Empirical Analysis of Reputation-aware Task Delegation by Humans from a Multi-agent Game

(Extended Abstract)

Han Yu, Han Lin, Su Fang Lim, Jun Lin, Zhiqi Shen, and Chunyan Miao
LILY Research Centre, Nanyang Technological University, Singapore
{han.yu, linhan, sufang, jlin7, zqshen, ascymiao}@ntu.edu.sg

ABSTRACT

What are the strategies people adopt when deciding how to delegated tasks to agents when the agents' reputation and productivity information is available? How effective are these strategies under different conditions? These questions are important since they have significant implications to the ongoing research of reputation aware task delegation in multi-agent systems (MASs). In this paper, we conduct an empirical study to address the aforementioned research questions by providing a gamified mechanism for people to report the reputation-aware task delegation strategies they adopt. The findings from this empirical study may help MAS researchers develop multi-agent trust evaluation testbeds with more realistic simulated human behaviours.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents, Multiagent systems*

Keywords

Human-agent interaction; crowdsourcing; human computation; reputation; trust; task delegation; multi-agent game

1. INTRODUCTION

Trust is an important mechanism facilitating interaction among people who may not be familiar with each other in the beginning. Multi-agent trust research consists of three major directions [5]: 1) *reputation evaluation*: evaluating agents' reputation based on their past performance (e.g., [1]); 2) *task delegation*: delegating tasks based on agents' reputation; and 3) *performance evaluation*: comparing the performance of various proposed multi-agent trust approaches. In recent years, as crowdsourcing becomes a popular new form of computation, new research challenges for reputation aware task allocation have emerged. As crowdsourcing workers are human beings, they have limited availability and productive capacities to work on tasks delegated to them [6]. These workers can be referred to in general as "resource constrained trustees". Thus, multi-agent trust research needs to find ways to balance the workload workers in order to obtain high quality results from them in a timely manner.

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.*

As this research topic is relatively new, the proposed approaches (e.g., [3, 4, 7]) mostly rely on simulations for evaluation purposes. In order to construct realistic simulations to comprehensively assess the performance of a proposed approach, it is important to understand what task delegation strategies are used by people and how effective they are under different conditions. In this paper, we present results from a one year empirical study involving over 400 participants on the strategies adopted by human beings when making reputation aware task delegation decisions using a multi-agent game [8]. To the best of our knowledge, this research is the first empirical study with regard to this research problem. This paper makes the following contributions: 1) it proposes a taxonomy for reputation aware task delegation strategies; and 2) it provides empirical data and analysis on the performance of various strategies under different conditions.

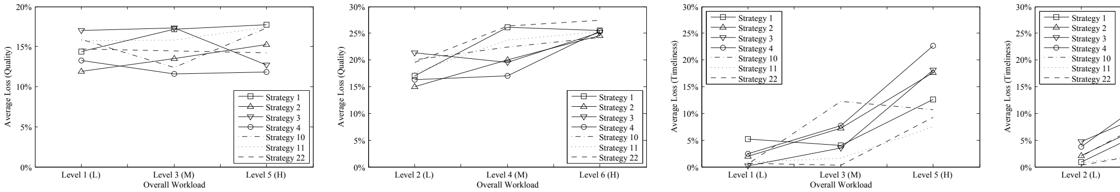
2. STUDY DESIGN

To study the strategies used by people to make reputation-aware task delegation decisions with resource constrained trustee agents, we use a multi-agent game - *Agile Manager* (AM) [8] - which provides opportunities for players to showcase their own strategies while unobtrusively collects data necessary for analysis. The game adopts the design concept of implicit human computation [2] through which players contribute data which are valuable for research.

This study consists of six settings as illustrated in Table 1. These six settings have been implemented as the six levels in the AM game. The *quality-quantity (QQ) tradeoff* variable is used to control the behaviour of programmer agents (PAs). If its value is set to 1, PAs' skill levels and productivity in the corresponding game level are negatively correlated; otherwise, PAs' skill levels and productivity in that game level are positively correlated. Through controlling the number of tasks which need to be assigned to different resource constrained agents and the average productivity output of the PAs, the *overall workload* compared to the PA team productive capacity that a player must delegate to the PAs has been divided into *Low* (L), *Medium* (M) and *High* (H) levels. At the end of each game session, a player is required to

Table 1: Study Settings

	Quality-Quantity (QQ) Tradeoff	
Overall Workload	0	1
Low (L)	Game Level 1	Game Level 2
Medium (M)	Game Level 3	Game Level 4
High (H)	Game Level 5	Game Level 6



(a) *Without* QQ Tradeoff (b) *With* QQ Tradeoff (c) *Without* QQ Tradeoff (d) *With* QQ Tradeoff

Figure 1: Strategy Effectiveness.

report the strategy he used during the game session. He can choose a mixture of available descriptive options.

