1-6-2009

A Multiagent Framework for Human Coalition Formation

Nobel Khandaker
University of Nebraska, knobel@cse.unl.edu

Leen-Kiat Soh
University of Nebraska, lsoh2@unl.edu

University of Nebraska–Lincoln, Computer Science and Engineering
Technical Report TR-UNL-CSE-2009-0003
Issued Jan. 6, 2009

A Multiagent Framework for Human Coalition Formation

Nobel Khandaker and Leen-Kiat Soh

Computer Science and Engineering
University of Nebraska-Lincoln
256 Avery Hall, Lincoln, NE 66588-0115

(402) 472-6738

{knobel, lksoh}@cse.unl.edu

ABSTRACT

Human users form coalitions to solve complex tasks and earn rewards. Examples of such coalition formation can be found in the military, education, and business domains. Multiagent coalition formation techniques cannot be readily used to form human coalitions due to the unique aspects of the human coalition formation problem, e.g., uncertainty in human user behavior and changes in human user behaviors due to human learning. Thus, a multiagent system designed to form human coalitions has to solve a learning problem, that is further made difficult by the limited learning opportunities and usability issues (i.e., actions or decisions being perceived as not useful due to loss of immediate rewards while the agents are learning or exploring) intrinsic to the human coalition formation process. We propose and design a multiagent framework that distinguishes the impact of the *model of a human user* from that of the *agent support for that model*. This novelty allows an agent to (1) better compute the types of support it should provide to its assigned user and (2) more accurately estimate the value of a coalition by its ability of (a) solving the current task and (b) improving the coalition members' behavior due to learning. In our design, each agent models its environment using a Bayesian network and forms human coalitions for its assigned user using a negotiation-based protocol. Each coalition balances the immediate and future rewards by analyzing the benefits of solving tasks and facilitating human learning. To evaluate the proposed framework, we have built a comprehensive simulation where agents support students to form coalitions in a collaborative learning environment. Our results show that the framework is able to form successful coalitions that facilitate student learning while solving tasks, leading to overall better rewards for student coalitions.

Categories and Subject Descriptors

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

General Terms

Algorithms, Design, Experimentation, Human Factors.

Keywords

Multiagent, Coalition Formation, Human Coalition Formation, Computer-Supported Collaborative Learning.

A Multiagent Framework for Human Coalition Formation, Author(s), *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

1. INTRODUCTION

Human users form coalitions to solve complex tasks and to earn rewards. Examples of such coalition formation process can be found in the business (e.g., organizations forming coalitions to earn better discounts), education (e.g., students forming coalitions to solve problems and gain knowledge), and in the military (e.g., soldiers forming coalitions to complete difficult missions) domains. With the advancements in the computer and communication technologies, human coalitions now span space (members spread across the globe), and time (group members communicating asynchronously). Although human coalitions are becoming common in various domains, achieving optimal outcome in a human coalition is not automatic and depends on three aspects: (1) the coalition structure, (2) the uncertainty in the environment, and (3) changes in human behavior. The coalition structure defines the distribution of the human users of different skills to the coalitions; the uncertainty in the environment and in the human user behavior determines how they use those skills; and changes in human behavior are due to how human skills improve due to learning.

Multiagent researchers have designed frameworks where agents take the uncertainty in the environment and the changes in other agents' behavior into account and generate coalition structures. However, those multiagent coalition formation methods cannot be readily used to form human coalitions for three reasons: (1) human changes due to learning, (2) short- and long-term rewards, and (3) human user modeling:

- Since humans learn, a multiagent solution for the human coalition formation problem (HCFP) must consider two issues. *First*, the skills and experience gained by a human user while working in a coalition can help him or her perform tasks better—and thus obtain higher rewards—in future coalitions. So, the agents that are helping the human users form coalitions need to model their own human users' as well as the other human users' *learning* to accurately estimate the outcome of a coalition. *Second*, the agents helping the human users form coalitions can *provide support* to the human users and *exploit* the human users' learning capability to further improve the outcomes of a coalition. So, while forming coalitions, the agents also need to contemplate the support that they provide to the human users to more accurately estimate the value of a coalition in terms of its outcome.
- Furthermore, the rewards of a typical HCFP can be of two types: *current-task* and *future-task* rewards. The complications lie with the future-task or the long-term rewards. In a multiagent system (MAS), long-term rewards are often calculated as discounted short-term rewards (e.g., [1]). However, in a HCFP, long-term rewards typically depend upon factors such as the openness—how often the tasks and users change in the environment—and how fast or how well the human users are able to ac-

quire new skills through learning. So, computing long-term rewards in a HCFP will need to consider these factors integrally.

- Finally, unlike conventional multiagent systems, the actors of the human coalition formation problem are human users (not agents). The implication is that the actors in the human coalition formation problem may not always behave rationally or consistently. So, a MAS that forms human coalitions must model such uncertainty regarding human user behaviors while reasoning to form coalitions.

In this paper, to solve a HCFP, we propose a multiagent framework called Multiagent Human Coalition Formation (MHCF). In this framework, agents work as assistants to the human users and help them form coalitions and solve assigned tasks to earn rewards. The strength of our novel approach lies in our combined use of multiagent and generic human-behavior-related knowledge to address the aforementioned key issues (i.e., human type transformation due to learning, reward structures, and user modeling). Briefly, first, MHCF agents address the type transformation issue by considering (1) the agent support, (2) the model of the assigned human user, and (2) the change in that model due to his or her learning, when estimating the value of a human coalition. The agents address the reward structure issue by using a probabilistic model of the environment that take the openness of the system and the learning of the human users into account to estimate the future-task rewards. Finally, the agents address the user modeling issue by cooperatively learning a probabilistic model of its assigned human users and the dynamics of the environment, where the cooperative learning enables the user agents to (1) better estimate the uncertain behavior of his or her assigned human user and (2) learn about the environment dynamics faster. To investigate the effectiveness of the MHCF framework, we have used it to form student coalitions in SimCoL (designed by [4]) —a simulation of students and computer-based scaffolding in a computer-supported collaborative learning environment. Our results show that the framework is able to form successful coalitions that facilitate student learning while solving tasks, leading to overall better rewards for student coalitions.

