

# Affect Interpretation in Metaphorical and Simile Phenomena and Multithreading Dialogue Context

Li Zhang

*School of Computing, Engineering and Information Sciences,  
University of Northumbria, Newcastle,  
UK*

## 1. Introduction

The detection of complex emotions and value judgments from open-ended text-based multi-threaded dialogue and diverse figurative expressions is a challenging but inspiring research topic. In order to explore this line of research, previously we have developed an affect inspired AI agent embedded in an improvisational virtual environment interacting with human users. The human players are encouraged to be creative at their role-play under the improvisation of loose scenarios. The AI agent is capable of detecting 25 affective states from users' open-ended improvisational input and proposing appropriate responses to stimulate the improvisation.

We notice in the collected transcripts, metaphors and similes are used extensively to convey emotions such as "mum rocks", "u r an old waiter with a smelly attitude", "I was flamed on a message board", "u stink like rotten meat", "a teenage acts like a 4 year old" etc. Such figurative expressions describe emotions vividly. Fainsilber and Ortony (1987) commented that "an important function of metaphorical language is to permit the expression of that which is difficult to express using literal language alone". There is also study on general linguistic cues on affect implication in figurative expressions as theoretical inspiration to our research (Kövecses, 1998; Barnden, 2007; Zhang et al., 2009). Thus affect detection from metaphorical and simile phenomena draws our research attention. In this chapter, we particularly focus on affect interpretation of a few metaphors including cooking and sensory metaphors, and simile expressions with the usage of comparative 'like' prepositional phrases.

Moreover, our previous affect sensing is conducted purely based on the analysis of each turn-taking input itself without using any contextual inference. However, most relevant contextual information may produce a shared cognitive environment between speakers and audience to help inference affect embedded in emotionally ambiguous input and facilitate effective communication. As Sperber & Wilson (1995) stated in Relevance theory "communication aims at maximizing relevance and speakers presume that their communicative acts are indeed relevant". Such relevant contextual profiles have also been employed in our present work to model cognitive aspect of personal and social emotion and

assist affect sensing from literal and figurative input. Also we only focus on 'neutral' and 9 most commonly used emotions (disapproving, approving, grateful, happy, sad, threatening, regretful, angry, and caring) out of the 25 affective states on contextual emotion analysis and prediction. We used a school bullying<sup>1</sup> and a crohn's disease<sup>2</sup> scenario in our previous user testing and the AI agent played a minor role in the improvisation of both scenarios. In this chapter, we mainly use the collected transcripts of both scenarios for the illustration of metaphor and simile phenomena recognition and contextual affect analysis.

## 2. Related work

Much research has been done on creating affective virtual characters. Indeed, emotion theories, particularly that of Ortony et al. (1988) (OCC), are used widely in such research. Egges et al. (2003) provided virtual characters with conversational emotional responsiveness. Aylett et al. (2006) also focused on the development of affective behaviour planning for their synthetic characters.

Text-based affect detection becomes a rising research branch recently (Shaikh et al., 2007; Zhang, 2010; Liu & Singh, 2004). Façade (Mateas, 2002) included shallow natural language processing for characters' open-ended input. But the detection of major emotions, rudeness and value judgements was not mentioned. Zhe and Boucouvalas (2002) demonstrated an emotion extraction module embedded in an Internet chatting environment. However the emotion detection focused only on emotional adjectives, and did not address deep issues such as figurative expression of emotion. Also, the concentration purely on first-person emotions is narrow. Context-sensitive research is also employed to detect affect. Ptaszynski et al. (2009) developed an affect detection component with the integration of a web-mining technique to detect affect from users' input and verify the contextual appropriateness of the detected emotions. However, their system targeted conversations only between an AI agent and one human user in non-role-playing situations, which greatly reduced the complexity of the modeling of the interaction context. Moreover, the metaphorical description of emotional states is common in literature and has been extensively studied (Fussell and Moss, 1998), for example, "he nearly exploded" and "joy ran through me," where anger and joy are being viewed in vivid physical terms. In the work of Zhang & Barnden (2010), a few other metaphorical affective expressions (such as animal metaphor ("X is a rat") and food metaphor ("X is walking meat")) were intensively studied and affect was derived from such simple metaphorical expressions.

There are also well-known cognitive theories on emotion modeling. The OCC model provided cognitive appraisal theories for 22 emotional states, while Gratch and Marsella (2004) also presented an integrated emotion model of appraisal and coping, in order to reason about emotions and to provide emotional responses, facial expressions and potential social intelligence for virtual agents. However, there is very limited research on contextual

---

<sup>1</sup> It is mainly about the bully, Mayid, is picking on a new schoolmate, Lisa. Elise and Dave (Lisa's friends) and Mrs Parton (the school teacher) are trying to stop the bullying.

<sup>2</sup> Peter has Crohn's disease and has the option to undergo a life-changing but dangerous surgery. He needs to discuss the pros and cons with friends and family. Janet (Mum) wants Peter to have the operation. Matthew (younger brother) is against it. Arnold (Dad) is not able to face the situation. Dave (the best friend) mediates the discussion.

emotion modeling in cognitive science to guide our practical development on contextual affect analysis. Lopez et al. (2008) have proposed an emotion topology integrated with the consideration of contextual and multimodal elements to facilitate computerization.

Our work thus distinguishes on the following aspects: (1) affect detection from metaphorical and simile expressions; (2) affect sensing for basic and complex emotions in improvisational role-play situations; (3) affect detection for second and third person cases (e.g. 'you', 'she'); and (4) affect interpretation based on contextual profiles.

