

Metaphor Interpretation and Context-based Affect Detection

Li Zhang

School of Computing

Teesside University

l.zhang@tees.ac.uk

Abstract

Metaphorical and contextual affect detection from open-ended text-based dialogue is challenging but essential for the building of effective intelligent user interfaces. In this paper, we report updated developments of an affect detection model from text, including affect detection from one particular type of metaphorical affective expression and affect detection based on context. The overall affect detection model has been embedded in an intelligent conversational AI agent interacting with human users under loose scenarios. Evaluation for the updated affect detection component is also provided. Our work contributes to the conference themes on sentiment analysis and opinion mining and the development of dialogue and conversational agents.

1 Introduction

Affect sensing from open-ended text-based natural language input is a rising research area. Zhang et al. (2008a) reported an affect detection component on detecting simple and complex emotions, meta-emotions, value judgments etc from literal expressions. Recently, metaphorical language has drawn researchers' attention since it has been widely used to provide effective vivid description. 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". In Wallington et al's work (2008), several metaphorical affective expressions (such as animal metaphor ("X is a

rat") and affects as external entities metaphor ("joy ran through me")) have been intensively studied and affect has been derived from some simple animal metaphorical expressions.

The work presented here reports developments on affect detection from one particular comparatively complex metaphorical phenomenon with affect implication, i.e. the cooking metaphor ("the lawyer grilled the witness on the stand", "I knew I was cooked when the teacher showed up at the door") (http://knowgramming.com/cooking_metaphors.htm). Since context plays an important role in the interpretation of the affect conveyed by the user during the interaction, we have used linguistic contextual analysis and cognitive emotional modeling based on Markov chain modeling and a dynamic algorithm to interpret affect from context in our application.

Our developments have been incorporated into an affect detection component, which can detect affect and emotions from literal text input and has been embedded in an intelligent conversational agent, engaged in a drama improvisation with human users under loose scenarios (school bullying and Crohn's disease). The conversational AI agent also provides appropriate responses based on the detected affect from users' input in order to stimulate the improvisation. In both scenarios, the AI agent plays a minor role in drama improvisation. E.g. it plays a close friend of the bullied victim (the leading role) in school bullying scenario, who tries to stop the bullying.

We have also analyzed affect detection performance based on previously collected (other) transcripts from user testing by calculating agreements via Cohen's Kappa between two human judges and between human judges and the AI agent with and without the new devel-

opment respectively in order to verify the efficiency of the metaphorical and contextual affect sensing.

The content is arranged as follows. We report relevant work in section 2 and the new developments on affect detection from the cooking metaphor in section 3. Contextual affect sensing is discussed in section 4. System evaluation and conclusion are presented in section 5.

2 Related Work

There is well-known research work in the related fields. ConceptNet (Liu and Singh, 2004) is a toolkit to provide practical textual reasoning for affect sensing for six basic emotions, text summarization and topic extraction. Shaikh et al. (2007) provided sentence-level textual affect sensing to recognize evaluations (positive and negative). They adopted a rule-based domain-independent approach, but they haven't made attempts to recognize different affective states from open-ended text input.

Although Façade (Mateas, 2002) included shallow natural language processing for characters' open-ended utterances, the detection of major emotions, rudeness and value judgements is not mentioned. Zhe and Boucouvalas (2002) demonstrated an emotion extraction module embedded in an Internet chatting environment. It used a part-of-speech tagger and a syntactic chunker to detect the emotional words and to analyze emotion intensity for the first person (e.g. 'I'). The detection focused only on emotional adjectives and first-person emotions, and did not address deep issues such as figurative expression of emotion. There is also work on general linguistic cues useful for affect detection (e.g. Craggs and Wood, 2004).

In addition, there is well-known research work on the development of emotional conversational agents. Egges et al. (2003) provided virtual characters with conversational emotional responsiveness. Aylett et al. (2006) also focused on the development of affective behavior planning for their synthetic characters. Cavazza et al. (2008) reported on a conversational agent embodied in a wireless robot to provide suggestions for users on a healthy living life-style. Hierarchical Task Networks (HTN) planner and semantic interpretation have been used in this work. The cognitive planner plays an important

role in assisting with dialogue management. The user's response has also been considered for the generation of a new plan. However, the system will hesitate when open-ended user input going beyond the planner's knowledge has been used intensively during interaction. The system we present here intends to deal with such challenge.