In developing taxonomies for reputation aware task allocation strategies, we adopted a combination of "top-down" and "bottom-up" approaches. Strategies commonly mentioned in related literatures have been incorporated into the AM game as choices for self-reporting purposes ("top-down"). Then, we extend the taxonomies based on how frequent people utilize these strategies, either in their pure forms or as mixed strategies ("bottom-up").

Five different strategic choices are available for players to select. They are:

1. *Random Approach* (R): where a player delegates tasks to PAs at random without considering either its reputation or its current workload;
2. *Reputation-based Approach* (RA): where a player delegates more tasks to PAs with higher reputation (i.e., PAs with more stars in the context of the AM game);
3. *Equality-based Approach* (EA): where a player delegates among PAs as equally as possible;
4. *Load Balancing* (LB): where a player delegates more tasks to PAs which are having low current workload to avoid overloading any PAs as much as permitted by the situation;
5. *Others*: where a player may adopt strategies not listed in the choices and should provide additional descriptions about the adopted strategies.

A total of $2^5 - 1 = 31$ combinations are possible (a player must select at least one of the choices). The list of frequently used strategies is shown in Table 2.

3. KEY FINDINGS

In terms of average loss due to poor result quality (Figures 1(a) and 1(b)), Strategy 4 (LB) consistently achieves good performance. This is especially the case under medium to high workload conditions with familiar worker agents. The performance of Strategy 2 (RA) improves with decreasing workload under both familiar and unfamiliar worker agents conditions. From these results, it can be observed that under low overall workload conditions, the simple strategy of delegating more tasks to PAs with high reputation works well when players can become familiar with the PAs' performance over time. However, when the overall workload level becomes higher, load balancing or a mixture of reputation-based approach with loading balance is required in order to reduce the chances of delegating tasks to less reliable PAs.

Table 2: Strategy Taxonomy

	Random Approach	Reputation-based	Equality-based	Load Balancing	Others
1	✓				
2		✓			
3			✓		
4				✓	
10		✓	✓		
11		✓		✓	
22		✓	✓	✓	

In terms of average loss due to delays (Figures 1(c) and 1(d), the performance of the strategies is inversely related to the overall workload. Under all conditions studied, Strategies 11 and 22 achieves the best performance with Strategy 22 performing slightly better. From these results, it can be observed that the simple strategy of delegating more tasks to PAs with high reputation is not able to reduce overloading some PAs. However, simply adopting the equality-based approach or loading balancing without considering reputation also do not work well, especially under medium to high overall workload conditions. A mixture of reputation-based approach with loading balance is required in order to reduce the chances of overloading PAs and resulting in delays.

Acknowledgements

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office.

REFERENCES

[1] S. Liu, H. Yu, C. Miao, and A. C. Kot. A fuzzy logic based reputation model against unfair ratings. In *AAMAS'13*, pages 821–828, 2013.

[2] A. J. Quinn and B. B. Bederson. Human computation: A survey and taxonomy of a growing field. In *CHI'11*, pages 1403–1412, 2011.

[3] H. Yu, C. Miao, B. An, C. Leung, and V. R. Lesser. A reputation management model for resource constrained trustee agents. In *IJCAI'13*, pages 418–424, 2013.

[4] H. Yu, C. Miao, B. An, Z. Shen, and C. Leung. Reputation-aware task allocation for human trustees. In *AAMAS'14*, pages 357–364, 2014.

[5] H. Yu, Z. Shen, C. Leung, C. Miao, and V. R. Lesser. A survey of multi-agent trust management systems. *IEEE Access*, 1(1):35–50, 2013.

[6] H. Yu, Z. Shen, C. Miao, and B. An. Challenges and opportunities for trust management in crowdsourcing. In *IAT'12*, pages 486–493, 2012.

[7] H. Yu, Z. Shen, C. Miao, and B. An. A reputation-aware decision-making approach for improving the efficiency of crowdsourcing systems. In *AAMAS'13*, pages 1315–1316, 2013.

[8] H. Yu, X. Yu, S. F. Lim, J. Lin, Z. Shen, and C. Miao. A multi-agent game for studying human decision-making. In *AAMAS'14*, pages 1661–1662, 2014.