### **3. The original affect detection processing and the system architecture**

As mentioned earlier, our original system has been developed for secondary school students to engage in role-play situations in virtual social environments. Without pre-defined constrained scripts, the human users could be creative in their role-play within the highly emotionally charged scenarios. After an inspection of the recorded improvisational transcripts, we noticed that the language used for improvisation is complex and idiosyncratic, e.g. often ungrammatical and full of abbreviations, mis-spellings, etc. Several pre-processing procedures have been developed in our application previously to deal with misspellings, abbreviations, letter repetitions, interjections and onomatopoeia etc. Moreover, the language contains a large number of weak cues to the affect that is being expressed. These cues may be contradictory or they may work together to enable a stronger interpretation of the affective state. In order to build a reliable and robust analyser of affect it is necessary to undertake several diverse forms of analysis and to enable these to work together to build stronger interpretations. Also our previous affect detection has been performed solely based on the analysis of individual turn-taking user input without any contextual inference. Overall, we have adopted rule-based reasoning, robust parsing, pattern matching, semantic and sentimental profiles (e.g. WordNet (Fellbaum, 1998) and WordNet-Affect (Strapparava and Valitutti, 2004)) for affect detection analysis. Jess, the rule engine for Java platform, has been used to implement the rule-based reasoning while Java has been used to implement other algorithms and processing with the integration of the off-the-shelf language processing tools, such as WordNet.

In our turn-taking based affect interpretation, we have considered the following affective expressions. We found that one useful pointer to affect was the use of imperative mood, especially when used without softeners such as 'please' or 'would you'. Strong emotions and/or rude attitudes were often expressed in this case. Expression of the imperative mood in English is surprisingly various and ambiguity-prone. We have used the syntactic output from the Rasp parser (Briscoe and Carroll, 2002) and a semantic resource (Esuli and Sebastiani, 2006) to deal with certain types of imperatives. In an initial stage of our work, affect detection was based purely on textual pattern-matching rules that looked for simple grammatical patterns or templates partially involving specific words or sets of specific alternative words. As mentioned above, Jess is used to implement the pattern/template-matching rules in the AI agent allowing the system to cope with more general wording and ungrammatical fragmented sentences. The rules conjectured the character's emotions, evaluation dimension (negative or positive), politeness (rude or polite) and what response the automated actor should make. However, it lacked other types of generality and could be fooled when the phrases were suitably embedded as subcomponents of other grammatical structures.

In order to go beyond certain such limitations, sentence type information obtained from Rasp was also adopted in the rule sets. Such information not only helped the agent to detect affective states from the input (such as the detection of imperatives), and to decide if the detected affective states should be counted (e.g. affects detected in conditional sentences were not valued), but also contributed to proposing appropriate responses.

The results of this affective analysis were then used to (see Figure 1):

1. Control the automated actor (EMMA) that operates a character in the improvisation. I.e. the detected affective states enable the AI agent to make appropriate responses to stimulate the improvisation.
2. Additionally, drive the animations of the avatars in the user interface so that they react bodily in ways that is consistent with the affect that they are expressing, for instance by changing posture or facial expressions.

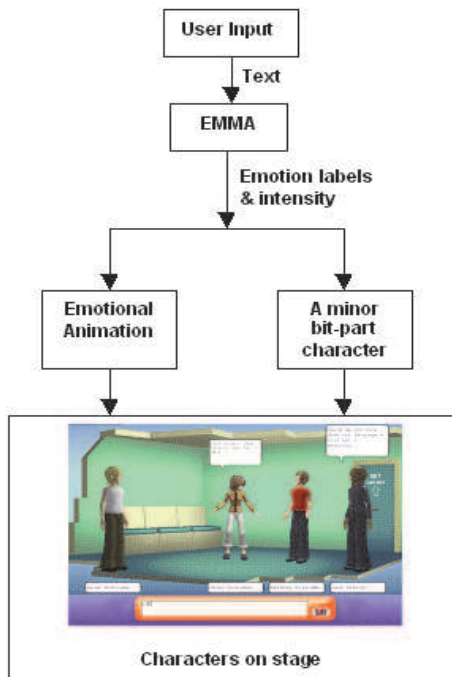


Fig. 1. Affect detection and the control of characters.

We have also developed responding regimes for the AI actor. EMMA normally responds to, on average, every Nth speech by another character in one improvisational session, where N is a changeable parameter (currently usually set to 3). However, it also responds when EMMA's character's name is mentioned, and makes no response if it cannot detect anything useful in the utterance it is responding to. The one-in-N average is achieved by sampling a random variable every time another character says something. We also have N dynamically adjustable according to how confident EMMA is about what it has discerned in the utterance at hand so that it is less likely to respond if it has less confidence. EMMA makes a

random response from several stored response candidates that are suitable for the affective quality it has discerned in the utterance it is responding to.

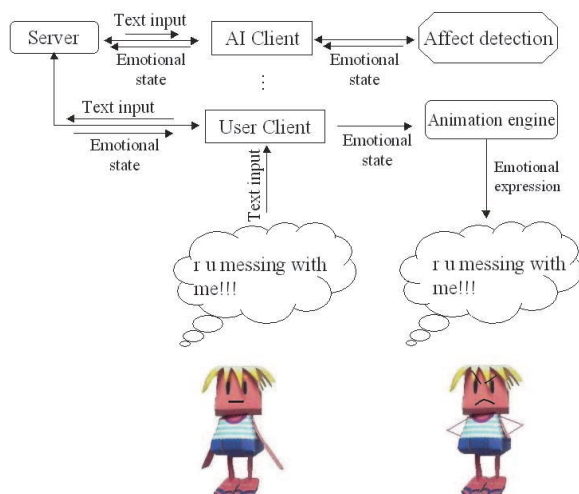


Fig. 2. An example of real-time interaction.

