Multi-Advisor Dynamic Decision Making

Doctoral Consortium

Zhaori Guo
University of Southampton
United Kingdom
zg2n19@soton.ac.uk

ABSTRACT

Being able to infer the ground truth from the answers of multiple imperfect advisors is a problem of crucial importance in many decision-making applications, such as lending, trading, investment, and crowd-sourcing. It is important to make multiple decisions over time in a sequential decision-making setting. Crucially, we assume no access to ground truth and no prior knowledge about the reliability of advisers. Specifically, our research considers how to (1) learn the trustworthiness of advisers dynamically without prior information by asking multiple advisers and (2) make optimal decisions without access to the ground truth and improve this over time. To address these problems, we proposed a new method, which combines the Bayesian Weighted Voting ensemble method and Subjective Logic. It can aggregate binary answers from multiple imperfect advisors for truth inference and model the trustworthiness of advisors. We address two problems based on our method. The first is a multi-trainer interactive reinforcement learning system; the second is multi-advisor dynamic binary decision-making by maximizing the utility. The experimental results show that our approach outperforms other state-of-the-art methods.

KEYWORDS

Bayesian inference, interactive reinforcement learning, trustworthiness

ACM Reference Format:
Zhaori Guo. 2023. Multi-Advisor Dynamic Decision Making: Doctoral Consortium. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Many situations rely on expert advice to make decisions, and often there is no objectively correct answer. Examples are wide-ranging and include crowdsourcing, machine learning ensemble models, or loan approvals. In such settings, and following the principles of the wisdom of the crowd [6, 11], it may be better to rely on the expertise of multiple advisers, especially if the stakes are high. However, it is unrealistic to assume that all people have the same level of knowledge, so we should model the quality of experts to determine their significance in the decision-making process. In addition, typically, multiple sequential decisions are made, and the reliability of individual advisers can be learned over time.

Some research involves aggregating answers to infer the ground truth [1, 2, 9–11]. They usually make decisions by dictatorship, majority voting, weighted voting, and expectation-maximization (EM) methods. However, these methods have disadvantages such as misleading, imprecise, and large computing power requirements. For example, maximum likelihood estimation methods have a large deviation between the estimated trustworthiness distribution and the real one [7] when the data set is small. This deviation can mislead future decisions and samples.

To address these challenges, we design a novel method, "Multi-Advisor Dynamic Decision-Making," for sequential, multi-advisor decision-making problems for settings with no ground truth. The method consists of two parts. The first part is the trust model, which takes care of understanding which advisors are more reliable than the others. We express the trustworthiness of the advisor through a parameter, which roughly estimates the probability that the report made by the advisor is correct. The second part is the decision model, which, given the feedback and trustworthiness of the advisors, allows the system to make decisions and provide new evidence for advisors' trustworthiness updating.

2 PROBLEM FORMALIZATION

Let D be the set of decisions, and let X be a set of advisors. For every decision $d \in D$, the decision-maker needs to choose a unique answer with a binary value, namely $a_d \in \{-1, 1\}$. For simplicity but without loss of generality, we assume that the correct value, i.e. the ground truth, denoted by a_d^*, is positive, i.e. $a_d^* = 1$. For any given $d \in D$, there is a subset of advisors $Y_d \subseteq X$. For every advisor, $x \in X$, τ_x is its trustworthiness, which is updated after every decision for which that advisor is consulted. Finally, we denote with $\vec{\tau}$ the vector containing all the advisors' trustworthiness values.

For any given $d \in D$, we denote with $P_d \subseteq Y_d \subseteq X$ the set of advisors who give positive answers to decision d. Similarly, we denote with $N_d \subseteq Y_d \subseteq X$ the set of advisors who give a negative answer to decision d. Note that $P_d \cap N_d = \emptyset$ and $P_d \cup N_d = Y_d$ for every $d \in D$.

