Incorporating Observation Error when Modeling Trust between Multiple Robots Sharing a Common Workspace

(Extended Abstract)

Shreyasha Paudel
Department of Aeronautics and Astronautics
Stanford University
Stanford, CA
spaudel2@stanford.edu

Christopher M. Clark
Department of Engineering
Harvey Mudd College
Claremont, CA
clark@hmc.edu

ABSTRACT

Most multi-robot systems assume all robots are programmed to cooperate and complete tasks without consideration of dishonest robots seeking to maximize their own benefit. The work presented here develops and analyzes a distributed trust estimation framework that allows robots to estimate the trustworthiness of other robots in the community, and share these estimates to enable cooperative trust estimation. Our previous work has shown that when observation errors were ignored, the trust-estimation error converges to a steady state value over time. We extend this previous work, by now considering the effects of observation error and proving that trust-estimation error converges.

Keywords

Cooperation, Multi-robot systems,Trust and Reputation

1. INTRODUCTION & RELATED WORK

In Multi-Robot Systems (MRS), it is common for robots to communicate and share information. However, it is not always clear if robots can trust the information being shared by another robot, especially if that robot was deployed by a different and possibly competing organization. To model and establish the trust between robots in this context, a framework was presented in [4], (see Fig. 1). Building on this previous work, proposed here is the addition of observation error. With this new addition comes the following contributions of this work: 1) introduction of observation error into the trust framework, and 2) analysis of the effects of observation error on trust estimation error convergence.

MRS have been developed for a wide variety of applications, often involving cooperative search and mapping [1, 2]. These systems typically use wireless communication and information sharing to minimize the individual robot path lengths required to complete the exploration of a workspace [2]. Lacking has been the ability to determine if the information shared between robots can be trusted.

Within the field of multi-*agent* systems, as opposed to MRS, reputation and trust models have become an integral part of the systems [6, 5]. Previous works range from simple online reputation models [6] which simply classify positive, neutral and negative to a sophisticated model such as FIRE [3] and TRAVOS [9] which computes a trust value for each agent and a reliability measure taking into account both direct experience, witness information and third party certification. Recently, the work in [8, 7] used an observation based trust model for detecting unreliable team members in a multi-robot patrolling task. While our work is similar to [8] in that we consider a binary observation model, our model does not require a dedicated observer to detect false observations. Our work also differs in that it considers cases in which some robots attempt to provide false information.

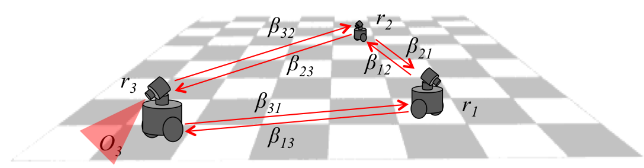

Figure 1: Illustration of the trust framework. Each robot r_i estimates the likelihood $\beta_{i,j}$ that robot r_j broadcasts the true value of its observations $O_{j,k,t}$ of cell k at time t.

2. PROBLEM DEFINITION

Given: There exists a set $R = \{r_1, r_2, ..., r_n\}$ of n robots sharing a common workspace. Each robot's goal is to estimate a set of K binary variables of interest $p_k \in \{0, 1\}, k = 1..K$. Each variable has an associated workspace position X_k. Each robot r_i is equipped with the neccessary sensors to take observations $O_{i,k,t} \in [0, 1]$ of p_k at time t when within distance d of X_k.

Using inter-robot communication, robots can share observations, and reduce the number of locations to visit in the workspace. However, each robot r_i may not always send truthful information, i.e. it only broadcasts the actual set of mean measurements $\bar{O}_{i,k,t}$ with probability β_i. Otherwise, it broadcasts 1-$\bar{O}_{i,k,t}$ with probability 1-β_i. To accommodate this, each robot must determine how much they trust shared observations by estimating the likelihood β_j that each robot j sends its actual set of mean observations. Specifically, we say that robot r_i's estimate of β_j at time t is $\hat{\beta}_{i,j,t}$.

If robots also share estimates of $\hat{\beta}_{i,j,t}$ through wireless

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

communication, they can more quickly estimate β_j values. We must introduce $\hat{\beta}'_{i,j,t}$ as the value robot r_i broadcasts to other robots as its estimate of β_j. Note that, $\hat{\beta}'_{i,j,t} = \hat{\beta}_{i,j,t}$ only if robot r_i is truthful. Deceiving robots inflate (or deflate) $\hat{\beta}'_{i,j,t}$ by $\epsilon_{\beta,i}$. That is $\hat{\beta}'_{i,j,t} = \hat{\beta}_{i,j,t} + \epsilon_{\beta,i}$.

Given this scenario, each robot r_i in R must calculate an estimate $\hat{\beta}_{i,j,t}$ of the likelihood that each of the other robot r_j is broadcasting truthful observations at every time step t.