Moreover, our system employs a client/server architecture for implementation. The conversational AI agent and other human-controlled characters consist of clients. The server broadcasts messages sent by one client to all the other clients. Thus user's text input from normal user client is sent to the AI agent client via the server. Then the AI agent, who plays a minor role in the improvisation with other human-controlled characters, analyzes the user's text input and derives the affective implication out of the text. Then the AI agent also searches its knowledge base to provide a suitable response to the human players using the detected affective states. We have particularly created the AI agent's responses in a way which could stimulate the improvisation by generating sensitive topics of the storyline. Then an XML stream composed of the detected affective state from one user input and the AI agent's response is dynamically created and broadcasted to all other clients by the server. The users' clients parse the XML stream to obtain the information of the previous "speaker's" emotional state and the current AI character's response. An animation engine has embedded in each user client which updates the user avatars' emotional facial and gesture animation on each user's terminal. Therefore, if the previous human-controlled character expresses 'anger' affective state by saying "r u messing with me!!!", the animation engine in each user client updates emotional animation of that character on each terminal using cross behavior via simple facial and gesture animation (see Figure 2). In each session, up to five characters are engaged in.

#### 4. Metaphorical affect interpretation

Metaphorical language can be used to convey emotions implicitly and explicitly, which also inspires cognitive semanticists (Kövecses, 1998). Examples such as, "he is boiling mad" and "Lisa fired up straightaway", describe emotional states in a relatively explicit if metaphorical way. But affect is also often conveyed more implicitly via metaphor, as in "his

room is a cesspit”: affect (such as ‘disgust’) associated with a source item (cesspit) gets carried over to the corresponding target item (the room). There are also cooking metaphor examples implying emotions implicitly, such as “he is grilled by the teacher”, “he knew he was going to be toast when he got home”. In these examples, the suffering agents have been figuratively conceptualized as food. They bear the results of intensive or slow cooking. Thus, these agents who suffer from such cooking actions tend to feel pain and sadness, while the cooking performing agents may take advantage of such actions to achieve their intentions, such as persuasion, punishment or even enjoyment. We detected affect from such metaphorical expressions previously and used the AI agent as a useful application of theoretical inspiration for figurative language processing generally.

Especially we notice sensory and another type of cooking metaphors not only implying emotions but also sharing similar linguistic syntactical cues. The sensory metaphor we are interested in includes temperature, smell, taste, and light metaphors. We gather the following examples for the study of the semantic and syntactical structures of such metaphorical expressions, including cooking metaphor: “the news inflamed her temper”, “he dishes out more criticism than one can take”, “she was burned by a shady deal”; light metaphor: “you lighted up my life”; temperature metaphor: “they are kindling a new romance”; taste metaphor: “bittersweet memories” and smell metaphor: “love stinks”, “the stench of failure” etc.

In the above cooking metaphor examples, the cooking actions are performed on cognitive abstract entities (‘temper’, ‘criticism’) or human agents (‘she’) [physical cooking actions + abstract entities/human agents]. Sometimes, human agents are also the objects of cooking actions performed by abstract subject entities (“she was burned by a shady deal”), which may lead to human agents’ negative emotional experience. Similarly in the sensory metaphor examples, the light and temperature metaphors show similar syntactical structures with actions conducting respectively on existence (‘my life’) or relationship abstract entities (‘romance’) [physical actions + abstract entities]. Emotion abstract entities are also used as subjects that are capable of performing actions such as love in smell metaphors [abstract subject entities + physical actions]. Overall, the above cooking and sensory metaphors indicate that: abstract entities are able to perform physical actions while they can also be the objects of physical actions. Also examples show cognitive abstract entities may also have characteristics of food, temperature, taste or smell (‘adj. + abstract entities’). In another word, some cognitive abstract entities could be un-cooked (“a raw talent”), tasty (“bittersweet memories”) or have temperature (“heated debate”, “burning love”) or smell (“the stench of failure”). We use such semantic preference violations to sense these metaphor phenomena and their affective states.

First, we use Rasp (Briscoe & Carroll, 2002) to indentify each subject, verb phrase and object in each sentence. Then we particularly send the main terms in these three components to WordNet (Fellbaum, 1998) to recover their hypernyms. We also focus on the analysis of phrases with a structure of ‘adjective + noun’ by deriving the synonyms or related nouns for the adjective and hypernyms for the noun term using WordNet. If the inputs indicate structures of ‘abstract subject entities + actions’, ‘physical actions + abstract object entities’ or ‘temperature/smell/taste/cooking adjectives + abstract entities’, then the inputs are recognized as metaphorical expressions. The detailed processing is also shown in Figure 3.

For example, the AI agent carries out the following processing to sense the metaphorical expression “they are kindling a new romance”.

1. Rasp: the input -> ‘subject PPHS2 (they) + VBR (are) + VVG (kindle + ing) + AT1 (a) + JJ (new) + object NN1 (romance)’
2. WordNet: ‘kindle’ -> hypernym: FLARE UP, LIGHT. ‘Romance’ -> relationship -> abstract entity;
3. An evaluation profile (Esuli & Sebastiani, 2006) determines: LIGHT-> positive; ‘romance’ -> positive.
4. The input indicates -> ‘third person subject performs a ‘positive’ action towards an abstract entity (romance) -> it is recognised as a metaphorical input.
5. The third person human subject (they) may experience a positive emotion by boosting up a ‘positive’ relationship abstract entity.

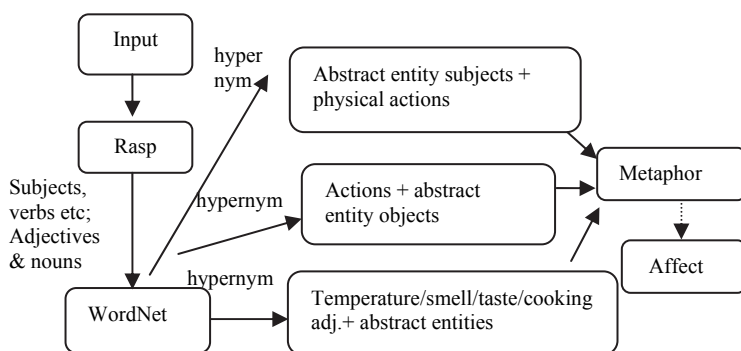


Fig. 3. Metaphor interpretation and detection.