2. Human Coalition Formation

2.1 Basic Assumptions and Definitions

Reflecting on the human coalition formation problem (HFCP), we have the following basic *Assumptions*: (1) (**Task**) There is a set of tasks in the environment that the human users in the environment can solve to earn rewards; (2) (**Coalitions**) The human users can form disjoint coalitions to earn rewards; (3) (**Behavior**) The solution of each task requires a coalitional action which is composed of a series of individual actions from the members of the formed coalition; (4) (**Learning**) The human users participating in the task solving activities of their respective coalitions are capable of learning from their experiences and improving their behavior, (5) (**Uncertainty**) The behaviors of the human users, i.e., their individual actions, in a coalition cannot be accurately modeled, and can only be modeled probabilistically; and (6) (**Reward**) The rewards achievable by a human coalition are of: (1) current-task (short-term) rewards and (2) future-task (long-term) rewards. Further, the current-task reward is a human user's share in the coalitional reward his or her coalition has earned by solving an assigned task and can be *exactly* calculated as soon as the coalition completes the assigned task. The future-task reward is an estimation of the current-task rewards of the future tasks that is expected to be available as a human improves as he or she learns.

Definition 1 (**Human Coalition Formation Problem**) Given a set of human users and a set of tasks, the human coalition formation problem (HCFP) refers to the division of that set of human users into disjoint groups in such a way that the division optimizes the current-task rewards of all the human users over the *entire* set of tasks. Since at any given time, the set of future tasks and the set of human users in the environment are not necessarily known, the core of HCFP lies at the tradeoff between maximizing current-task rewards and maximizing future-task rewards when forming the coalitions.

2.2 Environment

In this section, we further define the environment through a set of assumptions, which are essential in addressing the unique challenges involving human coalition formation. We also define agent support, a novel component in a MAS where human users and user agents co-exist. With this definition, it allows us to consider the impact and cost of agent support to address user needs.

Assumption 7 (**Environment**) The MHCF environment $E = \langle H, U, T, A_c, S, EVM \rangle$ consists of a set of human users $H = \{h_1, \dots, h_{nh}\}$, a set of user agents $U = \{u_1, \dots, u_{nh}\}$, a set of tasks $T = \{T_1, \dots, T_{nt}\}$, a finite set of human user actions A_c that are required to solve the tasks, a system agent S that conducts the assignment of tasks and distributes rewards, and an evaluation metric function in the form $EVM: T_j \times \alpha \to R$ where $\alpha \subseteq A_c$ and $R \in \mathbb{R}$.

Assumption 8 (**Task**) The tasks in the MHCF environment are denoted by: $T_j = \langle st_j, tl_j, tac_j, tcc_j \rangle$, where st_j is the time the coalition can start working on the task, tl_j is the time limit within which the coalition must complete the task, tac_j is the actionset(s) required to solve this task, and $tcc_j \subseteq tac_j$ is the task actionset(s) that are required and completed. The required actionsets also identify the type of a task. Furthermore, we define the *degree* of type similarity between two tasks T_j and T_k as: $Sim_{tac}(T_j, T_k) = 1 - \left(|tac_j \backslash tac_k| + |tac_k \backslash tac_j|\right) / \left(|tac_j| + |tac_k|\right)$ (1)

According to Eq. 1, the similarity between the types of two tasks is equal to the ratio of the number of required actionsets that are common between them (calculated by taking the set difference operator \) and the total number of required actionsets they have.

Assumption 9 (**Assistant**) Each user agent u_i in the MHCF environment is assigned to a human user h_i to help that human user form coalitions and earn rewards.

Assumption 10 (**Modeling**) Each user agent u_i maintains a model $hm_{i,t}^i$ of its assigned human user h_i and maintains a vector of models $\boldsymbol{hm}_{s,t}^i$ of a subset of human users $H_s \subseteq H$. A human user model $hm_{i,t}^i$ is a tuple of attributes $\langle ha_{i,t,1}^i, \dots, ha_{i,t,na}^i \rangle$ where $ha_{i,t,k}^i \in Ha \; \forall k$.

Assumption 11 (**Coalition**) A coalition $C = \langle H_c, U_c, \alpha_c, \boldsymbol{d}_c, g \rangle$ is defined as a five-tuple containing a set of human users H_c, their respectively assigned user agents U_c, a joint coalitional actionset $\alpha_c = \{\alpha_i, \dots, \alpha_k\}$ for individual actions α_i agreed to be performed by human users $h_i \; \forall \alpha_i \in \alpha_c$ and $\forall h_i \in H_c$, and a demand vector $\boldsymbol{d}_c = \{d_1, \dots, d_{nd}\}$ consisting of agreed demands of the reward which is available upon completing the joint coalitional actionset α_c. The goal of the members of a coalition is to perform the joint coalitional actionset $\alpha_c = \{\alpha_1, \dots, \alpha_{nc}\}$ where $\alpha_i \in A_c \; \forall \, i$ to receive coalitional reward $EVM(T_c, \alpha_c) = R_c$. This coalitional reward is then divided proportionally among the members: $R_i = R_c \cdot d_i / \sum_{d_i \in D_c} d_i$ where d_i and R_i are the demand and the individual current-task reward of human user h_i for the task T_c. Finally, g is a group agent—its role to be specified in Section 2.4—that helps the user agents learn.

Assumption 12 (**Time**) We assume that the agents' and human users' actions occur in discrete time quanta $t_E = t_0, \ldots, t_\infty$. Furthermore, we assume that the tasks are assigned to the coalitions in a time line $\tau_T = \{\tau_0, \ldots, \tau_{nt}\}$. At any point of time τ_i there is a subset of tasks T_{τ_i} where $Sim_{tac}(T_j, T_k) > 0 \; \forall \, T_j, T_k \in T_{\tau_i}$. That means, at any given point of time, there are single instances of multiple tasks available for the human coalitions to solve. In this environment, each user agent tries to optimize the rewards for its assigned human user by finding the most reward-yielding task and the most reward-yielding human coalition for that task.

