Emotionally interactive agents

Models in language processing have researched how words are interpreted by humans. Many models presume the ability to correctly interpret the beliefs, motives and intentions underlying words. The interest relies also on how emotion motivates certain words or actions, inferences, and communicates information about mental state. As we will see below, some tutoring systems have explored this potential to inform user models. Likewise, dialogue systems, mixed-initiative planning systems, or systems that learn from observation could also benefit from such an approach.

As these experimental data show, activating accessible constructs or attitudes through one set of stimuli can facilitate cognitive processing of other stimuli under certain circumstances, and can interfere with it under other circumstances. Some of the results support and converge on those centered on the constructs of current concern and emotional arousal.

Future research has to take seriously into account this question: how to develop models where emotion interacts with cognitive processing. One example could be the work of Pitterman et al. (2010) where it is combined speech-based emotion recognition with adaptive human-computer modeling. With the robust recognition of emotions from speech signals as their goal, the authors analyze the effectiveness of using a plain emotion recognizer, a speech-emotion recognizer combining speech and emotion recognition, and multiple speech-emotion recognizers at the same time. The semi-stochastic dialogue model employed relates user emotion management to the corresponding dialogue interaction history and allows the device to adapt itself to the context, including altering the stylistic realization of its speech.

Interpreting the mix of audio-visual signals is essential in human communication. Researchers have to take into account the advances in the development of unimodal techniques (e.g., speech and audio processing, computer vision, etc). In traditional human-computer interaction, the user faces a computer and interacts with it via a mouse or a keyboard. In the new applications (e.g., multiple agents, intelligent homes) interactions are not explicit commands. Some of the methods include gesture, speech (Potamianos et al., 2004), eye movements (Grauman et al., 2003), etc.

We can interpret the suggested selection mechanism as an information filter. This information filter only selects the measurement for the required features and passes them to the memory system. Features that do not contribute in solving a given task are discarded. This also requires a dynamical and flexible system architecture that allows for a demand-driven combination of processing modules. We have proposed such architecture for the congruent emotion of word processing. To acquire more complex information, the system needs to combine those procedures in a suitable way within memory representation. Beside this, the system has to decide which properties it has to measure for solving the current task. The resulting representation is demand related, as only the pieces of information to solve the task is acquired. This task driven representation can serve as a foundation for learning new relations between words and emotions and for interpreting current interactions.

Zhanj & Barnden (2012) addressed the problem of the detection and revealing of the relevant "context" to inform affect detection. They implemented a context-based affect detection component embedded in an improvisational virtual platform. The software allows up to five human characters and one intelligent agent to be engaged in one session to conduct creative improvisation

within loose scenarios. Some of these conversations reveal personal subjective opinions or feelings about situations, while others are caused by social interactions and show opinions and emotional responses to other participant characters. In order to detect affect from such contexts, first of all a naïve Bayes classifier is used to categorize these two types of conversations based on linguistic cues. A semantic-based analysis is also used to further derive the discussion themes and identify the target audiences for the social interaction inputs. Then, two statistical approaches have been developed to provide affect detection in the social and personal emotion contexts. The emotional history of each individual character is used in interpreting affect relating to the personal contexts, while the social context affect detection takes account of interpersonal relationships, sentence types, emotions implied by the potential target audiences in their most recent interactions and discussion themes. The new development of context-based affect detection is integrated with the intelligent agent.

In this context, a psychological framework of emotional language processing is needed to describe the steps humans take when they interact with other computer systems or agents (e.g., Parkinson, 2009). This framework can be used to help evaluate the efficiency and naturalness of a user interface (e.g., design principles, emotional inferences, etc.). So, the key question is to represent, reason, and exploit various models of word processing to more effectively process input, generate output, and manage the dialog and interaction between different agents. The input data (words) should be, cognitive and emotionally, processed in a joint feature space according to a context-dependent model.