Collective Decision-Making as a Contextual Multi-Armed Bandit Problem

AXEL ABELS and TOM LENAERTS, Université Libre de Bruxelles and Vrije Universiteit Brussel, Brussels, Belgium
VITO TRIANNI, Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy
ANN NOWÉ, Vrije Universiteit Brussel, Brussels, Belgium

1. INTRODUCTION

We consider a collective decision making model in which a learner queries a set of experts for their advice about alternative solutions to a problem, which is referred to as *deciding with expert advice*. In this abstract we base the discussion on an extension of a state-of-the-art approach [Abels et al. 2020b], using both confidence and value estimates to solve contextual multi-armed bandit problems (CMAB) collectively. Although the resulting EXP4.P+CON algorithm is more robust to noisy confidence estimates than the well-known Weighted Majority Vote (WMV) [Marshall et al. 2017], it remains sub-optimal under specific conditions, requiring thus novel methods that remain effective under all conditions. We introduce the meta-CMAB approach to solving these problems, which reformulates the issue of deciding with expert advice applied to the CMAB context as a CMAB (hence meta-CMAB)[Abels et al. 2020a]. This problem transformation allows for the immediate use of established CMAB algorithms and we show that this method provides significant improvements on the state-of-the-art without the need for accurate confidence estimates.

When *deciding with expert advice* in CMAB [Li et al. 2010; Zhou 2015], a set of experts is queried for their advice about which arm to pull among K possible choices given the contextual information $\vec{x}_{k,t}$ associated with each arm k. Each expert provides advice in the form a probability distribution over the possible arms. Because the probability advice given for each arm is dependent on the assumptions made about all other arms, the alternative of value advice was introduced [Abels et al. 2020b], revealing a better performance in certain situations.

The learner's goal in the CMAB problem is no longer to approximate the problem's mapping but rather to determine the best combination of expert advice. State-of-the-art approaches based on EXP4 [Auer et al. 2002], such as EXP4.P [Beygelzimer et al. 2010], iteratively update the weights of experts to compute a weighted average of the received advice. By maintaining context-independent weights, EXP4.P implicitly makes the assumption that expert performance is homogeneous over the context space. This is a limiting assumption, as experts are likely to have expertise only for the region of the problem on which they were trained.

Querying human experts for (honest) confidence estimates about their given advice has been shown to improve performance on visual perception and value estimation tasks [Bang et al. 2017; Marshall et al. 2017]. The dominant approach in these tasks is to use confidence-weighted majority votes. If the confidence $c^n_{k,t}$ of an expert n at time t about context $\vec{x}_k$ is expressed in the range $[0, 1]$ wherein confidences of $1, 0.5$, and 0 correspond respectively to a perfect expert, a random expert, and the worst possible expert, we can follow a WMV by pulling the arm with the highest weighted sum [Marshall et al. 2017]: $\sum_{n \in N} ln(c^n_{k,t}/(1 - c^n_{k,t}))\xi^n_{k,t}$. EXP4.P+CON, introduced in [Abels et al. 2020b] extends EXP4.P to enable the exploitation of value advice and confidence estimates. With value advice, experts are

weighted in function of the cumulative squared error of their predictions as opposed to their expected cumulative reward. When confidence estimates are available, they are used as priors on the expert weights.

Now, an alternative approach to solving the problem of deciding with expert advice is to consider it as a secondary multi-armed bandit. More specifically, each expert in the set is a possible solution, and selecting that experts equates to applying its policy. The meta-MAB built this way can then be solved by standard MAB algorithms, such as Thompson Sampling [Thompson 1933]. The main drawback of this approach that decisions are always taken by following the advice of a single expert, which prevents the emergence of crowd wisdom. Here we expand this idea by considering the problem of deciding with expert advice as a CMAB, which is explained in more detail and compared to the state-of-the-art and the MAB approach in the following sections.

2. METHODS

In the meta-contextual multi-armed bandit (meta-CMAB), we make the assumption that there exists a function $\mathcal{V}$ which maps the experts' advice and confidence for a given context to an expected reward. In other words, given a context $\vec{x}_{k,t}$ (i.e., a context for a given arm k at time t) and a set of N experts wherein each expert n provides advice $\xi^n_{k,t}$ and confidence $c^n_{k,t}$ about that context, we assume the mapping function f can be approximated by $f(\vec{x}_{k,t}) \approx \mathcal{V}(\{\xi^1_{k,t}, c^1_{k,t}, ..., \xi^N_{k,t}, c^N_{k,t}\})$.