Definition 2 (**Agent Support**) We define **agent support** or **scaffolding** provided by a user agent u_i to its assigned human user h_i as a function $sc_i(hm_{i,t}^i, \alpha_i) = \{gct_1, \ldots, gct_{ic}\}$ where $gct_{i,t} \in SC \; \forall i$ while h_i is working in a coalition C, having agreed to perform action α_i where the user agent u_i supports its human user accordingly with action gct_i . We define the **cost of scaffolding** as: $Cost\left(sc_i(hm_{i,t}^i, \alpha_c)\right) \in \mathbb{R}$. Furthermore, we define the set of scaffolding provided by the user agents in a coalition C as: $SC_{c,t} = \{sc_i(hm_{i,t}^i, \alpha_i) | \forall u_i, h_i \in C\}$.

Assumption 13 (**State**) The states of the MHCF environment are denoted by $\omega = (CS, T_{CS}, \boldsymbol{d}_{CS}, \boldsymbol{\alpha}_{CS}, SC_{CS})$. Here CS is the coalition structure, $T_{CS} \subseteq T$ is the set of tasks assigned to the coalitions $C \in CS$, $\boldsymbol{d}_{CS}$ is the vector of demand vectors of the coalitions $C \in CS$, $\boldsymbol{\alpha}_{CS}$ is the vector of joint coalitional actionsets being carried out in coalitions $C \in CS$, and SC_{CS} is the set of scaffolding actions of the user agents u_i's working in the coalitions $C \in CS$. Furthermore, a state at time t is defined by: $\omega_t = (CS_t, T_{CS,t}, \boldsymbol{d}_{CS,t}, \boldsymbol{\alpha}_{CS,t}, SC_{CS,t})$.

Assumption 14 (**Observable State**) We assume that each agent is able to obtain *only* a partial view of the environment—specifically, it is able to obtain information regarding its own coalitions. The states of the MHCF environment that are observable by a user agent u_i at time t are denoted by $\omega_t^i = (C_t, T_{c,t}, \boldsymbol{d}_{c,t}, \alpha_{c,t}, SC_{c,t})$. Here C_t is the coalition, $T_{C,t} \in T$ is the set of assigned tasks, $\boldsymbol{d}_{C,t}$ is the demand vector, $\alpha_{C,t}$ is the joint coalitional actionset being carried out, and $SC_{C,t}$ is the set of scaffolding actions of the user agents in the coalition C_t. at time t.

2.3 Problem Definition

Definition 3 (**The Multiagent Human Coalition Formation (MHCF) Problem**): Given the environment $E = \langle H, U, T, A_c, S, EVM \rangle$, the MHCF problem refers to a partition of the set of human users into coalitions that optimizes the total reward earned by the human users over a set of tasks ,

$$\sum_{t \in \tau_T} \max_{CS} \sum_{h_i \in C_t \in CS_t} R_{i,j,t} \quad (2)$$

where $R_{i,j,t}$ is the reward of human user h_i while working in the coalition C_t to solve task T_j at time t. Solving this problem faces several challenges. First, the total number of coalitions possible for a set of human users H is 2^H. Therefore, finding the coalition structure that optimizes the total reward of the set of human users will be a computationally intractable problem [13]. Second, since the real-world systems are often open, it will be a waste of resources to compute the optimally rewarding coalitional structure for the entire set of tasks since the set of human users and the available tasks could change over time. Third, since human behavior changes over time, we argue that by making sequentially rational [2] decisions in forming coalitions, it is possible to achieve *near optimal* rewards for the human users over an entire set of tasks. Finally, due to uncertainty in modeling the human user behavior, it may not be possible to calculate the coalition structure that would maximize the reward of all the human users for any task.

Therefore, instead of finding an algorithm that achieves this goal of the human coalition formation problem but may be difficult to implement in a real-world system due to high computational complexity, we propose a modified version of the multiagent human coalition formation problem in which the *individual* agents try to make *sequentially rational* [2] decisions over a set of tasks to allow its assigned human user to join the highest-reward-yielding coalitions based on the information available to it during that round of coalition formation.

Definition 4 (**Multiagent Human Coalition Formation Problem for Individual Agents**) In the environment $E = \langle H, U, T, S, EVM \rangle$, the multiagent human coalition formation problem (from the perspective of an individual user agent u_i) refers to finding a C, identifying the necessary scaffolding $sc_i(hm_i^i, \alpha_i)$, and when considering action α_i at any state ω_t^i such that, the sum of the expected reward for joining the coalition for the current task T_j starting at time τ_j minus the cost of scaffolding the human user for the current task is maximized. This can be written as:

$$\max_C \max_{\alpha_i} \max_{sc_i} \left[R_{i,j,\tau_j + tl_j}(\omega_t^i) \right] \quad (3)$$

where, the total expected reward $R_{i,k,t'}(\omega_t^i)$ at some state ω_t^i is defined as:

$$R_{i,j,\tau_j+tl_j}(\omega_t^i) = Rct_{i,j,\tau_j+tl_j}(\omega_t^i) + \sum_{t', T_k \subseteq T} Rft_{i,k,t'}(\omega_t^i) - Cost\left(sc_i(hm_{i,t}^i, \alpha_i)\right) \quad (4)$$

where $t' = \tau_k + tl_k$, $Rct_{i,j,\tau_j+tl_j}(\omega_t^i)$ is the expected current-task reward of the human user h_i for task T_j assigned at time τ_j and whose reward is available at time $\tau_j + tl_j$. Furthermore, $Rft_{i,k,t'}(\omega_t^i)$ is the *expected* future-task reward for future tasks T_k that would be available to the human coalitions at time τ_k and whose reward would be available at time $t' \in \tau_T + tl_k$. Finally, $Cost\left(sc_i(hm_{i,t}^i, \alpha_i)\right)$ is the cost incurred by the user agent u_i for providing scaffolding to the human user h_i for contributing individual action α_i to the joint coalitional action α_c.

3 Design of the MHCF Framework

Here we describe two key components of the MHCF framework: negotiation and learning. The agents use negotiation to form coalitions. Each agent also learns the model of its assigned human user and those of other relevant human users, and the probabilities for the impact of agent-provided scaffolding.

3.1 Coalition Formation by Negotiation

Our proposed negotiation-based coalition formation method consists of three main stages: *initialization*, *negotiation*, and *finalization*.

Stage 1 (**Initialization**) In the initialization stage, the system agent communicates with all the user agents and announces the tasks the human users need to complete in the upcoming round of coalition formation.

