Apprentissage de connaissances d’adaptation à partir des feedbacks des utilisateurs

Abir Béatrice KARAMI, Karim SEHABA, Benoît ENECELLE

abir-beatrice.karami@liris.cnrs.fr
http://liris.cnrs.fr/abir-beatrice.karami

SILEX-LIRIS

Journées francophones d’Ingénierie des Connaissances

May 14, 2014
Project RoboPopuli

Project FUI-RoboPopuli

- Financed by the French ministry of industries and labelled by Imaginove and Cap Digital.
- **Consortium**: Awabot, Artefacts Studio, ENSTA, LIRIS-SILEX
- **Projet objective**: companion robot (EMOX) with adaptive and social behaviour
- **Objectif de SILEX**: Generate a robot behaviour that is adaptive to the situation of interaction and personalised to the user's profile, using interaction traces and user’s **feedbacks**.
Illustrative Scenario

Example: projecting a video to a user

For adaptive and personalised service, many parameters to consider:

- What video to project: a movie, an episode, a comedy or a cartoon?
- What level of brightness and volume to set?
- Should the robot propose to project the video in the living room or in the bedroom?

Decisions must take into account environment situation, user preferences, other rules (e.g. parental rules) ...
Markov models for adaptive robots

[Pineau et al. 2003, Karami et al. 2011]
- Adaptive actions, adaptive interaction type, inference of user needs.
- POMDP, approaches to decrease complexity.
- **No personalised behaviour according to user profile.**

Learn to maximise interaction return

[Tapus et al. 2008, Mitsunaga et al. 2005]
- Learn to maximise interaction results using PGRL.
- Adapt distance, speed and verbal content.
- **Manual selection of important attributes.**
Rule based adaptive robots

[Kanda et al. 2005]
- Rule based systems and memorised history of interaction.
- Personalised interaction with identified students.
- **Does not learn the preference of users.**

Shaping

[Knox et al. 2009]
- Training an Agent Manually via Evaluative Reinforcement (TAMER).
- Supervised learning to model a reinforcement function.
- **Does not handle preferences and environments with several users.**
Research Issues

Main Focus

- How the robot can adapt its behaviour according to the situation and the user, taking into account his profile (age, preferences, habits, . . . )?
  - MDP model representing the user profile and the activity in its state.
  - A correct MDP reward function generates an adapted and personalised robot behaviour.

- How the robot can learn and evolve its knowledge about its users and their preferences?
  - Analysing the traces of interaction between the robot and the users (robot’s actions and the users’ feedback) in order to determine the reward function.
  - Generalising the rewards by learning the dependence between the robot behaviour and certain information about the user or the activity, this helps the robot to adapt faster with new users.
General Architecture

A = \{a_1, a_2\}
AT = \{\star, \circ\}
\star \text{ domain of values} = \{\star, \heartsuit\}
\circ \text{ domain of values} = \{\bullet, \bigcirc, \lozenge\}

Examples of interaction traces:
- a_1, (\heartsuit, \bigcirc) = -1
- a_2, (\star, \bigcirc) = +0.5

Examples of adaptation rules:
- a_1, (\star, \bigcirc) = -1
- a_2, (\star, (\heartsuit \lor \bullet)) = +0.5
## Robot Planning

**Markov Decision Process**

\[ \text{MDP} = \langle S, A, T, R \rangle \]

- **S** set of states e.g. user age, user gender, noise level, daytime, brightness and phase.
- **A** set of actions e.g. all phases actions.
- **T** \( : S \times A \times S : pr(s'|s, a) \) gives the probability of ending in a state \( s' \) knowing the starting state \( s \) and the action \( a \).
- **R** the basic reward function includes rewards that helps in respecting the sequence of phases. Additional learned reward are added from the learning procedure.
Proposed Learning Algorithms

Direct Algorithm
- Exact learning from the feedback rewards.
- Slow and needs a lot of traces.

Generalised Algorithm
- Learns the important attribute to each possible decision/action.
- Generalise and use the learned reward on new situations.
- Fast but with risk.

