Identifying Beneficial Teammates using Multi-Dimensional Trust (short paper)

Jaesuk Ahn, Xin Sui, David DeAngelis, K. Suzanne Barber
The University of Texas at Austin
Austin, TX, 78712 USA
+1-512-471-5350
{jsahn, xsui, dave, barber}@lips.utexas.edu

ABSTRACT
Multi-agent teams must be capable of selecting the most beneficial teammates for different situations. Multi-dimensional trustworthiness assessments have been shown significantly beneficial to agents when selecting appropriate teammates to achieve a given goal. Reliability, quality, availability, timeliness and compatibility define the behavioral constraints of the multi-dimensional trust (MDT) model. Given the MDT model in this research, an agent learns to identify the most beneficial teammates by prioritizing each dimension differently. An agent's attitudes towards rewards, risks and urgency are used to drive an agent's prioritization of dimensions in a MDT model. Each agent is equipped with a Temporal-Difference (TD) learning mechanism with tile coding to identify its optimal set of attitudes and change its attitudes when the environment changes. Experimental results show that changing attitudes to give preferences for respective dimensions in the MDT offers a superior means to finding the best teammates for goal achievement.

Categories and Subject Descriptors
I2.11 [**Distributed Artificial Intelligence**]: Multiagent systems – *Multiagent systems*

General Terms
Algorithms, Design, Experimentation

Keywords
Coalition formation, Partner selection, Multi-dimensional trust

1. INTRODUCTION
Multi-agent systems have been applied to distributed problem solving applications because of their capability to overcome the limitation of individuals in solving a complex problem. Furthermore, increasing the number of software agents acting as problem solvers on a network suggests a potential marketplace, where large groups of self-interested agents interact with each other and solve problems over various domains by taking different roles and forming "temporary teams" [7]. In this kind of scenario, problems that agents encounter often have multiple requirements to be satisfied. Accordingly, when working as a team, teammates may possess different behavioral constraints governing how they complete given sub-tasks, thus affecting a team's overall problem solving performance.

Cite as: Identifying Beneficial Teammates using Multi-Dimensional Trust (Short Paper), J. Ahn, X. Sui, D. DeAngelis, and K. S. Barber, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16, 2008, Estoril, Portugal, pp. 1469-1472.

As shown in the previous work [1], an agent's behavioral constraints can be modeled using multi-dimensional trust (MDT). Many researchers have shown that modeling the trustworthiness of others according to multiple dimensions can significantly benefit partner selection [2, 3, 4]. In this research, an agent's behavioral constraints are modeled as a multidimensional trust (MDT) model, which includes both performance-related constraints (reliability, availability, timeliness, and quality of service) and relationship-related constraint (compatibility).

This research develops a teammate selection algorithm centered on proposed MDT models. During the teammate selection process, an agent learns to recognize the most "beneficial" agents given the situation by valuing each dimension in the trust model differently based on feedback from the environment. The concept of attitudes is used to value each dimension in the trust model. Attitude represents the tendency to act in a certain way towards the objects, which can be used as a good predictor of behavior [6]. In this sense, attitude can be described as a set of parameters to dictate an individual agent's behavior. Three attitudes are used here, which are attitude toward reward, risk, and urgency. In addition, an agent learns the optimal set of attitudes given different situations from interaction with the surrounding environment. Temporal-Difference (TD) learning is used to determine a set of attitudes given different situations, and tile coding is used as a function approximation method to simplify the learning state space [5]. Using the proposed TD learning with tile coding approximation, an agent can estimate the optimal attitude despite incomplete information about the environment. This paper demonstrates an agent can earn more rewards by 1) modeling other agents' behavioral constraints as multi-dimensional trust and 2) using attitudes to prioritize important factors during the teammate selection process.

2. TEAMMATE SELECTION
In the problem-solving domain, each problem can be regarded as a goal to achieve, and has the constraints (e.g., time and quality constraints) to be satisfied. In this paper, we assume the problem is already decomposed into the sub-tasks. When a team of agents completes all the sub-tasks in a problem within the given time constraints, the payoff is given to the agents who worked on each sub-task. The amount of payoff depends on the quality of a team's final solution. When a team fails to complete all the sub-tasks, a penalty is given to the team leader. In this sense, a problem has multiple requirements that affect the amount of payoff or penalty a team can receive. When agent a_i works on a task (t_k) in a problem (P_i) with n number of agents as a team, the actual payoff agent a_i can receive is defined as follows:

