Budget Limited Trust-Aware Decision Making

Taha D. Güneş, Timothy J. Norman(✉), and Long Tran-Thanh

Agents, Interaction and Complexity Group,
University of Southampton, Southampton, UK
{t.d.gunes,t.j.norman,l.tran-thanh}@soton.ac.uk

Abstract. Utilizing witness information to supplement direct evidence is commonly used to build assessments of the trustworthiness of agents. The process of acquiring this kind of evidence is, however, typically assumed to be cost-free. In practice, agents are budget-limited, and investments in acquiring witness (or reputation) information will affect the budget that can be used for direct interaction. At the same time, acquiring such witness information can help in making better trust decisions. We explore this trade-off, formalising it as a budget-limited multi-armed bandit problem, and evaluate the effectiveness of algorithms to guide this decision process.

1 Introduction

Models of trust in agent societies are designed to support decisions of who to interact with. To better choose interaction partners, historical information about their past performance is necessary for most trust models [1–4]. Accessing interaction histories is not always feasible, however, and may be costly. Agents that query reputation information providers to reduce the uncertainty associated with limited knowledge will incur costs, at least in terms of time to decision: evaluating the trustworthiness of others is resource-dependent [5]. The question then is how to take into account information retrieval costs, resource limits and the properties of an agent society to guide the process of deciding whom to interact with.

Numerous models have been proposed to effectively discover trustworthy partners that do not consider resource constraints. In a recent and insightful review, Yu *et al.* [1] characterise these approaches as *greedy* and *dynamic*. The most common greedy approach is, in general, to progressively pick the best option that the agent has. Agents start by exploring trustees randomly, gradually shifting towards those that have higher reputation. Dynamic approaches tend to divide effort between *exploration* and *exploitation*. According to Yu *et al.*, there are few dynamic approaches except for those that use reinforcement learning. A recent example is the model proposed by Sen *et al.* [6], in which they consider a supply chain and employ a Budget-Limited Multi-Armed Bandit BL-MAB algorithm [7] to manage the explore/exploit trade-off. This is one of the first approaches that considers cost associated with invoking the services of trustees. They do not, however, consider the process of acquiring witness information, or

G. Sukthankar and J. A. Rodriguez-Aguilar (Eds.): AAMAS 2017 Visionary Papers,
LNAI 10643, pp. 101–110, 2017.
https://doi.org/10.1007/978-3-319-71679-4_7

the costs associated with this. In other recent research, strategies for acquiring witness information within cost constraints have been explored that are robust to biases in reports due to effects such as hearsay evidence [8]. This research focusses exclusively on the trust assessment and information fusion problem, however, eschewing the question of deciding who to trust.

There are several different metrics used to judge an agent as trustworthy that have been considered in the literature. In this work, we consider an agent to be *trustworthy* if it acts according to the truster's expectation most of the time; e.g. by consistently providing a satisfactory service. Reputational reports from third parties can aid in the generation of these expectations, but may be misleading for decision makers if, for example, the witness is not reliable [5]. To limit the complexity of the problem we address, however, we assume that witness information is unbiased.

Our starting point in this research is algorithms developed to solve budget-limited multi-armed bandit (BL-MAB) problems. We explore algorithms that combine direct and indirect evidence of agent performance and evaluate trustworthiness on-the-fly within cost and budget constraints; the aim being to maximise the number of successful interactions. Our assumed setting is as follows. The decision maker has an infinite number of tasks that can only be completed through out-sourcing to service providers, but it has a fixed budget for the completion of tasks. Each service provider handles a task with some fixed cost. Ratings of prior performance of these service providers can be purchased from a central authority. Given these constraints, the decision maker's goal is to have as many tasks as possible completed within budget. Therefore, the decision maker must spend its budget strategically to identify and utilise high-performing service providers.

The rest of the paper is organised as: First, we formulate this decision-making problem in Sect. 2. Then the algorithms that we developed are proposed in Sect. 3, later in Sect. 4 we show our findings and in Sect. 5 we discuss them in detail. Lastly, we conclude our investigation in Sect. 6.