- Extract/Learn adaptation rules from interaction traces.
- Integrate the adaptation rules in the MDP reward function.
Generalised Algorithm

- Detects automatically the important attributes for each action/decision (minimise the complexity of representation and decision making)
**Generalised Algorithm**

Based on the detection of conflicts between the set of learned rules and a new interaction trace.

**New trace**

1. \( a_1, (\star, \bullet, \square) = +1 \)  \(\rightarrow\) No detected important attributes  \(\rightarrow\) \( a_1, (\star, \bullet, \square) = +1 \)

2. \( a_1, (\star, \bigcirc, \square) = -1 \)  \(\rightarrow\) \(\bigcirc, \square\) detected as important attributes  \(\rightarrow\) \( a_1, (\star, \bullet, \square) = +1 \)
\(\rightarrow\) No confirmation, both \(\bigcirc, \square\) are not important attributes  \(\rightarrow\) \( a_1, (\star, \bullet, \square) = +1 \)

3. \( a_1, (\bullet, \bigcirc, \square) = -0.5 \)  \(\rightarrow\) \(\bigcirc\) detected as important attribute  \(\rightarrow\) \( a_1, (\star, \bullet, \square) = +1 \)
\(\rightarrow\) Confirmation, \(\bigcirc\) is an important attribute  \(\rightarrow\) \( a_1, (\star, \bullet, \square) = +1 \)

\(\bigstar\) possible values: \(\bigstar, \bigstar\)
\(\bigcirc\) possible values: \(\bigcirc, \bigcirc\)
\(\square\) possible values: \(\square, \square\)

\(\text{AT} = \{\star, \bigcirc, \square\}\)

\(K_{a_1}\)

\( a_1, (\star, \bullet, \square) = +1 \)
\( a_1, (\star, \bullet, \square) = +1 \)
\( a_1, (\star, \bullet, \square) = +1 \)
\( a_1, (\star, \bullet, \square) = +1 \)
\( a_1, (\star, \bullet, \square) = +1 \)

\([\text{Karami et al. IC'14, RO-MAN'14}]\)
Simulations for projecting video activity

\[ |S| = \text{age}(3) \times \text{gender}(2) \times \text{noise}(3) \times \text{daytime}(5) \times \text{brightness}(3) = 270 \]

\[ |A| = 20 \]

The procedure of evaluation

1. Re/Calculate the MDP policy.
2. generate \( n \) traces using predefined rules of preference.
3. evaluate the \( n \) traces (number of negative actions).
4. extract feedback rewards from traces, then learn and update the MDP reward function.
5. repeat from step 1.
Results - Simulation

- Generalised algorithms converges to an optimal reward function with less than 50% traces.
- all important attributes were learned.
Results - Simulation

- Ambiguity: 2 identical situations with different user feedback.
Conclusion 1

- Proposed an adaptive companion robot decision model.
- Learning user’s preferences using interaction traces including users’ feedback.
- Two learning algorithms: certain and with generalisation.
- Proposed architecture and learning algorithms permit the companion robot to adapt to its users preferences and even adapt to first time users knowing only basic information about their profiles.

Real experiences with users.
Experiment with real users
25 adults (14 men and 11 women) × 4 = 100 traces

\[|S| = 32, |A| = 16\]