$if\ \forall t \in P_i\ is\ completed\ within\ r_t$

$$Rw^*_{a_i}(t_k) = \frac{\sum q_{a_j}}{n} * Rw(t_k)$$

$Otherwise,$

$$Rw^*_{a_i}(t_k) = 0$$

$where$

$r_t: time\ constraint\ of\ J_i$

$Rw(t_k): original\ payoff\ of\ task\ (t_k)$

$q_{a_j}: quality\ of\ solution\ each\ agent\ (a_j)\ provide$

$\frac{\sum q_{a_j}}{n}: averge\ quality\ of\ solution\ a\ team\ provide$

Accordingly, each potential teammate's behavioral constraints should be considered to find out whether they are able to meet all the required constraints (quality and time constraints) as a team to earn more payoffs. These behavioral constraints can be modeled as multi-dimensional trust.

2.1 Building Multi-Dimensional Trust Models

When it is difficult for an agent to know other agent's behavior constraints *a priori*, agents must build models of these constraints over time by interacting with agents. Multi-dimensional trust (MDT) models can represent other agents' behavioral constraints allowing an agent to identify beneficial teammates. Four dimensions are defined as performance-related MDT (P-MDT):

- Reliability (d_r): Probability to fulfill the commitment
- Quality (d_q): Quality of Service being provided
- Availability (d_a): Availability to be a teammate
- Timeliness (d_t): Time required to complete given task

In addition, an agent also has a behavioral constraint that affects its relationship with other agents. Since agents are required to work as a team, identifying a beneficial behavioral constraint to promote teaming ability is also important. One dimension is defined as relationship-related MDT (R-MDT):

- Compatibility (d_c): Preference to be a teammate

Compatibility can be built up through the continuous positive interactions between two agents. When compatibility of agents in the team is high, the team produces solutions faster. Compatibility increases when there are more positive interactions between two agents.

2.2 Selecting Teammates

An agent identifies the most beneficial teammates by prioritizing each dimension differently given the situation. The prioritization of each dimension can be adjusted based on an agent's attitudes to give the best estimation of other agent's MDT score in any given situation. When using other agents' MDTs to decide whom to select as teammate, the helpfulness of potential teammates is calculated as a weighted sum of multiple dimensions. The weighting parameters can be defined as an agent's attitudes, which influence its selection of beneficial teammates by determining which dimension is important to consider in any given situation. Attitude models are represented as follow:

- a_{rw}: Attitude toward reward [0, 1]: an agent's willingness to seek for agents with high quality of service
- a_{rsk}: Attitude toward risk [0, 1]: an agent's sensitivity to possible risk (unreliability and unavailability of agents)
- a_{time}: Attitude toward urgency [0. 1]: an agent's willingness to seek for agents who can provide solutions quickly

An agent *j* estimates helpfulness of an agent *i* using the proposed dimensions and attitudes. Since Quality of agent affects the amount of payoff a team gets, the quality dimension (d_q) is considered as potential reward. In addition, since low Reliability and Availability (d_r and d_a) can increase the possibility of team failure, these two dimensions are considered as potential risk. The Timeliness (d_t) and Compatibility (d_c) are considered when there is less time to complete a problem since an agent with high timeliness or high compatibility can provide solutions faster than other agents can. An agent *j* estimates the helpfulness of the agent *i* at the time *t* as follows:

$$H^i_j(t) = a_{rw} * d_q - \alpha * a_{rsk} * (d_r + d_a) + \beta * a_{time} * d_t + d_c$$

Once the helpfulness of agent *i* is calculated, an agent *j* calculates multi-dimensional trust (MDT) of the agent *i* as follows:

$$MDT^i_j(t) = H^i_j(k) * f\left(n^i_j(t)\right),\ where\ f\left(n^i_j(t)\right) = \log_\tau n^i_j(t)$$

$f\left(n^i_j(t)\right)$ is function of the number of times agent *j* has interacted with agent *i*, $n^i_j(t)$. This function weights helpfulness more heavily when there have been more interactions between two agents. Because an agent builds its MDT model based on the feedback from the previous interactions, it is difficult to build an accurate MDT model when there are fewer interactions between agents. Initially, when there are fewer interactions between agents, an agent explores unknown set of agents rather than relying on the MDT model to select teammates. The exploration scheme is based on the ε-greedy algorithm. An agent builds a list of agents in descending order of MDT values. Then, the agent chooses the potential teammate with probability of 1- ε, or chooses a random agent with probability of ε. The exploration rate ε determines the tradeoff between exploration and exploitation. Exploration rate ε decays over time. Since agents are able to build accurate MDT models of other agents when the number of interactions increases over time, an agent's value of ε decays with time to exploit its MDT models.