Moreover, the AI agent conducts the following processing to sense the metaphorical expression “Mayid has a smelly attitude”:

1. Rasp: ‘NP1 (Mayid) + VHZ (has) + AT1 (a) + JJ (smelly) + NN1 (attitude)’
2. WordNet: ‘attitude’ -> hypernym: psychological feature -> abstract entity. ‘smelly’ -> synonyms & related nouns: ill-smell, foul, and malodorous.
3. The evaluation profile indicates: foul, malodorous -> ‘negative’.
4. Part of the input is interpreted as: ‘a cognitive abstract entity has negative smell (i.e. a smell adj with negative indication + an abstract cognitive entity)’ -> identified as a smell metaphor with negative implication.
5. The input becomes: ‘NP1 (Mayid) + VHZ (has) + a smell metaphor with negative indication’ -> implies ‘insulting/angry’ of the speaker towards ‘Mayid’.

Although the above metaphor recognition is at its initial stage, the system is capable of performing affect sensing and metaphor recognition more robustly and flexibly. It can also recognize other metaphorical input such as “a warm reception”, “she is burnt by a shady deal”, “deep, dark thoughts”, “he stirred up all kinds of emotion” etc.

In the Crohn’s disease scenario, metaphorical expressions have also been used to indicate battles between family members and Peter’s stress towards his life changing operation. An example interaction taken from a recorded transcript is as follows:

Dave: what are your other options peter

Peter: im trying very hard but theres too much stuff blocking my head up

Peter: my plate is already too full.... there aint otha options dave

In the first input from Peter, 'thoughts' have been regarded as physical solid objects that can occupy physical space such as a plate or head. With the contextual inference, in Peter's second input, plate has also been metaphorically used to refer to one's head. Moreover, we can hardly consider the last input as a metaphorical expression if without any contextual inference. This also indicates directions of our future work on metaphor interpretation.

## 5. Affect sensing from simile expressions

In the collected transcripts from the school bullying scenario, we also notice similes are used to convey strong emotions by linking and comparing dissimilar concepts via prepositions such as 'like'. An example interaction is demonstrated as follows.

1. Mayid: you need to stop being such an ugly little girl [angry]
2. Mrs Parton: Mayid detection![threatening]
3. Mayid: the only problem with me is that lisa is so mean to me. [angry]
4. Lisa: you need to learn how to be nice and act your age. [neutral]-> [angry]
5. Mayid: look at you though [neutral] -> [angry]
6. Lisa: about what? I don't even know u. tell him miss. [neutral] -> [angry]
7. Mayid: *a teenager acts like a 4 year old* [neutral] -> [angry]
8. Mayid: I bet you still play with Barbie dolls
9. Mrs Parton: CALM DOWN NOW!
10. Elise: that crap really don't suit u mayid
11. Lisa: *a teenager acts like a heavyweight boxer*

Indicated by italics in the above example, Mayid used a simile to indicate 'insulting/angry' by comparing Lisa's behavior of calling the schoolteacher for help with that of a 4 year old, while Lisa also employed a similar simile expression to imply Mayid's inappropriate behavior such as threatening or beating other characters as if he were a heavyweight boxer. There are also other similar simile expressions implying emotions such as "you dance like an angel", "Tom ate like a pig", "she sings like a bird" etc. In our processing, we particularly focus on such similes with a syntactical structure of 'subject + verb + preposition (like) + object'. The subjective evaluation profile and WordNet are then used to further derive the affect attached with the simile input. For example, we use the following processing to interpret the simile expression "you dance like an angel":

1. Rasp: 'PPY (you) + VV0 (dance) + II (like) + AT1 (an) + NN1 (angel)'
2. WordNet: 'angel' -> good person -> physical entity.
3. The evaluation profile shows: 'angel' -> positive;
4. The input indicates: 'the performance of a second person subject is compared with that of another person with 'positive' implication'.
5. The input implies 'affectionate'.

For the example, "a teenager acts like a heavyweight boxer", the processing taken is in the following.



1. Rasp: 'subject NN1 (teenager) + VVZ (acts) + II (like) + AT1 (a) + object NN1 (heavyweight) + object NN1 (boxer)'
2. WordNet: 'heavyweight' -> 'wrestler' -> person; 'boxer' -> 'combatant', 'battler' -> person
3. The input becomes: 'the subject's action is compared with that of another person' -> recognized as a simile.
4. Since the evaluation profile cannot provide any positive/negative evaluation values for the noun terms: 'heavyweight', 'wrestler', 'boxer', 'combatant', and 'battler', the root forms of the noun terms are used to retrieve the evaluation values.
5. The evaluation profile determines: 'wrestle', 'combat', 'battle' -> 'negative'; thus 'heavyweight' and 'boxer' -> 'negative'.
6. The input implies: 'the subject's action is compared with that of another person with a negative indication' -> thus the input conveys 'insulting/angry'.

However, purely based on the analysis of the input "a teenage acts like a 4 year old" itself, the AI agent recognizes the simile expression but fails to determine the affect conveyed in it due to the following processing:

1. Rasp: 'subject NN1 (teenager) + VVZ (acts) + II (like) + AT1 (a) + MC (4) + NNT1 (year) + JJ (old)'
2. WordNet: 'old' -> age; the evaluation profile: age -> objective (i.e. non-emotional);
3. The input becomes: 'the subject's action is compared with that of a 'neutral' object' -> recognized as a simile.

Since interaction context plays important roles in discovering affect conveyed in emotionally ambiguous input, it is resorted to to further justify the affect detected from the analysis of individual turn-taking input. I.e. context-based affect detection is employed to justify the neutral expression drawn from the analysis of the input itself, such as affect justification for the above neutral simile expression. As Schnall (2005) stated that the intention of communication is to achieve the greatest possible cognitive outcome with the smallest possible processing effort, i.e. "to communicate only what is relevant". Thus in the following section, we discuss context-based affect detection and emotion modeling in personal and social interaction context to justify affect interpretation in literal and figurative expressions.