<table>
<thead>
<tr>
<th></th>
<th>M/F</th>
<th>V/NV</th>
<th>D/ND</th>
<th>L/D</th>
<th>S/W</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+M</td>
<td>Yes %</td>
<td>52/43</td>
<td>46/50</td>
<td>82/14</td>
<td>52/44</td>
</tr>
<tr>
<td>A+D</td>
<td>Yes %</td>
<td>0/9</td>
<td>6/2</td>
<td>2/6</td>
<td>0/8</td>
</tr>
<tr>
<td>M+D</td>
<td>Yes %</td>
<td>11/20</td>
<td>16/14</td>
<td>4/26</td>
<td>15/15</td>
</tr>
<tr>
<td>A+M+D</td>
<td>Yes %</td>
<td>38/27</td>
<td>32/34</td>
<td>12/54</td>
<td>33/33</td>
</tr>
<tr>
<td>GS</td>
<td>Yes %</td>
<td>60/63</td>
<td>98/26</td>
<td>60/62</td>
<td>59/64</td>
</tr>
<tr>
<td>SS</td>
<td>Yes %</td>
<td>26/23</td>
<td>2/47</td>
<td>27/22</td>
<td>20/30</td>
</tr>
<tr>
<td>CS</td>
<td>Yes %</td>
<td>14/14</td>
<td>0/28</td>
<td>12/16</td>
<td>22/7</td>
</tr>
<tr>
<td>MD</td>
<td>Yes %</td>
<td>30/32</td>
<td>(0/61)</td>
<td>27/36</td>
<td>33/29</td>
</tr>
<tr>
<td>VD</td>
<td>Yes %</td>
<td>70/68</td>
<td>100/39</td>
<td>73/64</td>
<td>67/71</td>
</tr>
<tr>
<td>F</td>
<td>Yes %</td>
<td>59/68</td>
<td>63/64</td>
<td>100/56</td>
<td>78/52</td>
</tr>
<tr>
<td>C</td>
<td>Yes %</td>
<td>41/32</td>
<td>37/36</td>
<td>0/44</td>
<td>22/48</td>
</tr>
<tr>
<td>RW</td>
<td>Yes %</td>
<td>18/9</td>
<td>6/22</td>
<td>4/24</td>
<td>2/25</td>
</tr>
<tr>
<td>WW</td>
<td>Yes %</td>
<td>9/5</td>
<td>12/2</td>
<td>2/12</td>
<td>4/10</td>
</tr>
<tr>
<td>B</td>
<td>Yes %</td>
<td>7/5</td>
<td>6/6</td>
<td>0/12</td>
<td>8/4</td>
</tr>
<tr>
<td>OW</td>
<td>Yes %</td>
<td>66/82</td>
<td>76/70</td>
<td>94/52</td>
<td>85/62</td>
</tr>
</tbody>
</table>
## Results Gen. Algo. [Karami et al. ICAR 2013]

<table>
<thead>
<tr>
<th>Action</th>
<th>Detected Important Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appetizer + Main Dish</td>
<td>diabetes, daytime</td>
</tr>
<tr>
<td>Appetizer + Dessert</td>
<td>vegetarianism</td>
</tr>
<tr>
<td>Main Dish + Dessert</td>
<td>daytime, diabetes</td>
</tr>
<tr>
<td>Appetizer + Main Dish + Dessert</td>
<td>gender, season</td>
</tr>
<tr>
<td>Green Salad</td>
<td>vegetarianism, diabetes</td>
</tr>
<tr>
<td>Salmon Salad</td>
<td>-</td>
</tr>
<tr>
<td>Chicken Salad</td>
<td>diabetes</td>
</tr>
<tr>
<td>Meat Dish</td>
<td>vegetarianism, season</td>
</tr>
<tr>
<td>Vegetables Dish</td>
<td>-</td>
</tr>
<tr>
<td>Fruits</td>
<td>vegetarianism, daytime</td>
</tr>
<tr>
<td>Cake</td>
<td>-</td>
</tr>
<tr>
<td>Red Wine</td>
<td>gender</td>
</tr>
<tr>
<td>White Wine</td>
<td>vegetarianism, diabetes, daytime, season</td>
</tr>
<tr>
<td>Beer</td>
<td>diabetes, season</td>
</tr>
<tr>
<td>Only Water</td>
<td>-</td>
</tr>
</tbody>
</table>
Ability of the robot learn, adapt and personalise its behaviour to its users.

Capability of our proposed algorithms to handle ambiguities in users’ feedback.

Most important attributes were detected using the algorithm with generalisation.

Complexity of learning : number of needed interaction traces.
Current and future work

- Experiments with EMOX robot (JINT).
- Complexity of learning (lack of traces with real users): traces number/state space (profile and environment)/important attributes.
- Weigh the important attributes.