We assume that, for any given decision, d, there exists a true answer a_d^*, but this ground truth is never revealed to the decision maker. Therefore, we use $a_d = f(P_d, N_d)$ to refer to the decision-making function of our inference model. This is a function of the responses of the advisors in P_d and N_d. If $a_d = a_d^*$, we say that the answer is correct. Otherwise, we say that the answer is wrong. Our goal is to maximize the number n of $a_d = a_d^*$ under the feedback set (P_d, N_d) by the function $f(P_d, N_d)$. It can express as:

$$f^* = \arg\max_f n(f(P_d, N_d) = a_d^*). \quad (1)$$

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

3 TRUST-INFORMED ADVICE AGGREGATION

The design of MADDM consists of three parts. The first part is a trustworthiness model that determines an advisor's trustworthiness, which can be used as weights in the decision model and to calculate the contributions of advisors in the advisor selection model. The second part is the decision model, which selects an answer after receiving the opinions of the advisors.

3.1 Trustworthiness Model

Following Jøsang [3], we build our trustworthiness model using a Beta distribution. Recall that we do not know the ground truth, and so, for every advisor, we associate two values, called *correct estimated evidence* α_x and *wrong estimated evidence* β_x. Now, for every advisor $x \in X$, we define its trustworthiness as $\tau_x = \alpha_x/(\beta_x + \alpha_x) \in (0, 1)$. Every advisor's trustworthiness τ_x is paired with a parameter θ_x, which quantifies the reliability of τ_x. We need to use θ_x to tune our decision-making method. As we acquire more evidence regarding an advisor x, this uncertainty will reduce. For every advisor $x \in X$, the uncertainty of x is $\theta_x = 2/(\alpha_x + \beta_x) \in (0, 1]$.

3.2 Bayesian and Weighted Voting Ensemble

We use the Bayesian and Weighted Voting Ensemble (BWVE) as the decision function f to make decisions. Essentially, it combines two decision procedures to improve the overall outcome. One is based on a Bayesian model, while the other follows a weighted voting decision method. If we know the real trustworthiness $\vec{\tau}$ of all the advisors, the Bayesian method will obtain higher accuracy than the weighted voting method. However, in the beginning, because the uncertainty of the trustworthiness is large, the Bayesian method is unstable, so BWVE relies more on the weighted voting method for decisions. With the decreasing of the average uncertainty, the Bayesian method has a better performance. So BWVE uses the average uncertainty to control the weights of Bayesian and weighted voting automatically.

Let $\bar{\theta}_d$ denote the average uncertainty. Let $P_d^{e+}, P_d^{b+}, P_d^{w+}$ respectively represent the probability of using BWVE, Bayesian and weighted voting methods to get $a_d^* = 1$ under the answer set (P_d, N_d), respectively. Let $P_d^{e-}, P_d^{b-}, P_d^{w-}$ respectively represent the probability of using BWVE, Bayesian and weighted voting methods to get $a_d^* = -1$. For the ensemble decision, and the given the answer set (P_d, N_d), the probability that $a_d^* = 1$ is $P_d^{e+} := P_b(a_d^* = 1|P_d, N_d)$, while $P_d^{e-} := P_b(a_d^* = -1|P_d, N_d)$is the probability that $a_d^* = -1$. They can be expressed as:

$$P_d^{e+} = (1 - \bar{\theta}_d)P_d^{b+} + \bar{\theta}_d P_d^{w+} \tag{2}$$

$$P_d^{e-} = (1 - \bar{\theta}_d)P_d^{b-} + \bar{\theta}_d P_d^{w-} \tag{3}$$

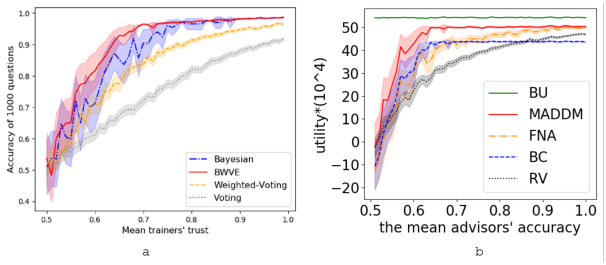
After getting P_d^{e+} and P_d^{e-}, the system needs to compare them. If $P_d^{e+} > P_d^{e-}$, the final answer $a_d = 1$. Otherwise, $a_d = -1$.

