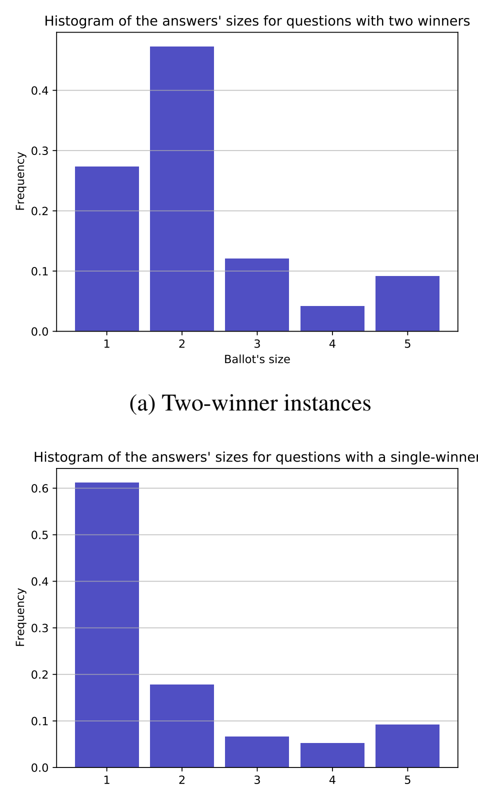
Multi-winner Approval Voting Goes Epistemic (Supplementary Material)

Tahar Allouche[1] **Jérôme Lang**[1] **Florian Yger**[1]

[1]LAMSADE, CNRS, PSL, Université Paris-Dauphine,

A DATA COLLECTION AND INCENTIVES

To see how the participants behave given the ranking incentives that we defined in the football quiz, we plotted the histogram of the sizes of the answers (see Figure 1). It appears that although the platform enables to select every alternative, only two voters did so for all the questions. Moreover, figures $1b$ and $1a$ show that the majority of the voters tend to select exactly the number of teams that appear in an image.



(a) Two-winner instances

(b) Single-winner instances

Figure 1: Histogram of the ballots' sizes

B INITIALIZING VOTERS' RELIABILITIES

Inspired by the *Anna Karenina Principle* in Meir et al. [2019], we devised an initialisation strategy for the voters' reliabilities. In his book, Leo Tolstoi stated that "Happy families are all alike; every unhappy family is unhappy in its own way". In the same spirit, it seems reasonable to make the hypothesis that accurate users tend to make similar answers, whereas inaccurate users have each their own way of being inaccurate.

Here follows an example of the Anna Karenina initialization scheme.

Example 1. *Consider following the approval profile (Table 1) for* 3 *voters,* 5 *alternatives and* 4 *Instances. Here we have*

	A^1	A^2	A^3	A^4
Voter 1	$\{a_1, a_4\}$	$\{a_1\}$	$\{a_3\}$	$\{a_1\}$
Voter 2	$\{a_2\}$	$\{a_5\}$	$\{a_4\}$	$\{a_1\}$
Voter 3	$\{a_2, a_3, a_4\}$	$\{a_2, a_3, a_5\}$	$\{a_2, a_3\}$	$\{a_3\}$

Table 1: Approval Ballots of 3 Voters on 4 Instances

that:

$$w_{max} = \frac{n}{n+1} = 0.75, w_{min} = \frac{1}{n+1} = 0.25$$

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

First, compute the mean Jaccard distance of all voters: $d_1 = 1.71, d_2 = 1.69, d_3 = 1.65$. *So* $d_{max} = d_1 = 1.71$ *and* $d_{min} = d_3 = 1.65$, *which means that voter* 3 *(the closest in average to all the voters) will get the biggest weight* $w_3 = w_{max} = 0.75$ *and voter* 1 *gets the smallest weight* $w_1 = w_{min}$. *Next, compute the weight that will be assigned to each voter, for instance:*

$$w_2 = (w_{max} - w_{min}) \frac{\frac{1}{d_2} - \frac{1}{d_{max}}}{\frac{1}{d_{min}} - \frac{1}{d_{max}}} + w_{min} = 0.38$$

Now we can set the initial values for the reliability parameters accordingly:

$$\hat{p}_2^{(0)} = \frac{1}{2}, \hat{q}_2^{(0)} = \frac{1 - \frac{e^{w_2} - 1}{e^{w_2} + 1}}{2}$$

We can check that these parameters are such that:

$$ln \left[\frac{p_2(1 - q_2)}{q_2(1 - p_2)} \right] = w_2$$

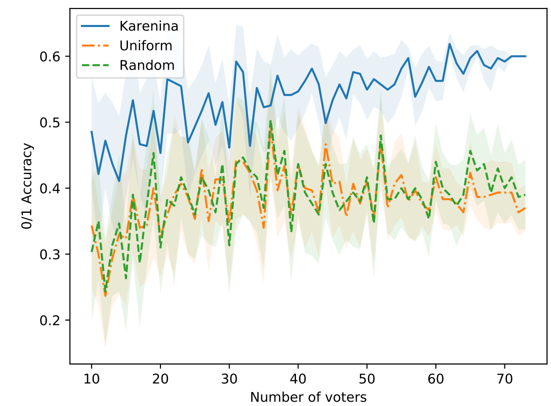
After proceeding in the same fashion with all the voters, we get the initial parameters:

$$\begin{cases} \hat{p}_1^{(0)} = 0.5 & \hat{p}_2^{(0)} = 0.5 & \hat{p}_3^{(0)} = 0.5 \\ \hat{q}_1^{(0)} = 0.44 & \hat{q}_2^{(0)} = 0.41 & \hat{q}_3^{(0)} = 0.32 \end{cases}$$