## 6. Context-based affect detection

Lopez et al. (2008) suggested that context profiles for affect detection included social, environmental and personal contexts. In our study, personal context may be regarded as one's own emotion inclination or improvisational mood in communication context and the social context may refer to other characters' emotional influence to the current speaker. We believe that one's own emotional states have a chain effect, i.e. the previous emotional status may influence later emotional experience. We make attempts to include such effects into emotion modelling. Bayesian networks are used to simulate such personal causal emotion context. E.g. we regard the first, second and third emotion experienced by a particular user respectively as A, B and C. We assume that the affect B is dependent on the first emotional state A. Further, we assume that the third emotion C, is dependent on both the first and second emotions, A and B. In our application, given two or more most recent emotional states a user experiences, we may predict the most probable emotion this user implies in the current input using a Bayesian network.

Briefly, a Bayesian network employs a probabilistic graphical model to represent causality relationship and conditional (in)dependencies between domain variables. It allows combining prior knowledge about (in)dependencies among variables with observed training data via a directed acyclic graph. It has a set of directed arcs linking pairs of nodes: an arc from a node X to a node Y means that X (parent emotion) has a direct influence on Y (successive child emotion). Such causal modelling between variables reflects the chain effect of emotional experience. It uses the conditional probabilities (e.g.  $P[B|A]$ ,  $P[C|A,B]$ ) to reflect such influence between prior emotional experiences to successive emotional expressions.

In our application, any combination of the 10 most commonly used emotional states could be used as prior emotional experience of the user. Also each conditional probability for each potential emotional state given two or more prior emotional experiences (such as  $P[\text{approval}|A,B]$  etc) will be calculated. The emotional state with the highest conditional probability is selected as the most probable emotion the user conveys in the current turn-taking. Moreover, it is beneficial that the Bayesian network allows us to use the emotional states experienced by a particular character throughout one improvisation as the prior input to the network so that our system may learn about this user's emotional trend gradually for future prediction. In detail, at the training stage, two human judges (not involved in any development) marked up 3 example transcripts of the school bullying scenario, which consisted of approximately 470 turn-taking inputs. For each character, we extract three sequences of emotions from the improvisation of the 3 example transcripts to produce prior conditional probabilities. We take a frequency approach to determine the conditional probabilities for each Bayesian network. When an affect is annotated for a turn-taking input, we increment a counter for that expressed emotion given the two preceding emotions. For each character, a conditional probability table is produced based on the training data. An example conditional probability table is presented in Table 1.

		Probability of the predicted emotional state C being:			
Emotion A	Emotion B	Happy	Approval	...	Angry
Happy	Neutral	P00	P01	...	P09
Neutral	Angry	P10	P11	...	P19
Disapproval	Disapproval	P20	P21	...	P29
Angry	Angry	P30	P31	...	P39

Table 1. An example conditional probability table for emotions expressed by one character

In the above table, the predicted emotional state C could be any of the most frequently used 10 emotions. At the training stage, the frequencies of emotion combinations in a  $10 * 10 * 10$  ( $(A*B)*C$ ) matrix are produced dynamically. This matrix represents counters ( $N_{CAB}$ ) for all outcomes of C given all the combinations of A and B. A one-dimensional array is also needed to store counters ( $N_{AB}$ ) for all the combinations of two prior emotions, A and B. Such a conditional probability matrix is constructed at run-time for each human-controlled character in the school bullying scenario based on the training emotional sequences.

For the prediction of an emotion state mostly likely implied in the current input by a particular character at the testing stage, the two prior recent emotional states are used to

determine which row to consider in the conditional probability matrix, and select the column with the highest conditional probability as the final output. The emotional sequences used for testing expressed by each character have also been used to further update and enrich the training samples so that these testing emotional states may also help the system to cope with any new emotional inclination because of each character's creative improvisation.

An example algorithm of the Bayesian affect sensing is provided in the following. For the initial run of the algorithm, A, B and C are initialized with the most recent affects detected for each character purely based on the analysis of individual input.

### **Pseudo-code for affect prediction using a Bayesian network**

Function Bayesian\_Affect\_Prediction

- ```
{
1. Verify the contextual appropriateness of the affect C predicted by the Bayesian reasoning;
2. Produce the row index, i, for any given combination of the two preceding emotional states A & B in the matrix;
3. Indicate the column index, j, for the recommended affect C;
4. Increment counters:  $N_{AB}[i]$  and  $N_{CAB}[i][j]$ ;
5. Update two preceding emotions by: Emotion A = Emotion B; Emotion B = The newly recommended affect C;
6. Produce the new row index, k, for any given combination of the updated two preceding emotional states A & B;
7. Calculate probabilities (i.e.  $P[C|A,B] = N_{CAB}[k][column]/N_{AB}[k]$ ) for the predicted emotional state C being any of the 10 emotions;
8. Select and return the affect with the highest probability as the predicted affect C; }
```

At the testing stage, when an affect is predicted for a user's input using the Bayesian network, the contextual appropriateness of the detected affect will be further justified. The verification processing using neural network-based reasoning, which will be introduced at a later stage, results in a final recommended affect. Then the conditional probability table obtained from the training stage is updated with the newly recommended affect and its two preceding emotions. The above processing is iterative to predict affect throughout an improvisation for a particular character based on his/her personal emotional profiles.

Moreover social emotional context also has great potential to affect the emotional experience of the current speaking character. E.g., a recent threatening input contributed by Mayid may cause Lisa and her friends to be 'angry'. A neural network algorithm, backpropagation, is used to model such an effect, which accepts two most recent emotions contributed by two other characters as input. The neural network implementation has three layers and 2 nodes in the input layer & 10 nodes respectively in the hidden and output layers indicating 'neutral' and the most commonly used 9 emotions in our application. Since it is a supervised learning algorithm, we use emotional context gathered from transcripts across scenarios as training data. This neural network-based reasoning may discover the emotional influence of other characters towards the current speaker as output.

