Multi User Learning Agent based on Social Interaction

D. Katagami
IGSSE
Tokyo Institute
of Technology
Yokohama, 226-8502

H. Ohmura
Advanced System
Development Center
Yamaha Corporation
Shizuoka, 438-0192

Y. Yasumura
Factory of
Engineering
Kobe University
Kobe, 657-8501

K. Nitta
IGSSE
Tokyo Institute
of Technology
Yokohama, 226-8502

Abstract

In this paper, we propose Multi User Learning Agent(MULA) to interact with various types of people effectively towards realization of Social Interaction. MULA is equipped with two learning functions for Social Inteaction. One is direct learning function using individualization of the user parameters, and another is indirect learning function using past experience and the similarity between each users. MULA adopts an extended classifier system which uses a classifier strength and a user profile for each people. Each person's classifier strength is enhanced by the interaction with the person and the past experience with similar type of people. We verified the effect of MULA by experimenting in interaction with six participants using a pet-type robot AIBO.

1 Introduction

In recent years, research of HAI (Human-Agent Interaction)[1] aiming at forming communication between an agent, such as a robot, and human is capturing the spotlight. However, the present HAI research is in the stage of a designing one-to-one adaptations between human and an agent, and the methodology with two or more users is not established. In this research, we aim at realizing a social interaction performed by the social learning which learns from interaction with two or more users using a constructive approach. Social Intearction is realized by two learning functions. One is direct learning function using individualization of user parameters, and another is indirect learning function using past ecperience and the similarity between each users. In particular, indirect learning uses experience with other users for the target user. We think it is one form of social learning to realize the suitable action from experience with two or more users to the current user. We named an agent equipped with Social Interaction as Multi User Learning Agent (MULA). MULA adopts an extended classifier system which uses a classifier strength and a user profile for each people. We veryfied MULA in experiments of a user's preference prediction task in a real world.

2 Implementation of MULA

2.1 A learning based on past experience

A target here is to realize an adaptation to the current user by using experience with another users. Then, we realized MULA which incorporated past experience using the function explained based on the model-based reinforcement learning (Dyna-Q algorithm [2]). In this paper, we define changing the action by the influence of the others as "social".

[Direct learning by individualization of user parameters]
Although the technique [3] of preparing knowledge and the rule set for each environment is common, the problem of efficiency, such as a problem of a memory and dignity attachment of the knowledge, occurs. Then, we adopt the approach individually prepared for every user about how to use knowledge (user parameters) instead of using two or more rule sets.

[Indirect learning using past experience]
Since it corresponds to two or more users, it is important to use past experience well. Here, the similarity between users is used like *collaborative filtering*. Especially this may become the indicator for the next interaction in a stage with little experience with a current user.

2.2 Social leraning using a interaction with users

In order to realize above two functions we propose MULA architecture (Fig. 1) and it's learning algorithm (Fig. 2).

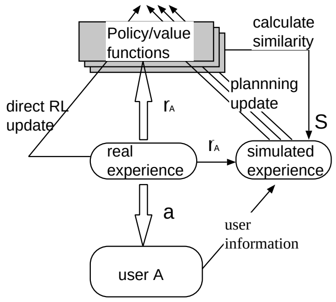

Figure 1: MULA architecture

Initialize $p(s, a, i)$ and $Model(s, a, i)$ for all s, a, i
Do forever
  user identification $i \in I$
  simple Q-learning
  $UserProfileR_i(s, a, i) \leftarrow r_i$
  for $j = 0$ *to* $j = n$ *do*
    Similarity $S_{ij} \leftarrow UserProfile\ R_i,\ R_j$
    $q(s, a, j) \leftarrow q(s, a, j) +$
      $\alpha S_{ij}(|r_i + \gamma max_{a'} q(s', a', i) - q(s, a, i)|)$

Figure 2: MULA algorithm

MULA specifies a user first. MULA updates directly individual user parameters of the user (it considers as User A temporarily) using simple Q-learning from an interaction (MULA receives remuneration r_A to act output a_A to certain state s_A) with the user. It is actual experience. Next, MULA calculates the similarity S_{AX} of $UserProfile$ between user A and another users X using users' information. MULA updates each $UserProfile$ and individual variables of all users indirectly using the interactive information $s_A, a_A, and\ r_A$ and this similarity with user A. It is indirectly experience.

2.3 The outline of the system

Fig.3 shows the outline figure of the system proposed by this research. This system has adopted XCS[4] as a fundamental learning mechanism, which is one of most useful classifier systems. XCS uses new three parameters instead of strength of classifier system for each classifier.
prediction$p$**:** prediction of classifier.

$$p_t = p_{t-1} + \alpha(|r_{t-1} + \gamma max_{a'} p_{t-1}(s', a') - p_{t-1}|) \qquad (1)$$

prediction errorε**:** An error between prediction p and actual measurement.