2 Budget-Limited Trust Decision Making

Given that our focus is on the problem of selecting good (trustworthy) agents within hard budgetary constraints, we intentionally use a simple model of evidence and trust assessment. We also formalise the model from the perspective of a single agent (the decision maker) making service selection decisions. The environment in which this agent operates consists of a set of agents, $\mathcal{A} = \{1, \ldots, n\}$, that offer functionally equivalent services, but that vary in performance. We assume that the performance of a service provider can be judged as success/failure once invoked by the decision maker. Given binary performance assessments, a common means to build a model to predict future performance (i.e. trustworthiness) is through Subjective Logic (SL) [9]; this is the trust assessment model we adopt.

In SL, an opinion held by some decision maker, i, about an agent, j, regarding some issue is a tuple $\omega_{i:j} = \langle b_{i:j}, d_{i:j}, u_{i:j}, a_{i:j} \rangle$, where $b_{i:j}$ is the belief mass

associated with i's view that j will succeed in future, comparable interactions (aka. *belief*), $d_{i:j}$ is that associated with future failure (aka. *disbelief*), $u_{i:j}$ is the belief mass associated with i's *uncertainty* where $u_{i:j} = 1 - (b_{i:j} + d_{i:j})$, and $a_{i:j} \in [0, 1]$ is the prior, or base rate. The evidence used to construct binomial opinions are represented as a pair $\langle r_{i:j}, s_{i:j} \rangle$ where $r_{i:j}$ is the number of *positive* interactions that i experienced with j and $s_{i:j}$ is the number of *negative* interactions. The belief masses, $b_{i:j}$, $d_{i:j}$ and $u_{i:j}$, are computed using the formulae:

$$b_{i:j} = \frac{r_{i:j}}{(r_{i:j} + s_{i:j} + 2)} \tag{1}$$

$$d_{i:j} = \frac{s_{i:j}}{(r_{i:j} + s_{i:j} + 2)} \tag{2}$$

$$u_{i:j} = \frac{2}{(r_{i:j} + s_{i:j} + 2)} \tag{3}$$

We can generate a single-valued, normalised trust assessment that can be used to rank and select from among individuals by distributing the uncertainty between belief and disbelief via our base rate, thus:

$$\tau_{i:j} = b_{i:j} + a_{i:j} \cdot u_{i:j} \tag{4}$$

Given that we consider the trust decision problem from the perspective of a single agent, we typically refer to τ_j as the trust that our decision maker has in agent $j \in \mathcal{A}$, that r_j is the number of positive experiences our decision maker has with j, etc. The exception is when we refer to an agent we call the *oracle*, $\mathcal{O}$.

As a proxy for querying for reputation reports from witnesses to the performance of agents in $\mathcal{A}$, we use a single reputation provider: the oracle. The oracle has some amount of evidence about each agent in the environment $\{\langle r_{\mathcal{O}:1}, s_{\mathcal{O}:1} \rangle, \ldots, \langle r_{\mathcal{O}:n}, s_{\mathcal{O}:n} \rangle\}$. The certainty of the opinions held by $\mathcal{O}$ is parameterized by K; i.e. for each $\langle r_{\mathcal{O}:j}, s_{\mathcal{O}:j} \rangle \in \mathcal{O}$, $r_{\mathcal{O}:j} + s_{\mathcal{O}:j} = K$. This is, of course, a significant simplification of the process of acquiring evidence from witnesses. Normally, it would be necessary for the decision maker to build a model of each other agent *as a witness* (a different issue from that of being a service provider), then use these in order to discount opinions from different sources. This would, however, introduce unnecessary complexity to our model; we argue that this simplification enables us to focus on our central question of budget-limited trust decision making.

We formalize our decision problem as a budget-limited multi-armed bandit [7], which aims to maximize the total amount of reward within a budget by pulling arms of a slot machine. Pulling an arm is a metaphor for interacting with either a reputation provider or invoking the service of some provider (trustee). The objective is to maximize the total number of successful interactions that truster agent makes with trustees given the available budget. The truster can request information about trustees from the *Oracle* or interact with agents directly. Information from *Oracle* supports future decisions only: it provides no reward. The truster, therefore, needs to decide how to invest its budget: querying the *Oracle* or directly interacting with trustees.