At the training stage, we have used 5 transcripts of the school bullying scenario collected in our previous user testing to generate the training data of the emotional contexts. Two

human judges have been used to provide affect annotations of these interaction contexts. After the neural network has been trained to reach a reasonable average error rate (less than 0.05), it is used for testing to predict emotional influence of other participant characters towards the speaking character in the test interaction contexts.

For the affect analysis of the above example transcript of school bullying scenario shown at the beginning of section 5, first of all, the AI agent performs affect sensing purely based on the analysis of the input itself without any contextual reasoning to annotate each user input. Therefore we annotate the affect conveyed from the 1<sup>st</sup> input to the 4<sup>th</sup> input. Since the 4<sup>th</sup> input from Lisa indicates non-emotional and generally statement inputs with second person subjects tend to convey emotions (e.g. “u r an angel”, “u aren’t needed here” etc), the contextual affect analysis based on the above description is activated. However since this is Lisa’s first input, we do not have any emotional profile yet to activate the improvisational mood prediction using the Bayesian approach. But we can still resort to the neural network-based social context modeling to reason the emotional influence from other characters to Lisa. With the most recent emotional context, ‘threatening (2<sup>nd</sup> input)’ and ‘angry (3<sup>rd</sup> input)’, provided by Mrs Parton and Mayid, as input to the Backpropagation reasoning, it deduces that in the 4<sup>th</sup> input Lisa has the highest probability (0.985) to be ‘angry’. Thus we adjust the affect implied in the 4<sup>th</sup> input to ‘angry’ caused by the social emotional context from other characters. Similarly for the 5<sup>th</sup> ‘neutral’ input from Mayid, the AI agent conducts the following processing:

1. The emotional profile of Mayid: ‘angry(1<sup>st</sup> input) and angry (3<sup>rd</sup> input)’ used as input to personal emotional context modeling via the Bayesian network -> ‘angry’ as the predicted most probable affect;
2. The social emotional context contributed by two other characters: ‘threatening (2<sup>nd</sup> input) and angry (4<sup>th</sup> input)’, used as input to Backpropagation reasoning -> ‘angry’ as Mayid’s mostly likely emotional inclination, which strengthens the output obtained from personal emotion context modeling.
3. The 5<sup>th</sup> input from Mayid is adjusted to be ‘angry’.

With the emotional context contributed by Mayid for the 3<sup>rd</sup> (angry) and 5<sup>th</sup> input (angry), the neural network-based social emotional context modeling also indicates ‘anger’ is implied in the 6<sup>th</sup> input from Lisa. For the 7<sup>th</sup> input “a teenager acts like a 4 year old”, we have the following procedure taken to detect affect from the simile expression.

1. The personal emotional profile of Mayid: ‘angry (1<sup>st</sup> input), angry (3<sup>rd</sup> input) and angry (5<sup>th</sup> input)’, as input to the Bayesian reasoning -> Mayid is most likely to be ‘angry’ again in the current 7<sup>th</sup> input;
2. The related social emotional context: ‘angry (4<sup>th</sup> input) and angry (6<sup>th</sup> input)’, as input to the neural network reasoning -> Mayid is most probable to be influenced to become ‘angry’.
3. Thus the simile input implies ‘anger’ other than being ‘neutral’.

Our AI agent can also sense other simile phenomena with similar syntactical structures and the affective states implied in them (“he walks like a lion”, “u stink like rotten meat” etc). The contextual affect detection based on personal and social cognitive emotion modeling has also been used to uncover and justify affect implied in other emotionally ambiguous metaphorical and literal input.

## 7. Evaluations and conclusions

As mentioned previously, the detected affective states from users' open-ended text input have also been used to produce emotional animation for human players' avatars. The emotional animation mainly includes expressive gestures and social attention (such as eye gazing). Thus, our processing has employed emotions embedded in the scenarios, dialogue and characters for expressive social animation generation without distracting users from the learning context. We also carried out user testing with 220 secondary school students from Birmingham and Darlington schools for the improvisation of school bullying and Crohn's disease scenarios. Generally, our previous statistical results based on the collected questionnaires indicate that the involvement of the AI character has not made any statistically significant difference to users' engagement and enjoyment with the emphasis of users' notice of the AI character's contribution throughout. Briefly, the methodology of the testing is that we had each testing subject have an experience of both scenarios, one including the AI minor character only and the other including the human-controlled minor character only. After the testing sessions, we obtained users' feedback via questionnaires and group debriefings. Improvisational transcripts were automatically recorded during the testing so that it allows further evaluation of the performance of the affect detection component.

We also produce a new set of results for the evaluation of the updated affect detection component with contextual and metaphorical affect interpretation based on the analysis of some recorded transcripts of the school bullying scenario. Generally two human judges marked up the affect of 400 turn-taking user input from the recorded 4 transcripts of this scenario (different from those used for the training of Bayesian and neural networks). In order to verify the efficiency of the new developments, we provide Cohen's Kappa inter-agreements for the AI agent's performance with and without the new developments for the detection of the most commonly used 10 affective states. The agreement for human judge A/B is 0.57. The inter-agreements between human judge A/B and the AI agent with the new developments are respectively 0.48 and 0.43, while the results between judge A/B and the agent without the new developments are only respectively 0.39 and 0.34.

Although future work is needed, the new developments on contextual affect sensing using both Bayesian and neural network based reasoning have improved the AI agent's performance comparing with the previous version. We have also provided evaluation results of the improvisational mood modeling using the Bayesian networks for the 3 leading characters in the school bullying scenario based on the analysis of the 4 testing transcripts. We have converted the recognized affective states into binary evaluation values and obtained the following accuracy rates shown in Table 2 by comparing with the annotation of one human judge.

|          | Mayid | Lisa | Elise |
|----------|-------|------|-------|
| Positive | 52%   | 46%  | 55%   |
| Negative | 94%   | 73%  | 86%   |
| Neutral  | 27%   | 35%  | 33%   |

Table 2. Accuracy rates of improvisational mood modeling using Bayesian networks.