Stage 2 (**Negotiation**) The negotiation is a three-step process that is carried out in rounds. These steps are: *proposition*, *consideration*, and *notification*. Next we describe these steps in details.

1. *Proposition*: We assume that at the beginning of the negotiation process, each user agent is in a singleton coalition. Then at each round of the negotiation process, one agent is randomly selected (with equal probability) to be a proposer who acts as the initiator of the negotiation process Once the negotiation process starts, the chosen proposer agent has the following options: (1) it can choose to stay in its own coalition with everything unchanged, i.e., relinquish its turn, (2) it can propose to modify its own coalition with a renewed demand**,** and (3) it can propose to form a new coalition with a new proposal with a new set of coalition members and a new demand. In our negotiation protocol, we assume that all agents, even those that have already formed a coalition, participate throughout the entire negotiation period. Furthermore, some agents may be in a singleton coalition during the negotiation rounds and it may take several rounds of negotiation for some agents to give up its singleton coalition and join some other large coalition. Finally, the proposer proposes to all of its potential coalition members at the same time. The proposal of a user agent is defined as the following:

Definition 5 (**Proposal**) The proposal of a user agent u_i to user agent u_j at time t is a 5-tuple $ps_{i,j,t} = \langle \boldsymbol{hm}^i_{c,t}, d_{j,t}, \alpha_i \in \alpha_c, sc_i(hm^i_{i,t}, \alpha_i)\rangle$ made by u_i to form coalition $C_t \in CS_t$, $\boldsymbol{hm}^i_{c,t}$ as the vector of human user models (without user agent/human user identities) that will be in the proposed coalition as viewed by user agent u_i, $d_{j,t}$ as the demand of reward for h_j determined by u_i "on behalf of" u_j respectively, α_c as the joint coalitional actionset of the coalition $C_t \in CS_t$ in which α_i is the action that h_i is going to contribute to the coalition, and $sc_i(hm^i_{i,t}, \alpha_i)$ is the scaffolding provided by u_i for its assigned human user h_i with user model $hm^i_{i,t}$.

Definition 6 (**Value of a Coalition**) The value of a coalition from the perspective of an agent u_i assigned to human user h_i is:

$$V^i_t(C_t|\boldsymbol{hm}^i_{c,t}) = \max_{\alpha\in A_c} \sum_{t\in\tau, T_j\in T} \Pr(\alpha_c|\boldsymbol{hm}^i_{c,t}, SC_{c,t})\Pr(\omega^i_t|\alpha_c) R_{i,j,t}(\omega^i_t) \quad (5)$$

So, the value of a coalition is the maximum sum of the expected reward it can generate over all future states in which task reward can be collected. The value of the expected reward is calculated by multiplying: (1) the probability of performing joint coalitional actionset for a given vector of human user models and the scaffolding agreed upon by the user agents to be provided to that group, (2) the probability of reaching a state after taking a joint coalitional actionset, and (3) the total reward available to the user agent at that state.

Furthermore, the proposer's choice of the demand (or offer from the proposer's perspective) for the responder, i.e., d_j , depends on the proposer agent's estimate of the value of the human user represented by the responder in the coalition, and is calculated as the following.

Definition 7 (**Value of a Human User**) The value of a human user h_j in the coalition C_t as calculated by the user agent u_i as:

$$V^i_t(hm^i_{j,t}|C_t) = \left[V^i_t(C_t|\boldsymbol{hm}^i_{c,t}) - V^i_t\left(C_t|\boldsymbol{hm}^i_{c\setminus\{h_j\},t}\right)\right] / V^i_t(C_t|\boldsymbol{hm}^i_{c,t}) \quad (6)$$

So, the value of the human user in the coalition is found by calculating the difference between the value of the coalition with that human user and the value of the coalition without that human user and then dividing that difference with the value of the whole coalition.

Assumption 15 (**Reward Valuation**) The human users are heterogeneous with respect to their valuation of current-task and future-task.

Definition 8 (**Total Reward**) The total reward for joining a coalition from the perspective of an agent u_i assigned to human user h_i for solving some task T_j in a coalition C is:

$$R^i_{i,j,t}(\omega^i_t) = vct^i_{i,j,t}(\omega^i_t) \cdot Rct^i_{i,j,t}(\omega^i_t) + vft^i_{i,j,t}(\omega^i_t) \cdot Rft^i_{i,j,t}(\omega^i_t) \quad (7)$$

According to Eq. 7, the total reward available to a human user at a state is the sum of current-task and future-task rewards weighted by its human user's valuations of those rewards.

Furthermore, the current-task reward in Eq. 7 is calculated as,

$$Rct^i_{i,j,t}(\omega^i_t) = d_i \cdot Rct^i_c(\omega^i_t) / \sum_{h_i\in H_c} d_i \quad (8)$$

where $d_i \in \boldsymbol{d}_c$ is the demand of the human user h_i (*Assumption* 11) for the coalitional reward. So, a human user's share of the current-task reward is calculated by dividing his or her demand with the total demand of the entire coalition which the coalition members agreed upon (*Assumption* 11) during the negotiation stage. Next, the future-task reward is defined as:

$$Rft^i_{i,j,t}(\omega^i_t) = \sum_{T_j\in T, t''>t'>t\in\tau} \Pr\left(hm'^i_{i,t'}|\omega^i_{\tau_j}\right) \Pr\left(\omega^i_{t''}|hm'^i_{i,t'}\right) Rct^i_{i,j,t}(\omega^i_{t''}) \quad (9)$$

Here, the future-task reward is calculated by multiplying: (1) the probability of the model of its human user being changed in state $\omega^i_{\tau_j}$ after completing the current task, (2) the probability of reaching future states due to that changed human user model, and (3) the current-task rewards. The inspiration of describing the future-task reward using Eq. 7 comes from *Assumptions* 3 and 4. We assume that human users are able to learn from their experiences of working in coalitions and are able to modify their behavior (e.g., acquire new capabilities) as a result of that learning. This future-task reward is designed to estimate how much reward in future a human user may earn from his or her experience of working in this coalition.

2. ***Consideration*:** Once approached by a proposer in the proposition stage, the responding agent has three types of replies: (1) accept the proposal as is, (2) reject the proposal outright, and (3) provide a counter-proposal to the proposer agent with the same coalition members but with an increased demand for itself.