Suppose that the cost of querying the *Oracle* is d, the cost of interacting with a trustee directly is c, and the agent has a budget, B. Given some algorithm A, the number of direct interactions N_i^B and witness information retrievals $N_{\mathcal{O}}^B$ are bounded by the budget B; that is:

$$P\left(\sum_{i}^{n} N_i^B(A) \cdot c + \sum_{i}^{n} N_{\mathcal{O}}^B(A) \cdot d \leq B\right) = 1. \tag{5}$$

The optimal algorithm, A^*, is an algorithm that maximizes total reward (total number of successful interactions), such that:

$$A^* = \arg\max_{A} \sum_{i}^{n} \mathbf{E}[N_i^B(A)] \cdot \mu_i - \sum_{i}^{n} \mathbf{E}[N_{\mathcal{O}}^B(A)] \tag{6}$$

3 The Algorithms

In this section, we formalise the algorithms that we investigated for this particular problem: the first (A_{greedy}) randomly picks trustees and tends to stick with honest agents, two other algorithms ($A_{\epsilon_{1,2}}$) are allocating budget for witness information to bootstrap their knowledge about the environment. All of the algorithms that are described below comply the restriction of not overspending the fixed budget (Eq. (5)). The normalised trust assessment calculation shown in Eq. (4) is used in each algorithm to calculate the density of reward.

A_{greedy}: The greedy algorithm is a popular approach for trust-aware decision making [1]. The version that we implemented is an extension of random exploration. Initially, normalised trust assessments of all agents are equal. For this reason, the first interaction that algorithm performs is to randomly pick a trustee agent. Based on the outcome of the first interaction, future iterations of A_{greedy} may be directed to explore other agents or stick with the same one. These selections are determined by picking the most dense arm which is $i = \text{argmax}_i(\tau_i)$, as in BL-MAB epsilon-first approaches [7]. The Oracle's opinions are not queried in this algorithm. We consider this algorithm as a baseline for other algorithms and formalised in Algorithm 1.

A_{ϵ_1}: The ϵ-first algorithm (shown in Algorithm 2) allocates its budget based on a ratio of exploration/exploitation, ϵ, where the exploration budget is ϵB and the remainder of the budget $B - \epsilon B$ is reserved for exploitation. In exploration as long as exploration budget is not exhausted, an agent is selected randomly and the reputation information about that agent is gathered from the *Oracle*. (The same agent is not queried twice.) The cost of witness information retrievals, d, is deducted from the exploration budget for each transaction. Depending on the exploration budget and the number of agents in the environment, there may be some budget left; if so, this is added to exploitation budget. The exploitation phase is then bounded by the remaining budget, where the cost of each interaction is c. This phase is identical to A_{greedy}, where the most dense arm is pulled and the density of this arm may change as a result.

A_{ϵ_2}**:** This algorithm (shown in the Algorithm 3) differs from A_{ϵ_1} in the exploitation phase only. Rather than looking for the densest arm, it randomly samples arms according to their density. This may lead to more information about other trustees being acquired, increasing the chance of exploring more of the population.

Algorithm 1. Trust-Aware Budget-Limited Greedy Algorithm - A_{greedy}

$t \leftarrow 1$;
Exploration phase:
Exploitation phase:
while $B_t \geq c$ **do**
    $i = \arg\max_{\mathrm{i}}(\tau_{\mathrm{i}})$;
    interact with i and update $\langle r_i, s_i \rangle$;
    $B_{t+1} \leftarrow B_t - c$;
    $t \leftarrow t + 1$;
end while

Algorithm 2. Deterministic Trust-Aware Budget-Limited ϵ-First Algorithm - A_{ϵ_1}

$t \leftarrow 1$;
$B^{explore} \leftarrow \epsilon B$;
$B^{exploit} \leftarrow B - B^{explore}$
Exploration phase:
$A = \mathcal{A}$
while $B_t^{explore} \geq d$ **and** $A \neq \{\}$ **do**
    randomly select $i \in A$
    $\langle r_{\mathcal{O}:j}, s_{\mathcal{O}:j} \rangle \leftarrow$ query $(\mathcal{O}, i)$
    $r_i \leftarrow r_i + r_{\mathcal{O}:j}$ $s_i \leftarrow s_i + s_{\mathcal{O}:j}$;
    $B_{t+1}^{explore} \leftarrow B_t^{explore} - d$;
    $A \leftarrow A \setminus \{i\}$;
    $t \leftarrow t + 1$;
end while
$B^{exploit} \leftarrow B^{exploit} + B^{explore}$;
Exploitation phase:
while $B_t^{exploit} \geq c$ **do**
    $i = \arg\max_{\mathrm{i}}(\tau_{\mathrm{i}})$;
    interact with i and update $\langle r_i, s_i \rangle$;
    $B_t^{exploit} = B_t^{exploit} - c$;
    $t \leftarrow t + 1$;
end while