Fitness$F$**:** It adopts as an evaluation value of the rule creation by GA.
[individual variables]
This system prepares the variables of a classifier (a prediction value p, a prediction error ε, and the degree of adaptation F), without preparing a new rule set for every user. We call it an individual prediction value, an individual prediction error, and an individual fitness, respectively. Moreover, we prepared **the individual evaluation value** R which records evaluation of the past interaction. Each classifier has an individual evaluation value, for each user. When positive evaluation is received, the value increases, and when negative evaluation is received, it decreases. We call these values **individual variables**.
[The procedure of the system]
MULA updates individual evaluation value R based on evaluation of a user's interaction, and the similarity is calculated by it each time. MULA updates individual variables $p,\ \varepsilon,\ and\ F$ other than an individual evaluation value R like the usual XCS does. However, in individual prediction value p, although individual prediction value p_A is directly updated according to the similarity between user A and other users. All of other users' individual prediction values p_X are also updated indirectly. The method how MULA learns the individual prediction value p of a classifier is updated by the interaction with users is explained in Fig.3.

(1) MULA acquires sensor information S_A (1001) by interacting with user A.

(2) MULA generates Match Set $[M]$ from sensor information. And the action a chosen from an action selection method is outputted.

(3) MULA gets evaluation for the action from user A, and updates prediction value p and other individual variables directly.

(4) MULA calculates all of the similarities S_{AX} from user A's updated profile $\mathbf{R}_A$ and other users' profiles $\mathbf{R}_X$.

(5) MULA updates all other users' inidividual prediction values p_X using the similarities S_{AX} indirectly.

We will explain about $UserProfile$, calculation of the similarity in (5) and indirect update of individual prediction in (6) as follows.
[User profile]
We regard the vector containing the individual evaluation value R which is an individual variable as **user**

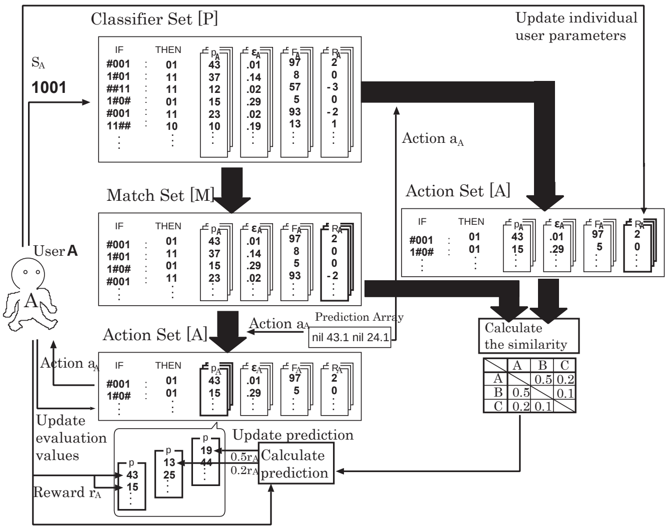

Figure 3: MULA based on XCS

profile. Therefore, the number of the individual evaluation values R in the profile to user A is the same as the number of classifiers (rules) which this system has. This system uses user profiles of two users by calculation of the similarity in the following subsection.

[**Calculation of the similarity**]

We adopt the vector space method as one of the simplest methods of calculating the similarity. Corresponding to the output action, this system makes the information of user's preference to an individual evaluation value, and acquires it for every classifier. The similarity S between user A and user B is then computed by the vector of individual evaluation value $\mathbf{R}_i$ and $\mathbf{R}_j$ in the following formulas from the user profile of each user which is used as the ingredient of the vector. **f** and **g** are two vectors (user profiles) to compare.

$$S(\mathbf{R}_i, \mathbf{R}_j) = \frac{(\mathbf{R}_i, \mathbf{R}_j)}{||\mathbf{R}_i||||\mathbf{R}_j||} = \frac{\sum_{m=1}^{K} R_i^m R_j^m}{\sqrt{\sum_{m=1}^{K}(R_i^m)^2}\sqrt{\sum_{m=1}^{K}(R_j^m)^2}} \quad (2)$$

[**Indirect update of individual prediction value**]

MULA introduces a social element into the update of individual prediction value p of each classifier. It is explained how to update this individual precidiction by interaction between human and MULA as follows.

Other users' individual prediction values p^j are updated by the following formulas according to the similarity S_{ij}. Here, p_j^t is a prediction value in step t to user j. α is a coefficient. r_j^t is the reward from human j in step t. r_max is maximum reward. Where $(S \leq 0)$, it decides r_i, r_{max} as criteria initial value $p_i^{t=0}$ in order to be able to reinforce the inverse direction. In this paper, we simply deal with single-step task in the experiment. Therefore update formula in Fig. 2 can express formulas below. Here, the formulas update users' individual prediction value indirectly by experience with user i.

$$p_j^t = \begin{cases} p_j^{t-1} + \alpha S_{ij}(r_i^{t-1} - p_i^{t-1}) \ (S > 0) \\ p_j^{t-1} - \alpha S_{ij}(r_{max} - r_i^{t-1} - p_i^{t-1}) \ (S \leq 0) \end{cases} \quad (3)$$

3 Experiments

3.1 Purpose of experiment and settings

We investigate MULA's efficiency in experiments of a user's preference prediction. In order to investigate the agent's effect intelligibly, we verified based on the following points.