The criterion for accepting a proposal is the following. A responding agent would only accept a proposal iff the value of the proposed coalition (Eq. 5) is strictly greater than the value of its current coalition. In other words, a responding agent u_j currently in the coalition C_t would switch coalition for a new proposal $ps_{i,j,t} = \langle id_i, \boldsymbol{hm}^{i}_{c',t}, d_{j,t}, \alpha_i \in \alpha_c, sc_i(hm^i_{i,t}, \alpha_i)\rangle$ *iff* $V^j_t\left(C'_t|\boldsymbol{hm}^j_{c',t}\right) > V^j_t\left(C_t|\boldsymbol{hm}^j_{c,t}\right)$.

If the responding user agent decides to revise the proposal, it will send out a counter-proposal to the proposer agent. We define the counter-proposal as the following:

Definition 9 (**Counter-Proposal**) The counter-proposal of a user agent u_j to user agent u_i is a 4-tuple $ps'_{j,i,t} = \langle hm^j_{j,t}, d_{j,t}, \alpha_j \in \alpha_c, sc_j\left(hm^j_{j,t}, \alpha_j\right), ps_{i,j,t_0}\rangle$ made by u_j to revise the proposal posted by u_i to form coalition $C_t \in CS_t$, d_j as the demand of reward for h_j determined by u_j. Notice that this new demand $d_{j,t} \propto V^j_t\left(hm^j_{j,t}|C_t\right)$ is calculated by the responding agent using Eq. 6. Furthermore, α_c is the joint coalitional actionset of the coalition $C_t \in CS_t$, α_j is the human user h_j's contribution to the joint coalitional actionset α_c, $sc_j\left(hm^j_{j,t}, \alpha_j\right)$ is the scaffolding provided by the user agent u_j to its assigned human user h_j, and ps_{i,j,t_0} is the original proposal put forth by the user agent u_i at time $t_o < t$.

3. ***Notification***: Once the proposer agent receives the counter-proposal, it first updates its own model of the human user represented by the responding agent from the information in the counter-proposal using the following formula.

$$ha^i_{j,t,k} \leftarrow \left(w_o \cdot ha^i_{j,t,k} + w_n \cdot ha^j_{j,t,k}\right)/(w_o + w_n)\ \forall ha^i_{j,t,k} \in hm^i_{j,t} \quad (10)$$

Notice that in Eq. 10, the proposer updates each attribute of its assigned human user by finding an weighted average of the: (1) the old value calculated by the proposer and (2) the new value described by the responder. This model exchange acts as a learning mechanism for all negotiating agents allowing them to update their models of the human users they would like to form coalitions with. Furthermore, the weights allow the proposer to balance the emphasis on its own model of a human user and the modeling of that human user by the responder agent. According to *Assumption* 4, the model of a human user changes over time due to his or her learning as he or she participates in joint coalitional actions. Furthermore, we assume that the user agents are able to observe the actions and behavior of the human users in its own coalition only. As a result, one of the two following situations may occur: (1) the proposer's assigned human user may not have ever been in a coalition with the responder's human user, i.e., the proposer does not have a model of the responder's human user; or (2) the proposer has an outdated model of a responder's human user.

Once the model is updated, the proposer has the following options: (1) accept the counter-proposal or (2) reject the counter-proposal. Whether the proposer accepts the counter-proposal depends on: (1) the discrepancy between the proposer and the responding agent's model of the responding agent's assigned human user (i.e., the difference between $hm^i_{j,t}$ and $hm^j_{j,t}$) and (2) the human users represented by the user agent in the coalition. We define the discrepancy between the models of a human user prepared by two different user agents as:

Definition 10 (**Model Discrepancy**) The discrepancy between the models of a human user h_i as prepared by the two different user agents u_i and u_j is:

$$MD\left(hm^i_{i,t}, hm^j_{i,t}\right) = \sum_{i=1,\ldots,|hm^i_i|} w_i \cdot CD\left(ha^i_{i,t}, ha^j_{i,t}\right) \quad (11)$$

where, $CD\left(ha^i_i, ha^j_i\right) \propto \left|ha^i_i - ha^j_i\right|$ denotes the difference between the attribute values modeled by the user agents u_i and u_j, i.e., difference between the values of ha^i_i and ha^j_i and w_is are weights since not all attribute difference would impact the performance of the human user in a coalition in the same way.

Once the model is updated, the proposer agent would accept the counter-proposal $ps'_{j,i,t}$ and revise its own proposal $ps_{i,j,t}$, iff,

$$V^i_t\left(hm^i_{j,t}|C_t\right) > d_{j,t}/\sum_{h_i \in C} d_i \in ps'_{j,i,t} \quad (11)$$

In Eq. 11, $V^i_t\left(hm^i_{j,t}|C_t\right)$ is calculated using Eq. 6. So, according to our reasoning, the responding agent's revised demand would be accepted by the proposer if that updated demand is less than the responding user agent's value in the coalition calculated with the proposer's *updated* model.

Once the proposer agent accepts or rejects the counter-proposal, it will communicate its decision to the responding agent and complete the negotiation round. If *all* of the user agents to whom the proposer proposed to, agrees to join its coalition, the proposer then forms the coalition and sends out the coalition formation message: $M_{cf} = \langle \boldsymbol{hm}^i_{c,t}, \boldsymbol{d}_c, \alpha_c\rangle$ to all of the proposed user agents.

Stage 3 (**Finalization**) In the finalization stage, the system agent signals the user agents to stop the negotiation rounds and assigns tasks to the formed coalitions. Once the tasks are assigned, the proposer agents send out a coalition revelation message: $M_{ir} = \langle U_i, \{\langle u_i, h_i, \alpha_i\rangle | h_i \in H_c, \alpha_i \in \alpha_c\}\rangle$ which reveals the identity of the members of the coalition (i.e., U_is) and the task assignments (i.e., $\langle u_i, h_i, \alpha_i\rangle$) to the member user agents.