4 Simulation Results

To evaluate our algorithms, we conducted experiments to investigate: the advantages and disadvantages of investing budget in acquiring witness information; choosing reputation versus direct experience in varying budget scenarios;

Algorithm 3. Trust-Aware Budget-Limited ϵ-First Algorithm - A_{ϵ_2}

Exploration phase:
Same as A_{ϵ_1}
Exploitation phase:
while $B_t^{exploit} \geq c$ **do**
    $i \leftarrow$ weighted random sample set $\{\tau_i, ..., \tau_n\}$;
    interact with i and update $\langle r_i, s_i \rangle$;
    $B_{t+1}^{exploit} \leftarrow B_t^{exploit} - c$;
    $t \leftarrow t + 1$;
end while

Table 1. Simulation environment

Description	Parameter	Value
Budget	B	300
Oracle knowledge	K	100
Direct interaction cost	c	3
Witness information cost	d	1
Total number of trustees	N	160

the knowledge acquired by each algorithm; and the factors that affect an optimal ϵ. Each experiment was repeated 1000 times and the average taken to minimise influence of noise. The parameters in Table 1 are selected for our experiments. We defined the behaviours of the agent as honest, random, malicious with numbers of 10, 50 and 100. Behaviours are distributed normally such that the mean of an honest agent is selected randomly from range $[0.5, 1.0]$ and for dishonest agents $[0, 0.5]$ with standard deviation 0.1. The amount of evidence from the Oracle is distributed normally with a mean 100 and a standard deviation 20.

4.1 Optimal ϵ

Our results indicate that investing some budget in acquiring witness information can yield an increase in reward. In Fig. 1a, the ϵ-first algorithm A_{ϵ_1} performed better than other algorithms for some values of ϵ for a budget of 300; Here A_{ϵ_1} gains the maximum reward with $\epsilon = 0.1$. The total reward is, however, sensitive to the choice of ϵ.

We then investigated whether the choice of a good ϵ, depends on the budget, B. As shown in Fig. 1b, we varied the budget up to 600 to explore how this affects the optimal ϵ. We found no clear dependency between budget and ϵ: a peak reward is obtained near to $\epsilon = 0.1$ in Fig. 1a, regardless of budget. We conclude that ϵ does not depend on available budget.

4.2 Environment Exploration

Exploration of the environment varied significantly in each algorithm. Since the budget is limited, all algorithms had to interact with a certain number of agents.