3. LEARNING AND ADAPTATION

The problem constraints may change over time with some degree of uncertainty in a dynamically changing environment. In order to address these changes, agents need to estimate the current state of the environment and learn the optimal set of attitudes given this estimation. In this research, single agent reinforcement learning is used to alter an agent's set of attitudes in order to find the most appropriate set to identify beneficial teammates. First, each set of two attitudes (attitude toward reward and risk) is defined as a strategy (s_k). An agent has a set of strategies (S); each strategy contains a unique set of two attitudes.

$$S = \{s_0, \ldots, s_n\},\ where\ s_i = \{a_{rw}, a_{rsk}\}$$

$$s_k = \{i, j\}, \qquad i, j \in [0, 10]$$

In this context, a strategy is the state of an agent, and choosing a strategy, or set of attitudes, can be defined as an action (a) of an agent. An agent selects a strategy (a set of two attitudes) to calculate the MDT value of potential teammates and rank them according to this value. In the experiments, the agent uses the Q-learning algorithm with linear tile-coding function approximation and accumulating eligibility traces. Since the state has two continuous variables (two attitudes), the state space can be tiled diagonally in a two-dimensional grid. Since Attitude toward reward and Attitude toward risk is presented as a weight parameter in building MDT model, the value difference between two attitudes is important to consider. In this sense, diagonal tiling gives a good generalization over this type of binary state space.

An agent uses the RL formula to update the estimated Q value associated with the outcome of the current action.

$$Q_k(a) = Q_{k-1}(a) + \alpha\big(p(a) - Q_{k-1}(a)\big)$$

where k is the number of timesteps and $p(a)$ is actually the payoff (or penalty) received on the k^{th} timestep using action a. An agent increases its attitude toward urgency when it experiences repeated failures due to unmet time constraints.

4. EXPERIMENTS

The experimental environment consists of a set of self-interested agents and problems (Table 1). Only a set of agents with the required capabilities can perform the sub-tasks in each problem. Problems have a time constraint. Each agent knows its own capabilities and the capabilities of all the other agents, and must form a team to complete the entire set of sub-tasks within the given time constraint. Otherwise, the leader of the team must pay a given penalty to the problem owner. The payoff is distributed to the agents who worked on each sub-task instance when all sub-task instances in a problem are successfully completed within the time constraint. Whenever a problem is completed or failed in the environment, a new problem is introduced as a replacement.

4.1 Effect of MDT and Attitudes

The objective is to see the possible effect of each attitude on the agent's outcome when choosing a teammate. Success is measured in terms of the total outcome, which subtracts the total penalty value from the total payoff. Agents are grouped into six classes, and four types of naïve agents were included (Table 2). Naïve agents do not have an attitude-based teammate selection mechanism, and always work as a pool of potential teammates. Figure 2 shows the average outcome each class earned during the experiment. When a time constraint is very tight (time constraint = 18), hiring unreliable members or rejection from unavailable agents might easily cause an agent penalty due to the problem completion failure. Therefore, the class 1 and class 4 agents with strong attitude toward risk can avoid agents with low reliability and availability. Even though class 1 and 4 earn less amount of payoff per problem than other classes, class 1 and class 4 achieve the highest outcome by avoiding a penalty due to the possible failure. On the other hand, the class 3 agents have a higher probability of failure under tight time constraints because of strong attitude toward reward. Class 3 is able to seek teammates with a higher timeliness value, who finish assigned sub-tasks faster than a class 0 agent's teammates. Therefore, a class 3 agent can have more opportunity to complete given problem within time constraints. Since the class 0, class 2, and class 3 pairing has a relatively strong attitude toward reward and weak attitude toward risk, class 0, 2, and 3 care less about reliability and availability of agents; thus increase possibility to fail.

Table 1. Common parameters and values

Parameters	Values
Number of Agents Number of Problems	54 (each has 1 capability) 8 (each has two tasks)
Time constraint in the problem	18, 45, 78, 90
Task Payoff (t_{payoff})	7.5
Penalty ($P_{penalty}$)	Varies between 0 and 4