3.2 Learning for Coalition Formation

The learning aspect of the MHCF framework is complex. First, the human users learn from their experience working in coalitions. Second, the agents learn about modeling their human users more accurately. Driving an agent's learning is its goal of selecting the right coalition, identifying the right task, and providing the right scaffolding for its assigned human user to achieve maximum total reward (per Def. 4, Eq.3):

Definition 11 (**Learning Goal**) The learning goal of an agent in the MHCF environment consists of (1) learning the model of its assigned human user, (2) learning the models of a subset of other human users with which its assigned human user may form coalition, and (3) learning the probabilities that capture the impact on the model (or type) of a human user due to the user-agent-provided scaffolding on: (a) completing the current task and earning current-task rewards and (b) improving the human user's model by learning and earning future-task rewards.

Our solution to the learning problem utilizes both model-based and cooperative (or divided [3]) learning. Using the model-based learning, the user agents learn the values of the different attributes of a given model of its assigned human user and the other human users in the MHCF environment through interactions. Each user agent, based on its own experience and through experience sharing with other agents, learn a model of the environment that (1) describes the generic dynamics of a human coalition formation environment and (2) links the task and the human user's models with their rewards (both current-task and future-task ones).

Learning Assigned Human User's Model.

Assumption 16 (**Learning Attribute Values through Observation**) The user agents are able to observe/learn the current state of its assigned human user's model periodically. For example, these observations can be in the form of psychometric tests or surveys that are performed by the user agents on their human users.

Assumption 17 (**Finite Attribute Categorization**) Each attribute regarding the human user behavior can be divided into a finite set of categories.

Here, finite attribute categorization (which is common in various psychometric tests) allows the user agents to overcome the so-called *continuous state problem* because, when every attribute in the human user model is divided into a finite set of categories, the total number of observable states (*Assumption 14*) for the user agents is reduced to a significantly smaller finite number. This reduction of the total number of observable states helps the user agent learn the probability values in the environment better which in turn helps their reasoning process while forming coalitions.

Learning Other Human Users' Models. As discussed earlier, the user agents have to learn other potential human user's models and they have to update those learned models since a user agent is able to observe only its assigned human user's model (*Assumption 16*). One way the user agents can learn the models of a subset of other human users is through the negotiation process. Notice that during the counter-proposal stage (Section 2.4) of the negotiation process, the responder provides the user agent an updated model of its assigned human user. Using this negotiation mechanism, the proposer agents can update their models of the subset of potential coalition members when those members' user agents act as the responders in the negotiation process.

Learning the Probabilistic Model of the Environment. In a way, the user agents learn the probabilistic model of the MHCF environment so that the user agents share the burden of learning among themselves. With this approach of cooperative learning, two questions arise: (1) how is the probabilistic model regarding the MHCF environment represented by individual agents? and (2) how do the agents share their learned probability values with each other? For modeling the probabilities of the environment, the user agents use *Bayesian networks*, and for sharing the learned probability values, the user agents use *group-agent-mediated cooperative learning*. A Bayesian network (BN) is a triplet $\langle V, G, \wp \rangle$ where V is a set of variables, G is a connected DAG whose nodes correspond one-to-one to members of V such that each variable is conditionally *independent* of its *non-descendants* given its parents, and $\wp$ is a set of probability distributions. Say the parent of nodes $v \in V$ is denoted by $\pi(v)$, then, $\wp = \{P(v|\pi(v))|\forall v \in V\}$ (15) and $P(V) = \Pi_{v \in V} P(v|\pi(v))$ (16). Figure 1 shows the BN structure of the MHCF environment as used by the user agents.

Fig. 1 shows the nodes and the edges that represent the generic structure of the BNs that is used by the user agents in the MHCF environment. Notice that Fig. 1 describes the BN of user agent u_i who is assigned to human user h_i with model $hm_{i,t}^{i}$ working in coalition C with other human users with model $\boldsymbol{hm}_{c,t}^{i}$ where the user agents in the group are providing scaffolding $SC_{c,t}$. The BN in Fig. 1 would allow the user agents to determine, probabilities like $\Pr(\alpha_c|\boldsymbol{hm}_{c,t}^{i}, SC_{c,t})$ (Eq. 5, 6) and $\Pr\left(hm'^{i}_{i,t'}|\omega_{\tau_j}^{i}\right)$ (Eq. 7, 9).

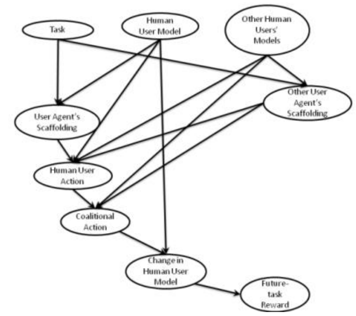

Fig. 1: Descriptive BN Structure of the MHCF Environment

Group-Agent-Mediated Cooperative Learning of BN. We use group agents to mediate cooperative learning within each coalition of agents. While the human users in a coalition are working, the user agents observe the actions of their assigned human users and update their own BNs. When the coalition's task is over and the rewards are distributed, the user agents upload their own BNs to the group agent assigned to their respective groups. Once the group agents have collected the BN from their user agents, they communicate among themselves and use the BN-Update algorithm (communication complexity $O(n_g)$), as shown in Table 1, to create an updated BN. The group agents then send the updated BN to each of the user agents in their group. Finally, the user agents combine the updated BN with their existing BN probability values to update their own BNs.

Note that with the above cooperative learning mechanism, there are the following advantages: (1) the user agents need to *communicate only twice* with the group agents, (2) the user agents have to *reveal only a small subset of probabilities* to the group agents, and (3) given a heterogeneous environment with different types of human users, tasks, and scaffolding, the user agents are able to *learn the probabilities faster and more accurately* together than they would have learned individually.