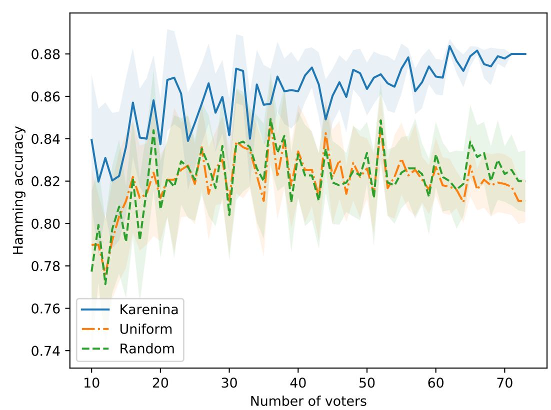
Since the AMLE only guarantees convergence to a local maximum, which makes the result depending on the initial point, we compared the results of this initialization (Anna Karenina) to other procedures to motivate its choice, see Figure 2, namely we tested:

- Uniform weights: Initially all the voters in the batch are given the same weight.
- Random weights: Initially, for each voter in the batch, p_i is randomly picked from $(0.5, 1)$ and q_i is randomly picked from $(0, 0.5)$.

We can notice that these two baseline procedures show very similar performances, and that they are both outperformed by the Anna Karenina initialization.



(a) 0-1 accuracy



(b) Hamming accuracy

Figure 2: Accuracies of different initializations

C LOSSES

C.1 HAMMING, HARMONIC AND 0-1 SUBSET METRICS

In addition to the Hamming and 0-1 subset accuracies, we introduced a new metric which can be considered as an intermediate one. The Hamming metric considers each label independently and the 0-1 subset loss considers them jointly in a strict fashion, whereas the harmonic accuracies that we introduced considers all the instance's labels jointly but with different convex weights depending on the number of

	$AMLE_c$	$AMLE_f$	Modal	Majority
Hamming	**0.88**	0.86	0.84	0.80
Harmonic	**0.78**	0.74	0.69	0.61
0/1	**0.60**	0.53	0.46	0.26

Table 2: Hamming and 0/1 accuracy for entire dataset

correctly predicted ones:

$$T(S, S^*) = \sum_{k=1}^{|S \cap S^*|} \frac{1}{6-k}$$

So out of the 5 labels:

- if 0 labels are correct then $T = 0$.
- if 1 labels is correct then $T = \frac{1}{5}$.
- if 2 labels are correct then $T = \frac{1}{5} + \frac{1}{4}$.
- if 3 labels are correct then $T = \frac{1}{5} + \frac{1}{4} + \frac{1}{3}$.
- if 4 labels are correct then $T = \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2}$.
- if 5 labels are correct then $T = \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + 1$.

Defined as such, this accuracy favours the estimators that are able to correctly estimate most of the instance's labels without being as rigid as the 0-1 subset accuracy.

This metric is reminiscent of the Proportional Approval Voting rule for multiwinner elections, which defines the score of a subset of candidates W for a voter as $1 + \frac{1}{2} + \ldots + \frac{1}{j}$, where j is the number of candidates in W approved by the voter. We could consider more generally a class of metrics defined by a vector $\vec{w}$, such that $T(S, S^*) = w_{|S \cap S^*|}$. This class generalizes Hamming, 0-1 and Harmonic and is reminiscent of the class of *Thiele* rules (see for instance Lackner and Skowron [2020] for an extended presentation of multiwinner approval-based committee rules).

C.2 RESULTS

We show in Table 2 the accuracies of the considered methods when applied to the entire annotation dataset. In Figure 3 we show the evolution of the Harmonic accuracies when the number of randomly picked voters in each batch increase.

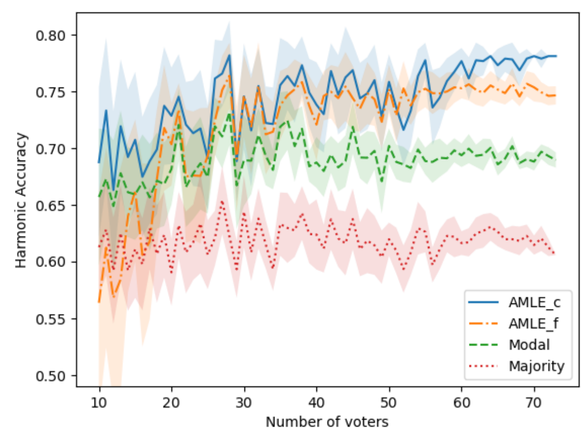


Figure 3: Normalized Harmonic accuracy

References

Martin Lackner and Piotr Skowron. Approval-based committee voting: Axioms, algorithms, and applications. *CoRR*, abs/2007.01795, 2020. URL https://arxiv.org/abs/2007.01795.

Reshef Meir, Ofra Amir, Gal Cohensius, Omer Ben-Porat, and Lirong Xia. Truth discovery via proxy voting. *arXiv:1905.00629*, 2019.