3. TRUST FRAMEWORK

In this framework, trust values are estimated by using two pieces of information - 1) trust values $\hat{\beta}'_{k,j,t}$ broadcasted by other robots and 2) individual robot measurements of truthfulness of broadcasted messages $z_{i,j,t}$. This information is incorporated into the proposed discrete time update rule:

$$\hat{\beta}_{i,j,t+1} = \frac{1}{n}\sum_{k=1}^{n}\hat{\beta}'_{k,j,t} + K_\beta(z_{i,j,t} - \frac{1}{n}\sum_{k=1}^{n}\hat{\beta}'_{k,j,t}) \quad (1)$$

In (1), n is the number of robots in R, K_β is a constant proportional gain hereafter referred as the trust constant. The variable $z_{i,j,t}$ is robot r_i's measurement of the fraction of robot r_j's statements that are true. The first term averages the trust estimates shared by all robots in R and is used as a predictive element. The second term uses the robot's individual observations as a corrective element.

A binary measurement system is assumed in which the i^{th} robot makes an observation $O_{i,k,t} \in \{0,1\}$ of some variable $p_k \in \{0,1\}$. The observation is subject to sensing errors $\epsilon_{s01,i}$ and $\epsilon_{s10,i}$. The framework incorporates the possibility of robots misinforming one another. Specifically, a robot r_j broadcasts to all other robots its mean estimated observation value $\bar{O}_{j,k,t}$ with a likelihood of β_j, and conversely $1 - \bar{O}_{j,k,t}$ with a likelihood of $1 - \beta_j$.

It can be shown that for constant β_j values, mean inflation error $\bar{\epsilon}_\beta$, and sensing errors ϵ_{s10i}, ϵ_{s01i}, the steady state error in robot r_i's trust estimation of robot r_j using the proposed framework converges over time and is equal to:

$$\hat{\beta}_{i,j,t\to\infty} = \beta_j + \frac{1-K_\beta}{K_\beta}\bar{\epsilon}_\beta + \epsilon_{s10,i} + \epsilon_{s10,j}(1-2\beta_j) \quad (2)$$

4. VALIDATION VIA SIMULATION

The trust framework was simulated within an exploration task. A square grid map of K cells was defined where each cell C_k was assigned a value $p_k \in \{0,1\}$. The $n = 4$ robots in R were tasked with estimating p_k for all the map cells. Each robot was assigned a $\beta_i \in [0,1]$. At each timestep, each robot broadcasted a message stating the mean observation value $\bar{O}_{i,k,t}$ of the cell C_k that it currently occupies. A robot r_i broadcasts its true observation with probability β_i. The robots also broadcasted truth estimates $\hat{\beta}'_{i,j,t}$ of all the robots in R.

Eq. (2) was validated with simulations where a set of robots were placed in the single cell and shared their observations of p_k where $k = 1$. Simulations were run for 20000 iterations, (i.e. time steps) with various values of β_i, K_β, $e_{s01,i}$, $e_{s10,i}$, and $\epsilon_{\beta,i}$. Figure 2 shows a plot of $\hat{\beta}_{i,j,t}$ vs Timestep for one of the single cell estimation simulations with sensing error $\epsilon_{s01,i} = \epsilon_{s10,i} = 0.1$ and β inflation $\epsilon_{\beta,i} = 0.05$ $\forall i$. As expected, $\hat{\beta}_{i,j,t}$ converges to the theoretical prediction.

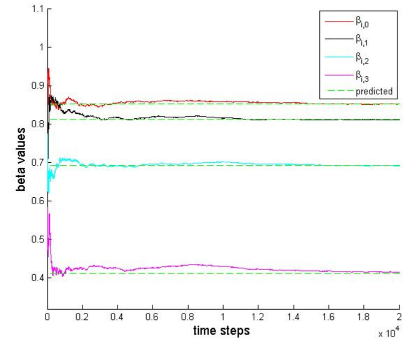

Figure 2: Trust estimate vs timestep for varying beta values in a simulation of single cell map. β estimates converge to theoretical prediction over time.

5. CONCLUSION

This work proposes a method for robots to identify trustworthy robots in a workspace where deceptive robots might be present. By monitoring the information communicated by robots, the level of trust in each robot can be estimated. The estimation of a robot's truthfulness is shown to converge to a bounded error, despite errors in sensing.

REFERENCES

[1] A. Birk and S. Carpin. Merging occupancy grid maps from multiple robots. *Proceedings of the IEEE*, 94(7):1384–1397, July 2006.

[2] W. Burgard, M. Moors, C. Stachniss, and F. Schneider. Coordinated multi-robot exploration. *Robotics, IEEE Transactions on*, 21(3):376–386, June 2005.

[3] T. D. Huynh, N. R. Jennings, and N. R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[4] K. T. Mantel and C. M. Clark. Trust networks in multi-robot communities. In *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, pages 2114–2119. IEEE, 2012.

[5] Z. Noorian and M. Ulieru. The state of the art in trust and reputation systems: a framework for comparison. *Journal of theoretical and applied electronic commerce research*, 5(2):97–117, 2010.

[6] I. Pinyol and J. Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review*, 40(1):1–25, 2013.

[7] C. Pippin and H. Christensen. Trust modeling in multi-robot patrolling. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2014)*, 2014.

[8] C. E. Pippin, H. Christensen, and L. Weiss. Dynamic, cooperative multi-robot patrolling with a team of uavs. volume 8741, pages 874103–874103–8, 2013.

[9] W. L. Teacy, J. Patel, N. R. Jennings, and M. Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.