Modeling the Logic of Emotion with Knowledge Engineering

William Jarrold
Counseling Psychology
University of Texas at Austin
william.jarrold@alum.mit.edu

Abstract

The aim of this work in progress is to implement a generative and validated model of Theory of Emotional Mind in a knowledge-based system. Model requirements are elucidated via human subject responses to autism therapy exercises.

Theory of Emotional Mind

(Baron-Cohen 1997) posits a mental module, the Theory of Mind Mechanism (ToMM), tasked with attributing mental states to social agents so as to predict and explain their behavior. My implementation should embody a theory of affective ToMM or Theory of Emotional Mind (ToEM). A successfully implemented ToEM may be applied in cognitive rehabilitation applications for people with autism (Galitsky 2000) or alexithymia.

Generative Capacity

(Ortony 2001) posits that generative capacity is critical to computational accounts of emotion. My work makes the modeling of this generativity its central goal. Consequently, ToEM should exhibit this capacity by demonstrating that it is capable of generating a multitude of appraisals with a variety of different foci and assumptions. At the same time ToEM should not overgenerate – i.e. it should not generate appraisals which are inconsistent with the target agent's desires. In other words, the "glass" may be legitimately appraised as "half-full" or "half-empty", but only certain explanations correspond to a given evaluation of a situation.

Assumptions and Scope

Certain scope delimiting assumptions are made about ToEM. As (Ortony 2001) suggests, registration of goodness or badness is fundamental to emotion. Correspondingly, scope is limited to computing event *desirability* (Ortony, Clore & Collins 1988). Secondly, appraisals are conceptualized as having two components – a judgment of valence (e.g. desirable/undesirable) and a justification of that judgment. Thirdly, scope is limited to a shared or "consensus reality" ToEM. Explaining individual differences between different agents' ToEMs is left as future work. Lastly, I model only what are termed atomic appraisals. By contrast, consider the following compound appraisal typical of responses in a recent University of Texas pilot study.

> Desire: Eric wants to ride in the car.
> Outcome: Eric is riding in the train.
> Question: Does Eric feel happy or sad? Why?
> Response: Eric is happy because he may be excited about the train because he did not know of his option to ride it and is still getting transported.

Such a compound appraisal may be broken into two positively valenced constituent atomic appraisals. One atom justifies its positivity thusly: "he did not know of his option to ride it". The other focuses on "he is still getting transported".

I assume that natural human appraisal output is obtained by generating several atomic appraisals and filtering/agglomerating the results into a final, possibly compound, appraisal. Compound appraisal generation is left as future work.

Elucidating Model Requirements

Requirements can be crystallized with a well-chosen criterial task. Howlin, et al. (1999), a clinical workbook containing social understanding exercises, provides such a task. Designed for children with autism, these exercises focus modeling effort on a fundamental level of social understanding – the child's level. Below is an adapted pilot study item from the workbook (p. 107).

Example A
Desire: Tracy wants to eat a banana.
Outcome: Mummy gives Tracy an apple.
Question: Does Tracy feel happy or sad? Why?

Pilot data on items like the above illustrates appraisal generativity – despite scenario simplicity, a wide variety of responses were btained. The criterial task is to generate a set of legitimate appraisals similar to the union of subjects' atomic appraisals. By modeling human data from such a task we reduce the chance that the final system will exhibit biases deriving from the modeler's favorite theory of emotion or other idiosyncrasies.

Knowledge Engineering

This model can be constructed in a knowledge based system such as Cyc[1] or KM[2]. This sketch exemplifies how ToEM components (bold) interact with background knowledge to exhibit the sort of generativity found in data described above. Take Example A. An appraisal type known as **Inferable Goal Failure** would produce a negative evaluation justified by background knowledge that apples are different from bananas. Additionally, pilot study data shows that several positively valenced **Goal Substitution** appraisals are generated for cases like Example A. Drawing on taxonomic knowledge that apples are a kind of food, Goal Substitution (Schank & Abelson 1977) construes the situation positively by assuming that what Tracy *really wanted* was food. Alternatively, knowledge that both bananas and apples are kinds of fruit can be used to generate another positive appraisal: "*at least* Tracy got fruit". An additional positive appraisal of this type can be generated using background knowledge to assume that Tracy has a tacit goal to receive attention from her parents. Appraisals based on comparison to counterfactuals such as "Tracy is happy – at least she wasn't given liver!" can be generated via **Downward or Upward Comparison** (Ben-Ze'ev 2000). As described above, background knowledge about object properties, taxonomic relationships and typical human goals applies in specific ways to particular examples of ToEM ontology components.

Evaluation

Having introduced the issues, a more precise statement of the claims and how to evaluate them is possible.

- Subjects should be nearly unanimous in their judgments of appraisal legitimacy.
- The ToEM implementation should generate all and only the legitimate atomic appraisals for situations.

The first hypothesis may be tested by asking subjects to distinguish legitimate from illegitimate appraisals. The second may be tested in two parts – the "all" part and the "only" part. The "only" part can be measured by the extent to which subjects find model generated appraisals sensible. The "all" part can be evaluated by the extent to which for every sampled human appraisal there is a model generated appraisal judged similar to it. **Computational Ablation** (Porter, Bareiss & Holte 1990) can be used to verify that the model is functioning as hypothesized. The basic idea is to compare performance between the full model implementation and an implementation which has been ablated or damaged in a specific way. If the model is correctly understood, ablated implementations should behave as predicted.

[1]http://www.cyc.com
[2]http://www.cs.utexas.edu/users/mfkb/RKF/km.html

References

Baron-Cohen, S. (1997) "How to build a baby that can read minds" chapter in *The Maladapted Mind*, Baron-Cohen, S. ed, Psychology Press, East Sussex UK.

Ben-Ze'ev, Aaron. (2000) *The Subtlety of Emotions* MIT Press, Cambridge Massachusetts.

Galitsky, Boris (2000) "Question-answering system for teaching autistic children to reason about mental states" DIMACS Technical Report 2000-25 September 2000.

Howlin P., Baron-Cohen S., Hadwin J. (1999) *Teaching children with autism to mind-read : a practical guide for teachers and parents* Chichester ; New York : J. Wiley & Sons.

Ortony, Clore and Collins (1988) *The Cognitive Structure of Emotion* Cambridge University Press, Cambridge England.

Ortony, A. (2001). to appear in R. Trappl & P. Petta (eds), *Emotion in Humans and Artifacts*, MIT Press, Cambridge, MA 2001.

Porter B. W., Bareiss E. R., and Holte R. C. (1990) "Concept learning and heuristic classification in weak-theory domains." *Artificial Intelligence*, 45(1–2).

Schank, R., Abelson, R. (1977) *Scripts Plans Goals and Understanding* John Wiley & Sons, New York.