If such a $\mathcal{V}$ exists, minimizing regret in the CMAB with contexts $\vec{y}_{k,t} = \{\xi^1_{k,t}, c^1_{k,t}, ..., \xi^N_{k,t}, c^N_{k,t}\}\ \forall k \in K$ and distribution function $\mathcal{V}$ is equivalent to optimising the CDM process (i.e., minimizing regret in the deciding with expert advice setting). This formulation makes it possible to select one of the many CMAB algorithms to solve the meta-CMAB, and consequently the original problem. Just as in standard CMAB, the choice of CMAB algorithm depends on the assumptions made about $\mathcal{V}$. For example, selecting LinUCB to solve the meta-CMAB, assumes that $\mathcal{V}$ is linear, i.e., there exists some $\vec{\theta} \in \mathbb{R}^{2N}$ such that $\mathbb{E}[f(\vec{x}_{k,t})] = \mathbb{E}[\mathcal{V}(\vec{y}_{k,t})] = \vec{\theta} \cdot \vec{y}_{k,t}$.

LinUCB provides a relatively easy to interpret linear relation between expert advice and expected reward.

To allow us to exhaustively test our methods we use a pool of N artificial KernelUCB [Valko et al. 2013] experts which solve an artificial CMAB. We consider a context space of $[0, 1]^d$ with $d = 2$. The value landscape is generated following Perlin noise [Lagae et al. 2010]. Values generated in this manner have an average reward of 0.5 and range from 0 to 1. When pulling an arm with context $\vec{x}$ in this space, the reward is sampled from a binomial distribution with probability of success $p(r = 1; \vec{x}) = f(\vec{x})$, where $f : [0, 1]^d \to [0, 1]$ is the function mapping the context to its value in the value landscape. We simulate prior knowledge by introducing each expert to 100 experiences covering approximately 25% of the context space. To evaluate the effect of imperfect confidence, we introduce noise in an expert's reported confidence. Concretely, if a^n_T is expert n's true confidence, her reported confidence is sampled as follows: $c^n \sim \beta(1 + a^n_T/\eta, 1 + (1 - a^n_T)/\eta)$, with η the noise level. We use LinUCB [Chu et al. 2011] to solve the meta-CMAB and Thompson Sampling [Thompson 1933] to solve the meta-MAB.

3. RESULTS AND DISCUSSION

Due to space limitations, the performance for the two extreme cases of many arms with few experts, and, few arms with many experts are shown here. Results are given in terms of the distance between the best expert and the CDM's performance, i.e., regret in relation to the best expert. A negative regret indicates that the collective results are better than those produced by the best expert.