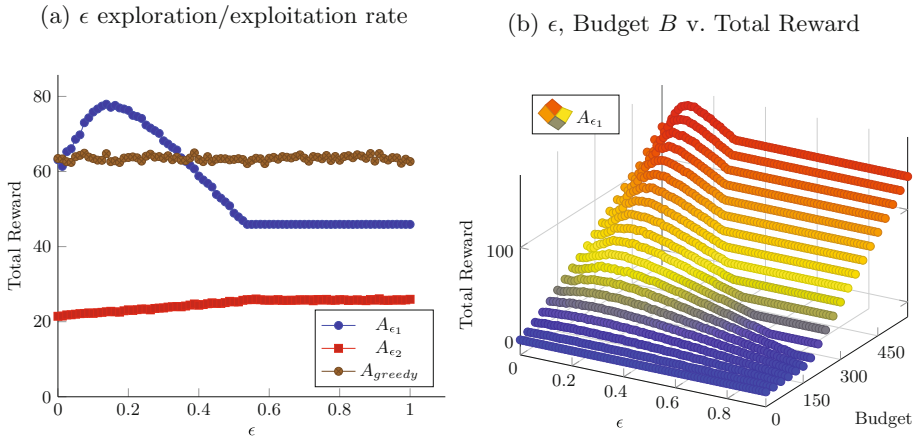

Fig. 1. Exploration vs. Exploitation comparison

As shown in Fig. 3[1], A_{ϵ_2} used its budget for more exploration than A_{ϵ_1} which spent more of its budget on exploitation, and hence acquired more evidence about specific service providers. Algorithm A_{greedy} followed a similar pattern as $A_{\epsilon_{1,2}}$ with the exception of the peak around 95. The reason for these peaks in the amount of evidence acquired about individual service providers is that both $A_{\epsilon_{1,2}}$ query the Oracle. The evidence that A_{greedy} acquires varies from 0 to 100 in an decreasing manner.

4.3 Performance over Time

We investigated the probability that an interaction is successful over time. In Fig. 2b, A_{greedy} became more successful over time as it starts to identify better performing service providers from a random initial selection; this drops to zero at the end simply because A_{greedy} has exhausted its budget. The other algorithms, $A_{\epsilon_{1,2}}$, invest budget at the start of the simulation on exploration (querying the Oracle), and hence receive no reward. During the exploitation phase, however, the probability of a successful interaction was relatively static for both ϵ-first algorithms.

The total reward acquired by A_{ϵ_1} was higher than our benchmark reference A_{greedy}, as shown in Fig. 2a, and this was consistently the case regardless of budget. On the other hand, the performance of A_{ϵ_2} was significantly worse than either of the other algorithms.

The formulation of our problem is the reason for the delayed-reward effect: all, zero reward, interactions with the Oracle occur before any exploitation of the knowledge acquired. This provides a reasonable outcome if the environment

[1] The maximum frequency in the figure is capped at 5 for clarity of presentation; the number of agents for which the decision maker has no evidence is often significantly higher than 5.

is static throughout; i.e. the availability of service providers does not change, and service providers have infinite capacity to complete tasks. In environments where agents may leave or join, or where their service offerings may change over time, strategic interleaving of exploration and exploitation may be beneficial.

(a) Cumulative Reward v. Time

(b) Probability of having successful interactions v. Time

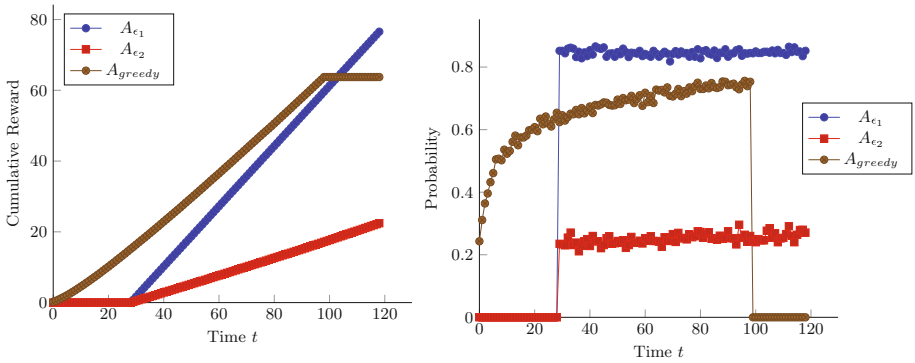

Fig. 2. Performance over time

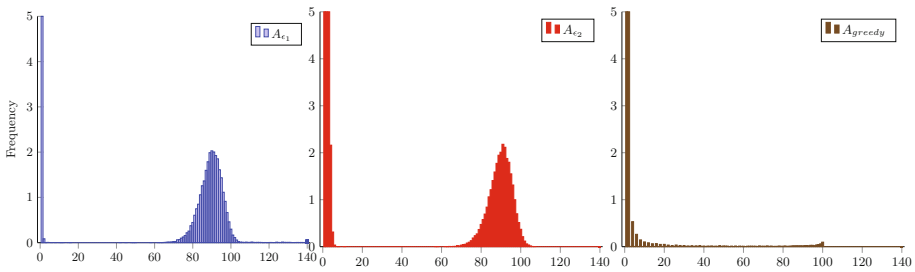

Fig. 3. Histogram of evidence $(r_i + s_i)$ in $[0, 140]$

5 Discussion

Optimal stopping in the class of problems referred to as the "secretary problem" resembles our problem of picking the right ϵ. In secretary problems [10] the applicants are interviewed one by one. The goal of the interviewer is to employ the best candidate. In these problems, however, each applicant is interviewed only once, and a decision to employ can only be made at that time. There are some similarities, however: the employer is not aware of the level of expertise of each applicant. Having agents leaving the environment is a challenging problem.