Table 1: BN-Update for Group Agent g_i

BN-Update for Group Agent g_i
1. Initialize: $P \leftarrow \phi, P_u \leftarrow \phi, P_o \leftarrow \phi$
2.Collect BN probability set $P = \{P_i, \dots, P_k\}$ from user agents $\{u_i, \dots, u_k\} \in G_i$
3.Unify the collected probabilities: $P_u \leftarrow UnifyBN(P)$
4.If received probability P_o' from group agent G_o
$P_o \leftarrow AssimilateBN(P_o, P_o')$
5.If available, collect BN token and group agent list
$G_c = \{g_1, \dots, g_k\}$
a. For each group agent $G_c = \{g_1, \dots, g_k\}$, Send P_u
b. Send BN token and $G_c \backslash \{g_k, g_i\}$ to g_k
6. Else If received BN token and group agent list G_c from g_{k+1}
a. For each group agent $G_c = \{g_1, \dots, g_k\}$, Send P_u
b. Send BN token and $G_c \backslash \{g_k\}$ to g_k
7. Update BN probability using $P_o \leftarrow UpdateBN(P_u, P_o)$
8. Send BN probability set P_o to user agents $\{u_i, \dots, u_k\}$

The key functions in Table 1 can be defined as follows:

$$UnifyBN(P_o, P_o')$$
$$= \left[\sum_{\forall P_i \in P} \gamma_i \cdot \Pr_i(x|y,z)\right] / \left[\sum_i \gamma_i\right] \ \ \forall x, y, z \in BN \ (17)$$

where γ_i is the number of observations that occurred in the group agent g_i's group (i.e., the group agent g_i's confidence on the calculated probability) and $\Pr_i(x|y,z)$ are the various probabilities measured by the user agents $u_i \in G_i \backslash \{u_i\}$.

$$AssimilateBN(P_o, P_o') = [\Pr_o(x|y,z) + \Pr_o'(x|y,z)]/2 \ \forall x,y,z \in BN \quad (18)$$

where $\Pr_o'(x|y,z)$ are the various probabilities measured by the user agents $u_o \in G_i \backslash \{u_i\}$

$$UpdateBN(P_u, P_o) = [\gamma \cdot \Pr_u(x|y,z) + (1-\gamma)\Pr_o(x|y,z)]/2 \ \forall x,y,z \in BN \quad (19)$$

3.3 Novelty of MHCF Design

The novelty of the MHCF framework lies in the design of its negotiation protocol and learning strategy. First, our design of the negotiation protocol allows the user agents to balance the current-task (estimated from their model) and future-task rewards (estimated from their model and the support of their agents) while forming coalitions. This novel design allows the user agents to *exploit* the learning capability of the human users to form coalitions that *improves* the model of human users while solving the current task. Second, unlike the typical multiagent coalition formation protocols, the negotiation protocol in MHCF allows the user agents to take the agent scaffolding into account while estimating the value of a human user or a coalition. This aspect of the design of the negotiation protocol allows the user agents to: (1) improve the current-task and future-task rewards of its assigned human user by capitalizing on the human learning and (2) more accurately judge the value of a coalition by considering the scaffolding provided by the agents of other human users, and (3) reason by balancing the cost and benefit of providing the scaffolding to the human users. Finally, the probabilistic calculation of the current-task reward and future-task reward allows the user agents to incorporate the openness of the environment and the learning capability of human users of different models to better estimate the value of a human user and that or a coalition.

Second, the novelty in the learning process of the user agents in the MHCF environment lies in the following. First, a model-based cooperative learning strategy allows the user agents to estimate the values of a human user and a coalition more accurately and at a faster rate than that of a regular reinforcement learning strategy. Second, the design of our model-based learning allows the user agents to take three important elements: (1) the model of a human user, (2) the support provided by his or her user agent, and (3) the models and the scaffolding of his or her group members into account while calculating the current-task and future-task rewards of a human coalition. Since the typical multiagent framework does not consider these three elements, it would be difficult for their agents to accurately estimate the short-term and long-term rewards available for a human user in a coalition.

Due to these novel approaches, the MHCF framework is better suited to solve the problem of forming human coalitions than are the typical multiagent coalition formation frameworks.

4. IMPLEMENTATION

To test the effectiveness of using MHCF in forming human coalition, we have adopted SimCoL [4]—a *multiagent application for simulating the collaborative learning of a set of students in the Computer-Supported Collaborative Learning (CSCL) environment*. SimCoL contains agents that act as the students collaborating in a CSCL classroom and an agent who act as the teacher of that classroom. In our implementation of MHCF in SimCoL, we have added agents to provide support to: (1) the students (student agents), (2) the teacher (teacher agent), and (3) the groups (group agents). In our implementation: (1) the student agents in SimCoL act as the user agents, (2) the teacher agent in SimCoL acts as the system agent, and (3) the group agent acts as the group agent. The tasks in SimCoL are typical classroom problems that contain several subtasks and those subtasks have difficulty values assigned by the teacher. Each student in SimCoL has a model that contains: knowledge, ability, motivation, emotion, and social relationship with others. The student agents in MHCF used the ability (calculated in SimCoL using knowledge, motivation, and emotion) and social-relationship of the students to build the student model. Furthermore, we have provided BN (like Fig. 1) to the student agents in SimCoL which contains the probability values they exchange with other student agents. Equipped with the student models and the BN, the student agents negotiate with other student agents to choose a subtask that is suitable for their assigned students. Once the coalitions are formed, the students in SimCoL collaborate with their peers to solve their assigned subtasks while the student agents monitor their activities and provide task-dependent scaffolding. This scaffolding improves the recipient student's expertise and helps him or her solve the assigned subtask better. When the collaborative session is over, the teacher agent rewards each student a score that is proportional to the sum of the difficulty of the assigned subtask and the final solution quality of that subtask.

5. EXPERIMENTS AND RESULTS

To test the effectiveness of MHCF framework, we have randomly generated 99 student models in SimCoL and 50 tasks where each task contained 3 different subtasks. Furthermore, $< 50\%$ of the total set of subtasks contained similar concepts and were randomly distributed (*not known* to the student agents forming the coalitions) among the 50 tasks. Then MHCF framework was used to form three-member coalitions for each task and the students collaborated to solve the task and earn rewards. We had two main foci of our investigation: first, we investigated the ability of the MHCF framework of forming student coalitions that are able to collaborate and earn rewards in the environment. Second, we investigated the MHCF framework's ability to exploit human learning and improve the total reward over the entire set of tasks. Fig.2 shows the average knowledge gain of the students for the 50 tasks they solved.