BWVE uses the absolute difference of P_d^{e+} and P_d^{e-} as the new estimated evidence to update α and β.

$$i_d = |P_e(a_d^* = 1|P_d, N_d) - P_e(a_d^* = -1|P_d, N_d)| \tag{4}$$

4 INTERACTIVE REINFORCEMENT LEARNING

In previous interactive reinforcement learning research, people often focus on the interaction between a single human trainer and an agent [4, 5, 8]. If the human teacher is not always reliable, then they will not be consistently able to guide the agent through its training. In this section, we propose a more effective interactive reinforcement learning system by introducing multiple trainers, namely Multi-Trainer Interactive Reinforcement Learning (MTIRL), which could aggregate the binary feedback from multiple non-perfect trainers into a more reliable reward for an agent training in a reward-sparse environment.



In Figure 1a, our results show that our aggregation method has the best accuracy when compared with the majority voting, the weighted voting, and the Bayesian method. Finally, we conduct a grid-world experiment to show that the policy trained by the MTIRL with the review model is closer to the optimal policy than that without a review model.

5 DECISION-MAKING BASED ON UTILITY

Being able to infer the ground truth from the answers of multiple imperfect advisors is a problem of crucial importance in many decision-making applications, such as lending, trading, investment, and crowd-sourcing. In practice, however, gathering answers from a set of advisors has a cost. Therefore, finding an advisor selection strategy that retrieves a reliable answer and maximizes the overall utility is a challenging problem. To address this problem, we propose a novel strategy based on MADDM for optimally selecting a set of advisers in a sequential binary decision-making setting, where multiple decisions need to be made over time without ground truth. Specifically, our approach considers how to select advisors by balancing the advisors' costs and the value of making the correct decisions.

In Figure 1b, the results show that our approach outperforms two other methods that combine state-of-the-art models.

6 CONCLUSION AND FUTURE WORK

In this paper, we introduce a MADDM, a novel approach for making dynamic decisions based on multiple imperfect advisors. It makes optimal decisions by multiple advisors without access to the ground truth and dynamically learns the trustworthiness of advisors without prior information. An interesting direction for future work is moving from binary answers to multiple answers, making our approach applicable to more scenarios. This requires changing the probabilities of the outputs from two to multiple.

REFERENCES

[1] Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. 2012. Zencrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In *Proceedings of the 21st international conference on World Wide Web*. 469–478.

[2] Meric Altug Gemalmaz and Ming Yin. 2021. Accounting for Confirmation Bias in Crowdsourced Label Aggregation.. In *IJCAI*. 1729–1735.

[3] Audun Jøsang. 2016. *Subjective logic*. Vol. 3. Springer.

[4] W Bradley Knox and Peter Stone. 2008. Tamer: Training an agent manually via evaluative reinforcement. In *2008 7th IEEE international conference on development and learning*. IEEE, 292–297.

[5] W Bradley Knox and Peter Stone. 2010. Combining manual feedback with subsequent MDP reward signals for reinforcement learning. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*. Citeseer, 5–12.

[6] Hélène Landemore. 2012. Collective wisdom: Old and new. *Collective wisdom: Principles and mechanisms* (2012), 1–20.

[7] Alexander Ly, Maarten Marsman, Josine Verhagen, Raoul PPP Grasman, and Eric-Jan Wagenmakers. 2017. A tutorial on Fisher information. *Journal of Mathematical Psychology* 80 (2017), 40–55.

[8] James MacGlashan, Mark K Ho, Robert Loftin, Bei Peng, Guan Wang, David L Roberts, Matthew E Taylor, and Michael L Littman. 2017. Interactive learning from policy-dependent human feedback. In *International Conference on Machine Learning*. PMLR, 2285–2294.

[9] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 627–635.

[10] Fangna Tao, Liangxiao Jiang, and Chaoqun Li. 2021. Differential evolution-based weighted soft majority voting for crowdsourcing. *Engineering Applications of Artificial Intelligence* 106 (2021), 104474.

[11] Yudian Zheng, Guoliang Li, Yuanbing Li, Caihua Shan, and Reynold Cheng. 2017. Truth inference in crowdsourcing: Is the problem solved? *Proceedings of the VLDB Endowment* 10, 5 (2017), 541–552.