The environment that is dynamic requires trusters to trade-off continuing to interact with the current, best service provider or trying others.

Our environment is not dynamic, and this is a very strong assumption. In any practical system agents may enter and leave the system. Indeed, malicious agents may exploit this ability: they may create new identities to *whitewash* a poor reputation, or even collude with other agents to increase their perceived standing [11]. An important avenue for future research is to investigate how algorithms are robust to these kinds of attacks.

We adopted Subjective Logic as the basis for our trust model. There are other models, however, that may be employed. Wang and Singh's model [12], for example, takes conflicting evidence into account in computing a trust rating. One area for future research is to explore the interactions between the trust model employed and the algorithm used to spend a limited budget on acquiring direct and indirect evidence.

We plan to try different scenarios of witness information propagation not only environments that have a global reputation provider, but also the environments such that trustees have opinions about each other. The challenge of having opinions of trustees about each other is difficult in trust aware decision making problems. Since it complicates the process of properly assessing an agent. Is an agent honest if most of the time it provides a good service or if the witness information it provides is good?

6 Conclusion

In this paper, we have introduced a challenging problem of having interaction costs and budget limitations in trust and reputation systems. We investigated the performance of some simple algorithms, adapted from existing Budget-Limited Multi-Armed Bandit (BL-MAB) models. We evaluated these algorithms in a simulated environment with a central reputation provider. This is the first, but a very initial investigation into the use of witness information in trust-aware decision making when the decision maker is budget-limited, and where acquiring witness information is not cost-free. We have provided some evidence that strategic gathering of witness information can increase the number of successful interactions, despite this incurring costs on a limited budget. In future research, we will investigate varying service and witness information costs, and develop techniques to interleave exploration and exploitation.

References

1. Yu, H., Miao, C., An, B., Leung, C., Lesser, V.: A reputation management approach for resource constrained trustee agents. In: International Joint Conference on Artificial Intelligence (2013)
2. Jøsang, A., Ismail, R.: The beta reputation system. In: Proceedings of the 15th Bled Electronic Commerce Conference, vol. 5, pp. 2502–2511 (2002)

3. Regan, K., Poupart, P., Cohen, R.: Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In: Proceedings of the National Conference on Artificial Intelligence, vol. 21, p. 1206 (2006)
4. Teacy, W.L., Luck, M., Rogers, A., Jennings, N.R.: An efficient and versatile approach to trust and reputation using hierarchical Bayesian modelling. Artif. Intell. **193**, 149–185 (2012)
5. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decis. Support Syst. **43**(2), 618–644 (2007)
6. Sen, S., Ridgway, A., Ripley, M.: Adaptive budgeted bandit algorithms for trust in a supply-chain setting. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, pp. 137–144 (2015)
7. Tran-Thanh, L.: Budget-limited multi-armed bandits. Ph.D. thesis, University of Southampton (2012)
8. Etuk, A., Norman, T.J., Şensoy, M., Srivatsa, M.: How to trust a few among many. Auton. Agents Multi-agent Syst. **31**(3), 531–560 (2016)
9. Jøsang, A., Hayward, R., Pope, S.: Trust network analysis with subjective logic. In: Proceedings of the 29th Australasian Computer Science Conference, pp. 85–94 (2006)
10. Stein, W.E., Seale, D.A., Rapoport, A.: Analysis of heuristic solutions to the best choice problem. Eur. J. Oper. Res. **151**(1), 140–152 (2003)
11. Wang, D., Muller, T., Liu, Y., Zhang, J.: Towards robust and effective trust management for security: a survey. In: Proceedings of the 13th International Conference on Trust, Security and Privacy in Computing and Communications, pp. 511–518 (2014)
12. Wang, Y., Singh, M.P.: Formal trust model for multiagent systems. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence, pp. 1551–1556 (2007)