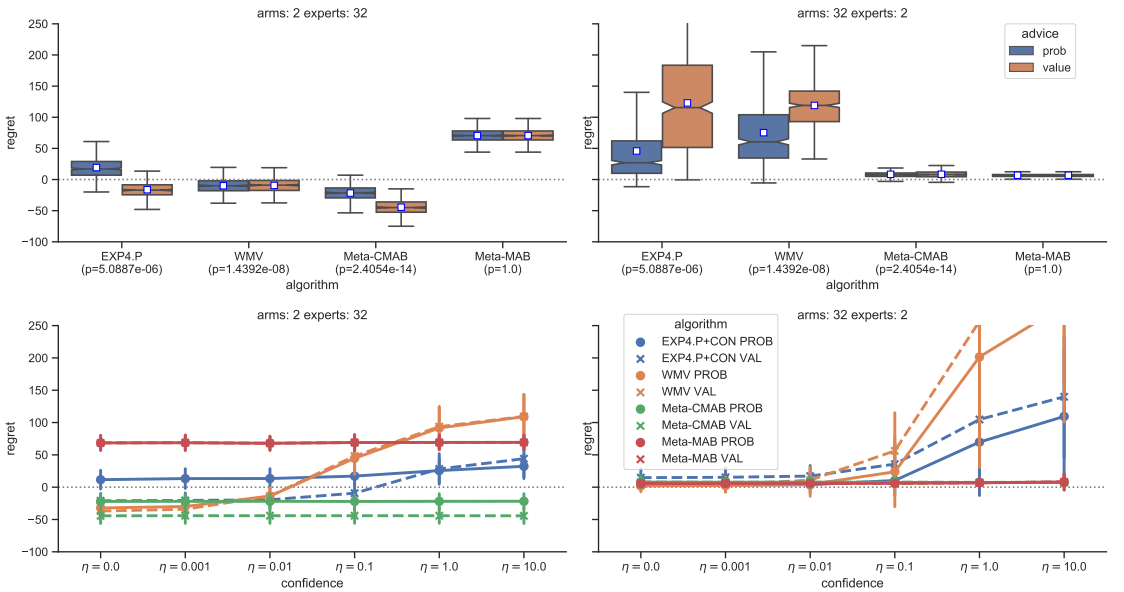

Fig. 1. Performance per advice type and confidence noise. Regret of different aggregation algorithms in function of advice and confidence noise. A value of 0 means the algorithm performs as well as the best expert. **(left) Performance without confidence grouped by advice type.** The white square marks the mean. The given p-value results from a Wilcoxon test on the results for probability and value advice. This plot presents performance when experts outnumber arms (top) and arms outnumber experts (bottom). **(right) Influence of noise on algorithm performance.** For each noise level (η) confidence is sampled from the beta distribution $\beta(1 + a/\eta, 1 + (1 - a)/\eta)$ wherein a is an expert's true confidence. A value of 0 means the algorithm performs as well as the best expert. Dashed lines use value advice, full lines use probability advice. This plot presents performance when experts outnumber arms (top) and arms outnumber experts (bottom).

When no confidence is provided (Figure 1, top) meta-CMAB significantly outperforms alternative methods when the number of experts is larger than the number of arms. Similar observations can be made for the opposite case of many arms, few experts, with the exception that meta-MAB performs similarly. These results strongly suggest that Meta-CMAB can learn an appropriate mapping from expert advice to expected outcome which it exploits to select the appropriate arms.

A first crucial observation is that even when perfect confidence estimates are provided (Figure 1, bottom, $\eta = 0$), the performance of meta-CMAB is comparable to the performance of the best performing alternative, WMV. Furthermore, as the noise in confidence estimates increases, the performance of meta-CMAB remains stable, while both EXP4.P+CON and WMV progressively degrade in performance. This strongly suggests that the WMV should only be preferred if the reliability of confidence estimates can be guaranteed and only probability advice is available. In all other cases, meta-CMAB should be preferred.

In this paper we briefly presented the drawbacks of existing approaches to deciding with expert advice and introduced meta-CMAB. Our experimental results show that this novel method provides significant improvements in performance when (i) value advice (as opposed to probability advice) is available, or (ii) confidence estimates are absent or noisy. Future work should explore whether a more purposeful integration of confidence into meta-CMAB can improve performance.

REFERENCES

Axel Abels, Tom Lenaerts, Vito Trianni, and Ann Nowé. 2020a. Collective Decision-Making as a Contextual Multi-Armed Bandit Problem. In *Submitted to the 12th International Conference on Computational Collective Intelligence (ICCCI 2020)*.

Axel Abels, Tom Lenaerts, Vito Trianni, and Ann Nowé. 2020b. How Expert Confidence Can Improve Collective Decision-Making in Contextual Multi-armed Bandit Problems. In *Submitted to the 12th International Conference on Computational Collective Intelligence (ICCCI 2020)*.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. 2002. The Nonstochastic Multiarmed Bandit Problem. *SIAM J. Comput.* 32, 1 (2002), 48–77. DOI:http://dx.doi.org/10.1137/S0097539701398375

Dan Bang, Laurence Aitchison, Rani Moran, Santiago Castanon, Banafsheh Rafiee, Ali Mahmoodi, Jennifer Lau, Peter Latham, Bahador Bahrami, and Christopher Summerfield. 2017. Confidence matching in group decision-making. *Nature Human Behaviour* 1 (05 2017). DOI:http://dx.doi.org/10.1038/s41562-017-0117

Alina Beygelzimer, John Langford, Lihong Li, Lev Reyzin, and Robert E. Schapire. 2010. An Optimal High Probability Algorithm for the Contextual Bandit Problem. *CoRR* abs/1002.4058 (2010). http://arxiv.org/abs/1002.4058

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. 2011. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 208–214.

Ares Lagae, Sylvain Lefebvre, Rob Cook, Tony DeRose, George Drettakis, David S Ebert, John P Lewis, Ken Perlin, and Matthias Zwicker. 2010. A survey of procedural noise functions. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 2579–2600.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. ACM, 661–670.

James A.R. Marshall, Gavin Brown, and Andrew N. Radford. 2017. Individual Confidence-Weighting and Group Decision-Making. *Trends in Ecology & Evolution* 32, 9 (2017), 636 – 645. DOI:http://dx.doi.org/https://doi.org/10.1016/j.tree.2017.06.004

William R Thompson. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25, 3/4 (1933), 285–294.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias N. Flaounas, and Nello Cristianini. 2013. Finite-Time Analysis of Kernelised Contextual Bandits. *CoRR* abs/1309.6869 (2013). http://arxiv.org/abs/1309.6869

Li Zhou. 2015. A Survey on Contextual Multi-armed Bandits. *CoRR* abs/1508.03326 (2015). http://arxiv.org/abs/1508.03326