

Application Scenarios: An Affective Dialogue System

Ian O'Neill

School of Computer Science, Queen's University Belfast

i.oneill@qub.ac.uk

1 Emotion-oriented systems and the EU HUMAINE Network of Excellence

The application scenario described in this paper, though not a deliverable of the recently established HUMAINE European Network of Excellence, is nevertheless informed by many of the issues that motivate HUMAINE. The HUMAINE network (Human-Machine Interaction Network on Emotion) is supported under the EU's Sixth Framework Programme as part of the 'IST Priority' (Information Society Technologies). Bringing together 27 partners from 11 countries, HUMAINE's aim is to lay sound foundations for emotion-sensitive computing in Europe. Among the areas being considered by the network are those where the emphasis is on the technological challenge: how to plan and anticipate emotion-related, or affective, actions and reactions; how to synthesise signs of emotion; and how to detect emotion. Other areas are more relevant to emotion-enabled applications: how to use affective language; how to ensure and how to evaluate usability. Still other areas are strongly rooted in psychology: theories of psychological processes; and the collection and analysis of empirical data. Impinging on all these issues are the philosophical questions of how to describe emotion, and how to develop and use emotion-enabled systems ethically. To give substance to its efforts, HUMAINE will develop a number of exemplars – not full-scale prototype systems, but rather a selection of technological illustrations at various levels of sophistication. That said, the affective dialogue scenario described in the following sections gives some idea of the potential of a full, working, emotion-enabled application.

2 Spoken dialogue systems and affective dialogue systems

Spoken dialogue systems are currently used as providers of information or services: they listen for key words in a user's utterances and make an appropriate spoken response. Such systems can be made to sound convincingly life-like: the system's voice will probably recreate the modulation of natural speech, and will probably be able to emphasise important phrases. If the system has an on-screen presence (an avatar or persona), this may be programmed to look pleased when it can fulfil a request or concerned when it has to reject a request, and so on.

An affective dialogue system (ADS) will have all these capabilities, but will add an important new dimension: it will be emotion-enabled. An ADS will assess the user's emotional state from the prosodic and semantic features of their speech, from their gestures, posture and facial expressions. In its turn it will use these same modalities to influence the user's attitude towards the system, determining on-the-fly the affective qualities it should adopt and the means it should adopt to realise them.

Key features of an affective dialogue system

As with a spoken dialogue system, an ADS will use a number of technologies in combination to give the impression of naturalistic dialogue. Like a conventional spoken dialogue system, the ADS will recognise speech and semantically parse recognised utterances. It will generate text strings of natural language and synthesise speech from them. An affective dialogue manager (ADM) at the heart of the system will decide what the system should do next, given a particular combination of inputs from the user. However, in the range of inputs that the system receives

from the user, and the range of resources it has at its disposal when it responds, the ADM will differ significantly from a more conventional dialogue manager (DM).

The affective dialogue manager will use probabilistic approaches to interpret the user's emotional state. It will do so in fairly broad terms given the current state of the art – e.g., given these spoken and visual signs (word choice, prosody, gesture, etc.), is the user more probably calm or agitated? For a more accurate assessment of the user's probable emotional state these emotion-related input streams can be correlated with the overall progress of a discourse towards some transactional goal – e.g., is the user effortlessly providing the information that the system would need in order to reserve a theatre ticket, or is each piece of information being repeatedly negated or amended by the user in an unproductive, protracted exchange? Active strategies can be used to clarify remaining ambiguity in the signs of emotion: if, in order to formulate an appropriate response, the system needs to know whether the user's apparent agitation is due to anger or confusion, it can, quite simply, ask.

Having assessed the state of the discourse and the emotional state of the user, the system may decide, on probability, to maintain its current dialogue strategies (all is progressing well, so just provide the information requested, adopting a confident manner...) or modify its strategies (the user seems confused, so provide the information requested, along with additional advice, and adopt a concerned manner...)

Wizard-of-oz experiments during initial phases of ADS development can be used to assess different users' verbal and non-verbal reactions to a simulated, emotionally-enabled system: such experiments are crucial for gauging the weights to be attached to signs of emotion in the different modalities. Moreover, the strategies that make human beings effective communicators of both emotion and information can be used by an ADS as well.

What are the likely advantages?

In stereotypical, short service transactions (e.g. buying a ticket) the utility of emotion-awareness is probably limited to the detection and avoidance of undesirable user states. Longer, mixed-initiative interactions, typically involving several interrelated dialogue topics, and entailing problem-solving and negotiation, offer greater scope for the deployment of emotion-enabled systems – e.g. planning a week-end break. In these contexts people expect a human service provider to show some emotional engagement: interest when they are talking, moderate pleasure when they succeed in achieving a goal, moderate concern when they have difficulties, and so on. An affective dialogue system should attempt to match that kind of behaviour. The system's reactions should be subtle and human-like: a usable affective dialogue system will have to do much more than synthesise and react to full-blown, archetypal emotions. An *appropriate* level of affect should make users feel more comfortable working with the system and improve the outcome of the dialogue, in much the same way as a conversation with a sympathetic dialogue partner tends to be a more satisfactory experience than a conversation with a disengaged, distanced interlocutor.

Timeliness of experimentation with affective dialogue

Efforts to improve individual emotion interpretation components (subsystems that can interpret the likely emotion-related significance of a facial expression, gesture or tone of voice) are already underway, not least in HUMAINE. However, since the performance of individual emotion interpretation components is unlikely to improve rapidly in the near future, an affective dialogue system provides an excellent context in which to investigate the combined capabilities of current, imperfect technologies, and to gauge the effect of changing a particular component's behaviour. Implementing and optimising an affective dialogue system is a challenge that is waiting to be tackled!