- MULA predicts a user's preference individually using past experience with users and their user profiles. We verify correlation between a user's true preference and a prediction value using past experience of a user and a user profile.
- The interaction (for example, evaluation and taste) from human is always not optimal. And it is changeable. We investigate changes of the similarity.

In this experiments, we introduced MULA into pet-type robot SONY AIBO. All experiments are conducted in the following procedure based on the interaction between a user and AIBO.

(1) A subject pushes the button of AIBO and AIBO outputs an action.

(2) The subject gives one evaluation from two kinds, "praise($r = 100$)" or "scold($r = 0$)", by preference to the action of AIBO. It returns to (1).

The above procedure is considered as *one procedure of an interaction*. In this experiment, the initial values of update parameters are $\alpha = 0.2$, $r_{max} = 100$, $p^{t=0} = 50$, $R^{t=0} = 0$, $S^{t=0} = 0$.

We prepared 5 kinds of actions (e.g. "bark","dance","greeting") of agent, and 3 kinds of situation ("back button","head button","jaw button"). According to Fig. 3, all user profiles are created by classifier set selected by the same act, the similarity S_{AX} is created from them, and a suitable act is predicted to the state.

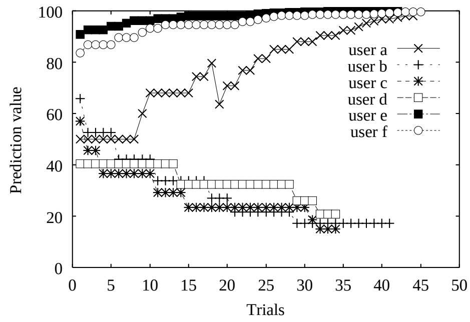

Figure 4: Changes of individual prediction values

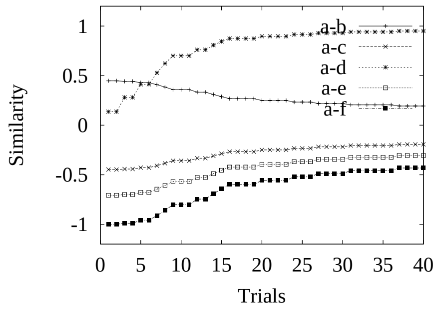

Figure 5: The changes of the similarity between user a and the others

Six subjects (a-f) participate in the experiments. The experiments divided into two phase. All subjects perform a profile creation phase first. They enter an evaluation phase in order. In an evaluation phase, a subject is influenced by the subjects who performed this phase before. For example, the 6th subject is influenced by five previous subjects.

3.2 Experimental results and discussions

Fig. 4 shows changes of individual prediction in evaluation phase. The prediction value of user e and f is large compared with the initial value $p^{t=0} = 50$, and direction of convergence "praise($r = 100$)" is in agreement, and you can see learning is promoted. In order to check the efficiency of learning, the prediction value of each action of user from b to f was investigated. The action which prediction was successful and accelerated learning was 74% and the action which prediction was not successful was 9%. The correlation coefficient of user's actual preference and a prediction value is 0.73 on the whole. Therefore, MULA has been predicted the suitable action to a user's preference.

Moreover, the similarity is always changing with the interaction between users and MULA dynamically. According to the past experience between users, the similarity will change dynamically and it will be convergenced to the fixed value in the situation of change of the similarity between users (Fig. 5). Thus, MULA calculates the optimal prediction value for the moment according to the calculated similarity for every trial. Therefore, finally exact anticipation like this experimental result was completed.

4 Conclusion

In this paper, we attempted to realize social interaction by constructing MULA which performs social learning from an interaction with two or more users. Two learning functions realized social learning. One is direct learning function using individualization of user parameters, and another is indirect learning function using past ecperience and the similarity between each users. Especially indirect learning uses experience with other users for the target user. We think it is one form of social learning to realize the suitable action from experience with two or more users to a current user. By the experiment in real environment, we verified that MULA can respond to dynamic change of users' preference, and perform efficient learning.

References

[1] S. Yamada, T. Yamaguchi : Mutual Learning of Mind Reading between a Human and a Life-like Agent, The Fifth Pacific Rim International Workshop on Multi-Agents, pp.138–150, 2002.

[2] R. S. Sutton: Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming, Proc. of the 7th International Conference on Machine Learning, pp.216-224, 1990.

[3] R. Parr and S. Russel: Reinforcement Learning with Hierarchies of Machines, Advances in Neural Information Processing Systems 10, pp.1043-1049, 1998.

[4] S. W. Wilson : Classifier Fitness Based on Accuracy, Evolutionary Computation, 3(2), pp.149-175, 1995.