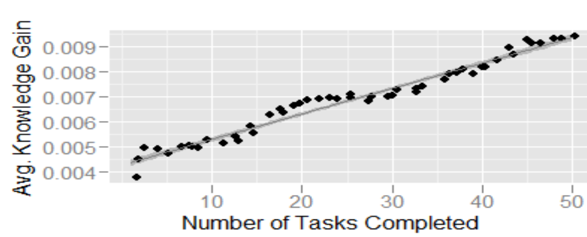

Fig. 2: Average Knowledge Gain over Time

The trend in Fig. 2 shows that over time, the students were able to gain more knowledge per classroom session. Furthermore, a similar trend is observed when the total reward earned by the students is plotted against the number of tasks completed. The outcome of the student coalitions consists of the knowledge and reward gain of the students. Furthermore, the setup of SimCoL dictates that student coalitions may gain knowledge and earn rewards when their members have high ability difference and high social relationship values. When this is not the case, the collaboration process among the members may break down and they may fail to earn reward or gain knowledge. So, our results imply that the MHCF framework was able to form student coalitions that lead to successful collaborations and the entire set of students were able to improve their knowledge and earn reward while solving the assigned tasks.

We also investigated the MHCF framework's ability to exploit the student learning to improve the total rewards earned over the entire set of tasks. The results show that over time: (1) *the standard deviation of the rewards earned by the student coalitions decreased* (slope = $-0.22, p < 0.01$) and (2) *the reward earned per interaction increased* (slope = $1.56, p < 0.01$). The reward (i.e., the score) of a student is proportional to the sum of (1) the difficulty of the subtask and (2) the solution quality achieved by the students who are solving it. Since the student agents in our implementation try to improve the total reward earned over the entire set of tasks, initially, some of the coalitions will contain students who may fail to attain high solution quality for the subtasks since they or their peers would have low ability. As a result, the members of those coalitions would have low rewards yielding (1) higher standard deviation of rewards and (2) low reward per interaction value for the entire classroom respectively. However, that tradeoff would improve the total reward earned by the entire classroom for future tasks since the members of those coalitions have improved their abilities and can now utilize that learned skill to earn better rewards. As a result, there would be fewer coalitions with low rewards and (1) the standard deviation of the earned reward would decrease and (2) the reward gain per interaction would increase, as indicated by the observed slopes above.

To further validate the MHCF framework's ability of improving the total reward over time, we have replicated the same experiment with randomly formed coalitions. The *statistically significant* results of our comparison experiment shows that (1) *MHCF-formed coalitions earn more rewards* ($p < 0.01$) *over time*, (2) *MHCF-formed coalitions reduce the standard deviation of the earned reward faster* (slope -0.22 vs. $0.01, p < 0.01$), and (3) *MHCF-formed coalitions improve the reward per interaction faster* (slope 1.56 vs. $-1.96, p < 0.01$).

In summary, our analysis of the results indicate that the MHCF framework is able to form student coalitions in such a way that the members of those coalitions achieve better rewards over time while solving the current task. Further, the MHCF framework is able to exploit the learning ability of the participating students and improve their total reward earned over a period of time.

6. RELATED WORK

Although most the typical multiagent coalition formation frameworks do not explicitly discuss human coalition formation problems, some of the researchers have discussed forming coalitions in the real-world uncertain environments which is relevant to the human coalition formation problem environment. Chalkiadakis and Boutilier [1] describe a negotiation-based algorithm and Bayesian learning to form agent coalitions in uncertain environments. Their research is very relevant to ours since it describes the coalition formation process where agents lack sufficient knowledge of the capabilities of their partner agents and the value of the coalitions is not known with certainty. In another research, Kraus, Shehory, and Taase [5] describe a negotiation-based coalition formation protocol for forming agent coalitions in real world uncertain environments. In their research, agents use simple heuristics to negotiate coalitions that allows them to solve complex tasks in the business domain under time constraints. However, the coalition formation methods described these research approaches are designed to be used in scenarios where the principal actors (i.e., who solve the tasks) are agents and these methodologies do not address the unique aspects of the human coalition formation problem. For example, these coalition formation methodologies are not designed to exploit the behavioral improvement of the coalition members and coalition formation methodologies do not calculate the long-term rewards as a function of the environment both of which are common in human coalitions. Finally, these coalition formation methods do not consider the agent support to the actors while forming the coalitions which could be an important factor in determining the short-term and long-term rewards in a human coalition formation problem.

7. CONCLUSIONS & FUTURE WORK

In this paper, we have described the design and implementation of MHCF, a multiagent framework that considers comprehensively (1) the roles of coalition structures and agent-provided scaffolding in facilitating human learning, and (2) the tradeoffs between maximizing current-task and future-tasks rewards taking into account environment openness and human changes, to form human coalitions. In our MHCF design, each human user is assigned a user agent that conducts negotiations and cooperative agent learning to form human coalitions. To validate the MHCF framework, we have used it to form coalitions in SimCoL, an environment that simulates the collaborative learning activities of students. Our results show that the agents in the MHCF framework can: (1) learn to form better coalitions over time and (2) enable the students to earn better rewards for a given set of tasks.

The problem of forming human coalitions is a relatively new research arena and there are a lot of open issues to address. Our future work includes: (1) improving the MHCF framework by implementing adaptive scaffolding and incorporating the cost of scaffolding and (2) conducting a large-scale real-world classroom experiment with the MHCF framework.

8. ACKNOWLEDGMENTS

The work reported here was partially supported by NSF Grant DMI-044129 and Microsoft Research.

9. REFERENCES

[1] Chalkiadakis, G. and Boutilier, C. 2008. Sequential Decision Making in Repeated Coalition Formation under Uncertainty, Proc. AAMAS.

[2] Sandholm, T., Larson, K., Andersson, M., Shehory, O., and Tohme, F. 1999. Coalition structure generation with worst case guarantees. J. Artificial Intelligence. 111(1999), 209-238.

[3] Weiss, G. and Dillenbourg, P. 1998. What is 'multi' in multi-agent learning? Collaborative learning. Cognitive and computational approaches, Pergamon Press, 1998, pp. 64-80.

[4] Khandaker, N. and Soh, L-K. 2008. On Incorporating Learning Theories to Simulate a Computer-Supported Collaborative Learning Environment. Department of Computer Science and Engineering. University of Nebraska-Lincoln.

[5] Kraus, S., Shehory, O., and Taase, G. 2003. Coalition formation with uncertain heterogeneous information. Proc. AAMAS, pp. 1-8.