Computational modeling of emotions

Christian Becker-Asano, Dr. rer. nat.
Albert-Ludwigs-Universität Freiburg
Computer Science Department
Foundation of Artificial Intelligence
Overview

a) Motivation: Why to simulate affect
b) Theoretical background: Emotion psychology
c) Implementation: WASABI & its emotion dynamics
a) Why to simulate affect

Burghouts, op den Akker, Heylen, Poel & Nijholt (2003):

1. The “Believable-agent-motive”:
   “Embodied conversational agents that show emotions in the way they act or behave in environments where they interact with humans [...] are more believable and engaging than similar agents that do not show emotions.”

2. The “Experimental-theoretical-motive”:
   “The system is built and used as an experimental environment to verify or falsify hypotheses based on the theoretical insights expressed in the emotion theory.”

The embodied conversational agent MAX:
1. as an interactive & believable embodied agent
2. as testbed for emotion theory
b) Theoretical background

1. Psychological background
   • “OCC-model” of emotions (Ortony, Clore & Collins 1988)
   • “Core Affect” (Russel & Feldmann Barrett 1999) 
     & Pleasure-Arousal-Dominance space (Russel & Mehrabian 1977)

   • Primary emotions
   • Secondary emotions

3. Social / interpersonal emotions 
   (Parkinson, Fisher & Manstead 2005)
   • sociomoral emotions: embarrassment, shame & guilt
b) The OCC-model (example) (Ortony, Clore & Collins 1988)

IF event-has-consequences-for-self AND prospects-relevant THEN

PROSPECT-BASED emotions cluster

- hope
- fear

CONFIRMED
- satisfaction
- fears-confirmed

DISCONFIRMED
- disappointment
- relief

OCC-model best suited to reason about emotions
The non-cognitive emergence of emotions is neglected
(E.g.: The experience of relaxation when sitting in front of a warm oven.)
b) Core affect
(Russel & Feldmann Barrett 1999)

Assumption for “Core Affect”:
→ Emotions not identifiable by distinct categories from the start
→ “Circumplex model of Core Affect” (Pleasantness & Activation)

<table>
<thead>
<tr>
<th>Term</th>
<th>Pleasure Mean</th>
<th>Pleasure SD</th>
<th>Arousal Mean</th>
<th>Arousal SD</th>
<th>Dominance Mean</th>
<th>Dominance SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>31. Happy</td>
<td>.81</td>
<td>.21</td>
<td>.51</td>
<td>.26</td>
<td>.46</td>
<td>.38</td>
</tr>
<tr>
<td>50. Anxious</td>
<td>.01*</td>
<td>.45</td>
<td>.59</td>
<td>.31</td>
<td>-.15*</td>
<td>.32</td>
</tr>
<tr>
<td>52. Surprised</td>
<td>.40</td>
<td>.30</td>
<td>.67</td>
<td>.27</td>
<td>-.13*</td>
<td>.38</td>
</tr>
<tr>
<td>82. Angry</td>
<td>-.51</td>
<td>.20</td>
<td>.59</td>
<td>.33</td>
<td>.25</td>
<td>.39</td>
</tr>
<tr>
<td>101. Fearful</td>
<td>-.64</td>
<td>.20</td>
<td>.60</td>
<td>.32</td>
<td>-.43</td>
<td>.30</td>
</tr>
<tr>
<td>151. Sad</td>
<td>-.63</td>
<td>.23</td>
<td>-.27</td>
<td>.34</td>
<td>-.33</td>
<td>.22</td>
</tr>
</tbody>
</table>

(Russel & Mehrabian 1977)
b) Primary & secondary emotions

1. Primary emotions (fear, anger, joy, ...):
   - fast, hard-wired stimulus response patterns
   - trigger fight-or-flight behaviors
   - ontogenetically earlier types of emotion

2. Secondary emotions (e.g., hope, shame):
   - lead to cognitively elaborated, deliberative behaviors
   - are based on memories and expectations
   - “social emotions” developed during infancy
   - “utilize the machinery of primary emotions”
b) Social emotions