Table 2. Agent's Class

Class	Attitudes (A)
Class 0	A-Reward (0.8), A-Risk (0.2), A-Urgency (0.2)
Class 1	A-Reward (0.2), A-Risk (0.8), A-Urgency (0.2)
Class 2	A-Reward (0.5), A-Risk (0.5), A-Urgency (0.5)
Class 3	A-Reward (0.8), A-Risk (0.2), A-Urgency (0.8)
Class 4	A-Reward (0.2), A-Risk (0.8), A-Urgency (0.8)
Class 5	Random Teammate Selection
Naïve 1	Reliability(0.2 / 0.5), Quality(0.8), Timeliness (0.1)
Naïve 2	Reliability(0.8), Quality(0.2), Timeliness (0.1)
Naïve 3	Reliability(0.2 / 0.5), Quality(0.8), Timeliness (0.9)
Naïve 4	Reliability(0.8), Quality(0.2), Timeliness (0.9)

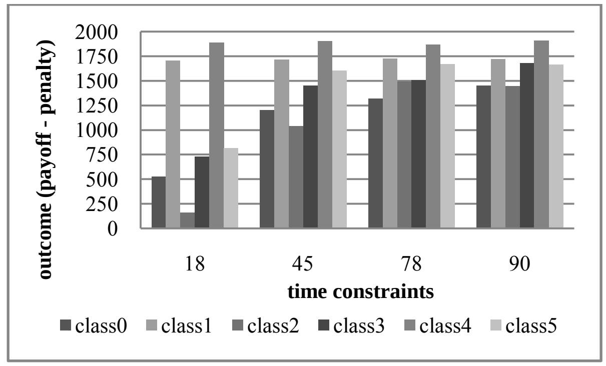

Figure 2. Average Outcome when the lowest reliability is 0.2 (Penalty=4, Payoff=7.5)

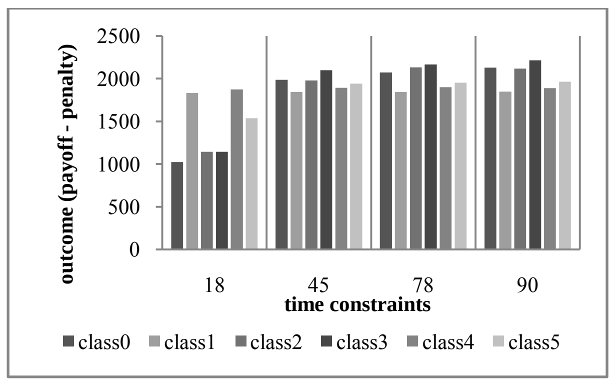

Figure 3. Average Outcome when the lowest reliability is 0.5 (Penalty=4, Payoff=7.5)

However, when time constraints relax, class 0, class 2 and class 3 agents are able to increase their outcome since there is more time to hire additional members even if some of them are unreliable or unavailable.

In the next experiment, we increased the reliability of naïve agent 1 and 3 from 0.2 to 0.5. Class 1 and class 4 still outperform others when a time constraints is 18, Classes 0 and 3 start outperforming class 1 and 4 after a time constraint increases (Figure 3). The result was statistically significant. Since the low reliability naïve agents (naïve agent 1 and naïve agent 3) have higher reliability (0.5) than the previous experiment (0.2), there is less chance of failure when agents select naïve agents with lower reliability Therefore, the class 0 and 3 are able to lower their failure rate when a time constraints is larger than 18.

4.2 Learning (Altering Attitudes)

In this set of experiments, we examine the effectiveness of our learning mechanism. Table 3 shows four cases of the agent's class used in the experiment. Class 0 is always a learning agent, and classes 1, 2, and 3 change their type per each case. In addition, behavioral constraints of naïve agents were assigned randomly to reflect the openness of the environment. A time constraint during the experiment also changes between 18, 45, and 90 to reflect the dynamics of environment. Figure 4 shows the experiment result for each case. The class 0 agent who is able to alter attitudes is able to perform well in most cases. Since the classes 0, 1, and 2 outperform class 3, the benefit of having MDT models and a set of attitudes is clear when the environment is dynamic. The class 0 performs well in case 1 when it competes with agents which care more about reliability and availability. In case 2, 3, and 4, the class 0 agents have a significant difference from the class 1 and 3 agents. This means that altering an agent's attitude to prioritize a subset of the three dimensions is better than equal weights on all dimensions or random selection. In case 2 and 4, average outcome of class 0 is slightly less than class 2 but it was not significant. Since class 2 has a strong attitude toward urgency from the initial state, class 2 was able to find agents who can solve remaining tasks quickly even though there were unavailable or unreliable agents during the initial stage of team formation. Having strong attitude toward urgency from the initial stage compensate class 2's weak attitude toward risks, thus decrease failure rate. Therefore, class 2 was able to perform as good as class 0 agents.