Generally negative emotions are well detected across testing subjects. Since in the school bullying scenario, the big bully tends to make other characters suffer, the improvisation tends to be filled with negative emotional expressions such as threatening, angry and fear. Although positive and neutral expressions are recognized less well, the percentages of the inputs indicating positive and neutral expressions based on the human judges' interpretation are respectively approximate 30% and 25%. Thus although there is room for further improvements, the performances of affect sensing from positive and neutral expressions are acceptable.

Moreover, we also provide accuracy rates for the performance of the affect sensing in social interaction context using neural networks. Approximate 100 interaction contexts taken from the selected 4 example transcripts of the school bullying scenario are used for testing. We have also converted the recognized affective states into binary evaluation values and obtained 69% accuracy rate for positive emotions and 88% for negative emotions by comparing with the annotation of one human judge. The results indicate that other characters' emotional influence to the speaking character embedded in the interaction context is well recovered in our application using neural net based inference.

From the inspection of the evaluation results, although contextual affect detection based on both personal and social interaction context is provided, there are still some cases: when the two human judges both believed that user inputs carried negative or positive affective states, the AI agent regarded them as neutral. One most obvious reason is that sometimes Bayesian networks failed to predict some of the positive affective states (e.g. grateful) due to their low frequencies presented in the training data. Also affect sensing based on the analysis of individual turn-taking input sometimes failed to uncover the affect embedded in emotionally ambiguous input due to characters' creative improvisation which may affect the performance of contextual affect sensing. We also aim to extend the evaluation of the context-based affect detection using transcripts from other scenarios.

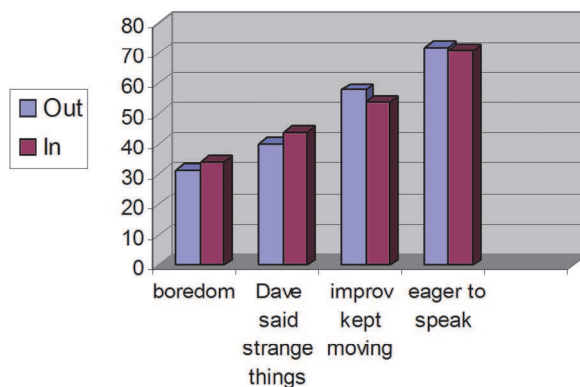


Fig. 4. Statistical results for 'boredom', 'Dave said strange things', 'improvisation kept moving' and 'eager to make own character speak' when EMMA is OUT (blue) or IN (purple) an improvisation.

Using a metaphorical resource (<http://knowgramming.com>), our approach for disease, cooking and sensory metaphor recognition obtains 50% average accuracy rate among the 80

testing examples. Also, we intend to use other resources (e.g. Wallstreet Journal and other metaphorical databases (e.g. ATT-Meta)) to further evaluate the metaphorical affect sensing. With a limited sample size of 40 simile examples extracted from the transcripts of school bullying and Crohn's disease scenarios, our approach for simile detection achieves 63% accuracy rate. The simile interpretation will also be further developed to accommodate more complex phenomena in future work.

Figure 4 also shows some evaluation results from a 'within-subjects' analysis looking at the difference made PER SUBJECT by having EMMA IN (= playing Dave, in either scenario) or OUT. When EMMA is out, the overall boredom is 31%. When EMMA is in, it changes to 34%. The results of 'human Dave and EMMA Dave said strange things' respectively are 40% and 44%. When EMMA changes from in to out of an improvisation, the results of 'improvisation kept moving' are respectively 54% to 58% and the results of 'the eagerness to make own character speak' are respectively 71% to 72%. Although the measures were 'worsened' by having EMMA in, in all cases the worsening was numerically fairly small and not statistically significant.

We have exploited emotion evolvment and prediction in personal and social context using the Bayesian reasoning and a supervised neural network. The conversational intelligent agent has also been equipped with the capabilities of metaphor & simile recognition and interpretation. Although the proposed approaches indicate initial exploration on context-based affect analysis and metaphor & simile inference, the implementation has enabled the AI agent to perform more effectively in affect detection tasks. In future work, we intend to employ emotion research of Hareli and Rafaeli (2008) and use Hidden Markov Models to further study and model how emotions evolve within individuals and in social communication context given various stimuli.

## 8. Appendix

Inspection of the transcripts collected indicates that EMMA usefully pushed the improvisation forward on various occasions. The following example transcript collected from the previous user testing shows how EMMA contributed to the drama improvisation in the Crohn's disease scenario. In the following interactions, Dave was played by EMMA.

DIRECTOR: are we ready to go to the restaurant? /\*Background changes to a 3D restaurant.\*/  
Janet Williams: this is nice  
Peter Williams: its a bit posh  
Arnold Williams: no one else is here  
Janet Williams: oh shut up  
Dave Atkins: Could we all tone down our language a bit? ppl r watching...  
Arnold Williams: dont boss me about wife  
Peter Williams: so i think its a good place to tell you im going to die unless i have an operation  
Janet Williams: i will thats why i married you  
Arnold Williams: peter talk about it another time  
Peter Williams: excuse me  
Janet Williams: no

... ..

Arnold Williams: your blood pressure will get too high

Peter Williams: im confused

Janet Williams: nobody cares what u think dad

Dave Atkins: Arnold, y u dont want 2 b involved in? Peter is ur son.

Arnold Williams: i just dont want to talk about it. i do care about him

DIRECTOR: remember, Peter, it's you're choice

Janet Williams: oh shut up

Peter Williams: its my choice

Arnold Williams: bossy

... ..

Peter Williams: I'M THE ILL ONE

Dave Atkins: Arnold, Peter needs ur support and u can't just ignore it.

Janet Williams: nobody cares peter

Arnold Williams: we know peter now stop talking about it

Peter Williams: yes i need your support

Dave Atkins: Hey guys, lets all calm down a bit.