- „The final recurrent theme is that emotions are not necessarily defined by the quality of the associated feeling state but may instead derive their identity from the **interpersonal dynamics** that provide the context for their subjective aspects.“ (Parkinson, Fisher, Manstead, 2005)
- Of particular interest here: „sociomoral emotions“
  - Embarrassment: „[..] an interruption of the orderly performance of social action.“
  - Shame: „[your] failure to live up to central standards of conduct [in the eyes of others].“
  - Guilt: „[..] blameworthy action is the key elicitor.“
Guilt

1. helps to repair relationships
2. encourages actions that maintain relationships
3. guilt induction serves as a way of influencing another’s conduct from a relatively powerless position

→ Guilt induction strategies:
   - If I know
     • that you are present and were able to help
     • that you don’t know
       - that I am present and need help
       - that you could help
   - then I should (at least) make you aware of me.
b) Conclusions

1. Primary emotions:
   - No memory, no expectations, no higher-order cognition
   - Elicitation of primary emotions in PAD space (Russel & Mehrabian)

2. Secondary emotions:
   - product of conscious appraisal based on memory, expectations and goal-conduciveness (Scherer)

3. Social emotions:
   - Rely on ability to reason about others' knowledge (Parkinson, Fisher & Manstead)

Implementation?
c) Nine emotions in PAD space

As of 2005 only primary emotions
c) Secondary emotions

Three prospect-based emotions as secondary emotions:

1. HOPE:
   MAX hopes that the human player plays one of her stock cards or her main card, because MAX could play his own main card afterwards.

2. FEARS-CONFIRMED:
   MAX sees his fears confirmed, because the human player played a card that MAX was afraid of before.

3. RELIEF:
   MAX realizes that the human player did NOT play a card that MAX was afraid could have been played.

Aspects of their connotative meaning represented in PAD space:
- ensuring mood-congruent elicitation
- enables calculation of their awareness likelihood
c) Secondary emotions in PAD space

Primary emotions:
- happy (50, 0, +/- 100)
- bored (0, -80, 100)
- concentrated (0, 0, +/- 100)
- depressed (0, -80, -100)

Secondary emotions:
- Relief
- Hope

Fears-confirmed:
- happy (80, 80, +/- 100)
- surprised (10, 80, +/- 100)
- angry (-80, 80, 100)
- annoyed sad (-50, 0, +/- 100)
- fearful (-80, 80, -100)
c) The WASABI architecture

Integration/categorization module

Appraisal module

- Reactive appraisal
  - intrinsic pleasantness
  - expectation deviation

- Cognitive appraisal
  - goal conduciveness
  - Dominance assessment

- Cognitive reappraisal
  - causal (mis)attribution
  - coping

Emotion dynamics

Pleasure Arousal

PAD space

Integration/categorization module

Perceive

Beliefs, desires, intentions

Goals, plans, expectations

Primary emotions

Emotional impulses

Dominance

Secondary emotions

Aware emotions

Emotion dynamics

Mood

X

Beliefs, desires, intentions

Goals, plans, expectations

Cognitive reappraisal

- goal conduciveness
- Dominance assessment

Cognitive reappraisal

- causal (mis)attribution
- coping

Emotion dynamics

Pleasure Arousal

PAD space

Integration/categorization module

Primary emotions

Emotional impulses

Dominance

Secondary emotions

Aware emotions
c) Secondary emotion example

Cognition: The prospect of an undesirable event was confirmed

→ trigger **Fears-confirmed**

awareness likelihood = (0.3 * fearful, 0.2 * sad, 0.6 * **Fears-confirmed**)
Future work!

Dynamic epistemic logic approach

Situation I:
- V knows that H present
- V not knows whether or not (H knows that V present)
  → V makes noise to induce guilt!

Situation II:
- H knows that V present
- H not knows whether or not (V knows that H present)
  → H avoids actions that imply (V knows that H present) to avoid feeling guilty

H = Helper, V = Victim
Thank you for your attention