \neq a^*) = p_i = \frac{\phi_i}{1 + \phi_i}, \forall a \in X$$

Theorem

For $m \geq 2$, we have that:

$$\mathbb{E}_p[|A_i|] = (m-1) - (m-2)p$$

Why is this interesting?

$$p_i = \frac{m - 1 - \mathbb{E}_{p_i}[|A_i|]}{m-2}$$

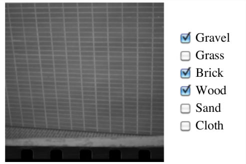
Experiments

We conducted experiments on data collected in [Shah et al., 2015]:

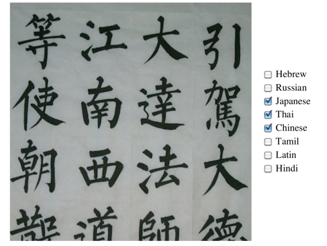
Experiments

We conducted experiments on data collected in [Shah et al., 2015]:

(a) Animals



(b) Textures



(c) Languages

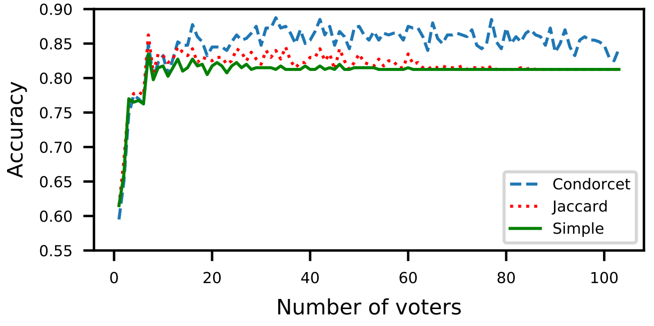


Figure 3: Guessing Animals from Images

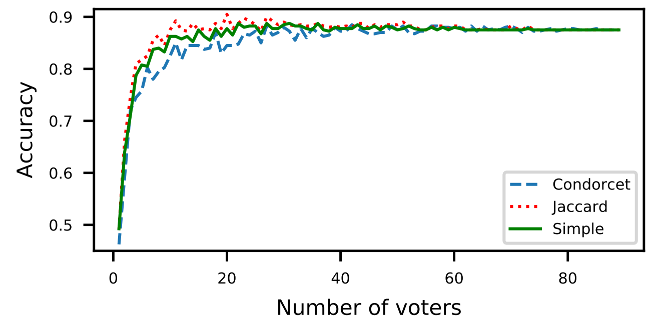


Figure 4: Guessing Textures from Images

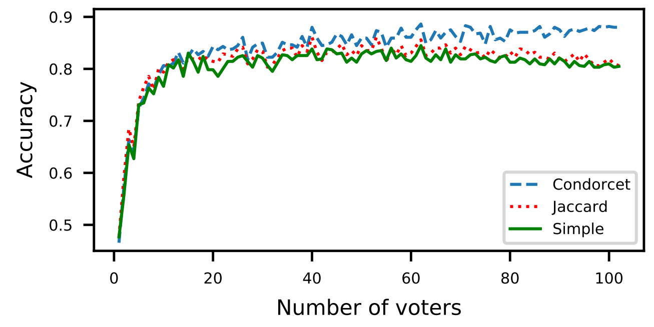


Figure 5: Guessing Languages from Images

Conclusion

Our main results are:

Main Ideas

Our main results are:

- **Homogeneous reliability:** We characterize all the functions d whose associated optimal rule is size-decreasing and we give the optimal weights.

Our main results are:

- **Homogeneous reliability:** We characterize all the functions d whose associated optimal rule is size-decreasing and we give the optimal weights.
- **Heterogeneous reliabilities:** We give a sufficient condition to d in order to have:

$$\frac{\partial \mathbb{E}_{\phi,d}[|A_i|]}{\partial \phi} \geq 0$$

Main Ideas

Our main results are:

- **Homogeneous reliability:** We characterize all the functions d whose associated optimal rule is size-decreasing and we give the optimal weights.
- **Heterogeneous reliabilities:** We give a sufficient condition to d in order to have:

$$\frac{\partial \mathbb{E}_{\phi,d}[|A_i|]}{\partial \phi} \geq 0$$

- **Condorcet noise:** In case d is the Hamming distance we give an explicit formula to estimate the reliability of a voter from her ballot's size.
- **Experiments:** We tested different size-decreasing rules on different image annotation datasets.

THANKS !

References I

Shah, N., Zhou, D., and Peres, Y. (2015).
Approval voting and incentives in crowdsourcing.
In *ICML*.