Arnold Williams: we will discuss it later

Janet Williams: well you have got mien

Peter Williams: help me daddy

Arnold Williams: not now son

Peter Williams: well when

Janet Williams: he is not your daddy

Arnold Williams: another time

Peter Williams: i dont like u

Arnold Williams: wife i need your support on this

Dave Atkins: I think we all should respect Peter's decision.

Peter Williams: u should love me

Janet Williams: peter lets go home u have spolit it dad

Arnold Williams: i do son

Janet Williams: i hate u dad

Dave Atkins: wat??

Arnold Williams: lets talk about it another time

Dave Atkins: Guys, let's try 2 sort this out calmly.

Arnold Williams: thats not very nice mum

Peter Williams: yes calmly

DIRECTOR: ok 2 mins left, help peter make a decision

Peter Williams: what shall i do??

Janet Williams: ok if dad if you stop arguing then i will stop

Peter Williams: have the operation??

## 9. References

- Aylett, A., Louchart, S. Dias, J., Paiva, A., Vala, M., Woods, S. and Hall, L.E. (2006). Unscripted Narrative for Affectively Driven Characters. *IEEE Computer Graphics and Applications*. 26(3). 42-52.



- Barnden, J.A. Metaphor, semantic preferences and context-sensitivity. (2007). In K. Ahmad, C. Brewster & M. Stevenson (Eds), *Words and Intelligence II: Essays in Honor of Yorick Wilks*, 39-62.
- Briscoe, E. & Carroll, J. (2002). Robust Accurate Statistical Annotation of General Text. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Gran Canaria. 1499-1504.
- Egges, A., Kshirsagar, S. & Magnenat-Thalmann, N. (2003). A Model for Personality and Emotion Simulation, In *Proceedings of Knowledge-Based Intelligent Information & Engineering Systems (KES2003)*, Lecture Notes in AI. 453-461.
- Esuli, A. and Sebastiani, F. (2006). Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of EACL-06*, Trento, IT. 193-200.
- Fainsilber, L. and Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity*, 2(4):239-250.
- Fellbaum, C. (1998). *WordNet, an Electronic Lexical Database*. The MIT press.
- Fussell, S. and Moss, M. (1998). Figurative language in descriptions of emotional states. In *Social and Cognitive Approaches to Interpersonal Communication*, S. R. Fussell and R. J. Kreuz, Eds., Lawrence Erlbaum.
- Gratch, J. and Marsella, S. (2004). A Domain-Independent Framework for Modeling Emotion. *Journal of Cognitive Systems Research*. Vol 5, Issue 4, pp.269-306.
- Hareli, S. and Rafaeli, A. (2008). Emotion cycles: On the social influence of emotion in organizations. *Research in Organizational Behavior*, 28, 35-59.
- Kövecses, Z. (1998). Are There Any Emotion-Specific Metaphors? In *Speaking of Emotions: Conceptualization and Expression*. Athanasiadou, A. and Tabakowska, E. (eds.), Berlin and New York: Mouton de Gruyter, 127-151.
- Liu, H. & Singh, P. (2004). ConceptNet: A practical commonsense reasoning toolkit. *BT Technology Journal*, Volume 22, Kluwer Academic Publishers.
- Lopez, J.M., Gil, R., Garcia, R., Cearreta, I. and Garay, N. (2008). Towards an Ontology for Describing Emotions. In *Proceedings of the 1st world summit on The Knowledge Society: Emerging Technologies and Information Systems for the Knowledge Society*.
- Mateas, M. (2002). Interactive Drama, Art and Artificial Intelligence. Ph.D. Thesis. School of Computer Science, Carnegie Mellon University.
- Ortony, A., Clore, G.L. & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge U. Press.
- Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. And Araki, K. (2009). Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States. In *Proceedings of IJCAI*.
- Schnall, S. (2005). The pragmatics of emotion language. *Psychological Inquiry*, 16, 28-31.
- Shaikh, M.A.M., Prendinger, H. & Mitsuru, I. (2007). Assessing sentiment of text by semantic dependency and contextual valence analysis. In *Proceedings of ACII 2007*, 191-202.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition* (2nd ed.). Oxford, UK: Blackwell.
- Strapparava, C. and Valitutti, A. (2004). WordNet-Affect: An Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal, 1083-1086.

- Zhang, L. (2010). Exploitation on Contextual Affect Sensing and Dynamic Relationship Interpretation. In *ACM Computers in Entertainment*. Vol.8, Issue 3.
- Zhang, L. & Barnden, J.A. (2010). Affect and Metaphor Sensing in Virtual Drama. *International Journal of Computer Games Technology*. Vol. 2010.
- Zhang, L., Gillies, M., Dhaliwal, K., Gower, A., Robertson, D. & Crabtree, B. (2009). E-drama: Facilitating Online Role-play using an AI Actor and Emotionally Expressive Characters. *International Journal of Artificial Intelligence in Education*. Vol 19(1).
- Zhe, X. & Boucouvalas, A.C. (2002). Text-to-Emotion Engine for Real Time Internet Communication. In *Proceedings of International Symposium on Communication Systems, Networks and DSPs*, Staffordshire University, UK, 164-168.



## **Human Machine Interaction - Getting Closer**

Edited by Mr Inaki Maurtua

ISBN 978-953-307-890-8

Hard cover, 260 pages

**Publisher** InTech

**Published online** 25, January, 2012

**Published in print edition** January, 2012

In this book, the reader will find a set of papers divided into two sections. The first section presents different proposals focused on the human-machine interaction development process. The second section is devoted to different aspects of interaction, with a special emphasis on the physical interaction.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Li Zhang (2012). Affect Interpretation in Metaphorical and Simile Phenomena and Multithreading Dialogue Context, Human Machine Interaction - Getting Closer, Mr Inaki Maurtua (Ed.), ISBN: 978-953-307-890-8, InTech, Available from: <http://www.intechopen.com/books/human-machine-interaction-getting-closer/affect-interpretation-in-metaphorical-and-simile-phenomena-and-multithreading-dialogue-context>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.