Table 3. Agent's Class

Cases	Case 1	Case 2	Case 3	Case 4
Class 0	Learning Agents			
Class 1	{0.5, 0.5, 0.5}*,	{0.5, 0.5, 0.5},	{0.5, 0.5, 0.5},	{0.5, 0.5, 0.5}
Class 2	{0.2, 0.8, 0.7},	{0.8, 0.2, 0.7},	{0.6, 0.4, 0.3},	{0.7, 0.3, 0.7}
Class 3	Random,	Random,	Random,	{0.3, 0.7, 0.3}
Naïve	Reliability, Quality, and Timeliness (Randomly Assigned)			

* {Attitude toward Reward, Risk, and Urgency}

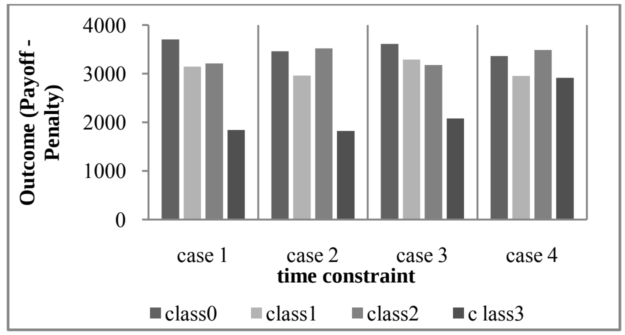

Figure 4. Average Outcome (Penalty=4, Payoff=7.5)

5. CONCLUSIONS

When forming teams, an agent needs to identify the helpfulness of other agents as potential teammates to maximize the reward it receives from solving a problem. Especially, when the problem has multiple constraints to be satisfied, an agent must consider the trustworthiness of potential teammates relative to multiple dimensions accounting for multiple problem requirements. This research endows agents with the ability to assert how much it should trust multiple facets of a potential teammate's behavior – the *availability* of an agent to deliver *quality* solution with *reliability* in a *timely* manner – in the context of multiple problem requirements. The teammate selection algorithm allows an agent to use multiple dimensions to estimate how much a potential teammate can be trusted to complete a problem within a time constraint. In this research, the attitude models play a role as a guide to shape an agent's teammate selection. Three attitudes are proposed: attitude toward reward giving preference to the quality dimension, attitude toward risk, giving preference to reliability and availability, and attitude toward urgency giving priority to both timeliness and compatibility dimensions. A method is proposed to build a multi-dimensional trust model using agents' attitudes to give priority to a subset of the five dimensions during the teammate selection process. The experiments show the clear effect of an agent's attitude on the outcome. In addition, given a simple reinforcement learning technique to alter its attitudes, an agent is able to identify the optimal set of attitudes (attitude toward reward, risk, and urgency) to solve the team formation problem.

6. ACKNOWLEDGMENTS

This research is sponsored in part by the Office of Naval Research Project # N00014-06-1-0062. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

7. REFERENCES

[1] Ahn, J, David DeAngelis, K. S. Barber, "Attitude Driven Team Formation using Multi-Dimensional Trust", To be appeared at the 2007 IEEE / WIC / ACM International Conference on Intelligent Agent Technology, Silicon Valley, CA, Nov 2-5.

[2] Griffiths, N., "Task Delegation using Experience-Based Multi-Dimensional Trust", *In the Proceedings of the Fourth International Conference on Autonomous Agents and Multi-agent Systems (AAMAS-05)*, Utrecht, The Netherlands, 2005, pp. 489-496.

[3] Gujral, N., D. DeAngelis, K. Fullam, and K. S. Barber. "Modeling Multi-Dimensional Trust," In *the Proceedings of The Workshop on Trust in Agent Societies* at *The 5th International Conference on Autonomous Agents and Multiagent Systems*, Japan, May 8-12, 2006; pp. 35-41.

[4] Maximilien, E. M. and M. P. Singh, "Agent-Based Trust Model Involving Multiple Qualities", *In the Proceedings of Fourth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-05)*, Utrecht, The Netherlands, 2005, pp. 519-526.

[5] Oliveira, E., J. M. Fonseca and N. R. Jennings, "Learning to be Competitive in the Market". in *Proceedings of AAAI Workshop on Negotiation: Settling Conflicts and Identifying Opportunities*, Florida, USA, 1999, pp. 30-37.

[6] Petty, R. E., D. T. Wegener, and L. R. Fabrigar, "Attitudes and Attitude Change", *Annual Review of Psychology,* 48, 1997, pp. 609-647.

[7] Sutton, R. S., Andrew G. Barto, "Reinforcement Learning: An Introduction", A Bradford Book, The MIT Press, Cambridge, Massachusetts, 1999.