Our work focuses on the following aspects: (1) affect detection from metaphorical expressions; (2) real-time 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 context profiles.

3 Further Development on Metaphorical Affect Detection

Without pre-defined constrained scripts, our original system has been developed for 14-16 year old school students to conduct creative improvisation within highly emotionally charged scenarios. Various metaphorical expressions were used to convey emotions (Kövecses, 1998), which are theoretically and practically challenging and draw our attention.

Metaphorical language can be used to convey emotions implicitly and explicitly, which also inspires cognitive semanticists (Kövecses, 1998). In our previous study (Zhang et al. 2008b; 2009), we detected affect from several comparatively simple metaphorical affective phenomena. Another type of comparatively complex metaphor has also drawn our attention, i.e. the cooking metaphor. Very often, the agent himself/herself would become the victim of slow or intensive cooking (e.g. grilled, cooked). Or one agent can perform cooking like actions towards another agent to realize punishment or torture. Examples are as follows, "he basted her with flattery to get the job", "she knew she was fried when the teacher handed back her paper".

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 carried out by other agents 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. The syntactic structures of some of the above exam-

ples also indicate the submissive stance of the suffering agents. E.g. in the instances, passive sentences (“he knew he was cooked when he saw his boss standing at the door”) have been used to imply unwillingness and victimization of the subject agents who are in fact the objects of the cooking actions described by the verb phrases (“X + copular form + passive cooking action”). In other examples, the cooking actions have been explicitly performed by the subject agents towards the object agents to imply the former’s potential willingness and enjoyment and the latter’s potential suffering and pain (“A + [cooking action] + B”).

Thus in our application, we focus on the above two particular types of expressions. We use Rasp (Briscoe & Carroll, 2002) to recognize user input with such syntactic structures (‘A + copular form + VVN’, ‘A + VV0/VVD/VVZ (verb) + B’). Many sentences could possess such syntactic structures (e.g. “Lisa was bullied”, “he grills Lisa”, “I was hit by a car”, “Lisa was given the task to play the victim role”, “I steamed it” etc), but few of them are cooking metaphors. Therefore we need to resort to semantic profiles to recognize the metaphorical expressions. Rasp has also provided a syntactic label for each word in the user input. Thus the main verbs were identified by their corresponding syntactic labels (e.g. ‘given’ labeled as ‘past participle form of lexical verbs (VVN)’, ‘likes’ and ‘grills’ labeled as ‘-s form of lexical verbs (VVZ)’ and the semantic interpretation for their base forms is discovered from WordNet (Fellbaum, 1998). Since WordNet has provided hypernyms (Y is a hypernym of X if every X is a (kind of) Y) for the general noun and verb lexicon, ‘COOK’ has been derived as the hypernym of the verbs’ described cooking actions. E.g. ‘boil’, ‘grill’, ‘steam’, and ‘simmer’ are respectively interpreted as one way to ‘COOK’. ‘Toast’ is interpreted as one way to ‘HEAT UP’ while ‘cook’ is interpreted as one way to ‘CREAT’, or ‘CHEAT’ etc. One verb may recover several hypernyms and in our application, we collect all of them. Another evaluation resource (Esuli and Sebastiani, 2006) is resorted to in order to recover the evaluation values of all the hypernyms for a particular verb. If some hypernyms are negative (such as ‘CHEAT’) and the main object of the overall input refers to first/third person cases or singu-

lar proper nouns (‘him’, ‘her’, or ‘Lisa’), then the user input (e.g. “he basted her with flattery to get the job”) conveys potential negative affect (e.g. pain and sadness) for the human objects and potential positive affect (e.g. persuasion or enjoyment) for the subjects. If the evaluation dictionary fails to provide any evaluation value for any hypernyms (such as ‘COOK’ and ‘HEAT UP’) of the main verbs, then we still assume that ‘verbs implying COOK/HEAT UP + human objects’ or ‘human subjects + copular form + VVN verbs implying COOK/HEAT UP’ may indicate negative emotions both for the human objects in the former and the human subjects in the latter. E.g. for the input “I was fried by the head teacher”, the processing is as follows:

1. Rasp identifies the input has the following structure: ‘PPIS1 (I) + copular form (was) + VVN (fried)’;
2. ‘Fry’ (base form of the main verb) is sent to WordNet to obtain its hypernyms, which include ‘COOK’, ‘HEAT’ and ‘KILL’;
3. The input has the following syntactic semantic structure: ‘PPIS1 (I) + copular form (was) + VVN (Hypernym: COOK)’, thus it is recognized as a cooking metaphor;
4. The three hypernyms are sent to the evaluation resource to obtain their evaluation values. ‘KILL’ is labeled as negative while others can’t obtain any evaluation values from the profile;
5. The input is transformed into: ‘PPIS1 (I) + copular form (was) + VVN (KILL: negative)’
6. The subject is a first person case, then the input indicates the user who is speaking suffered from a negative action and may have a ‘negative’ emotional state.

Although our processing is limited to the verb metaphor examples and hasn’t considered other instances like “tasty tidbits of information”, it points out promising directions for figurative language processing. After our intention to improve the performance of affect sensing from individual turn-taking input, we focus on improvement of the performance using context profiles. In future work, we intend to use a metaphor ontology to recognize metaphors.

4 Affect Sensing from Context Profiles

Our previous affect detection (Zhang et al. 2008a) has been performed solely based on in-

dividual turn-taking input. Thus the context information has been ignored. However, the contextual and character profiles may influence the affect implied in the current input. In this section, we will discuss relationships between characters, linguistic contextual indicators, cognitive emotion simulation from a communication context and our approach developed based on these features to interpret affect from context.

4.1 Relationship Interpretation

Relationships between characters in drama improvisation are very crucial for the contextual affect interpretation for the emotionally ambiguous users' input. During the improvisation of each scenario, like any other drama progression, normally the recorded transcripts for creative roleplays are composed of three main improvisational sections, including the starting of the drama, the climax and the final ending. Relationships in these three drama progression stages between characters are different from one another. E.g. in the climax of the improvisation of the school bullying scenario, we normally expect very negative relationships between the bully and the bullied victim (Lisa) & her friends since the big bully is very aggressive at Lisa and her friends who try to stop the bullying. Moreover, in nearly the end of the improvisational session, sometimes the big bully feels sorry for his behavior and is cared by Lisa and her friends since he is abused by his uncle. The intense negative relationships between the big bully and Lisa & her friends are changed to those with at least less negativity or even normal relationships. Because of the creative nature of the improvisation, sometimes the bully and the victim may even have a positive relationship towards the ending of the drama improvisation.

However in our current study, we only assume consistent negative relationships between the bully and the bullied victim & her friends throughout the improvisation to simplify the processing. We will report our work on relationship interpretation using fuzzy logic to dynamically capture the changing relationships between characters as the drama progresses in the near future.

4.2 Linguistic Contextual Indicators

In our study, we noticed some linguistic indicators for contextual communication in the rec-

orded transcripts. One useful indicator is (i) imperatives, which are often used to imply negative or positive responses to the previous speaking characters, such as "shut up", "go on then", "let's do it" and "bring it on". Other useful contextual indicators are (ii) prepositional phrases (e.g. "by who?"), semi-coordinating conjunctions (e.g. "so we are good then"), subordinating conjunctions ("because Lisa is a dog") and coordinating conjunctions ('and', 'or' and 'but'). These indicators are normally used by the current 'speaker' to express further opinions or gain further confirmation.

In addition, (iii) short phrases for questions are also used frequently in the transcripts to gain further communication based on context, e.g. "where?", "who is Dave" or "what". (iv) Character names are also normally used in the user input to indicate that the current input is intended for particular characters, e.g. "Dave go away", "Mrs Parton, say something", "Dave what has got into you?" etc. Very often, such expressions have been used to imply potential emotional contextual communication between the current speaking character and the named character. Therefore the current speaking characters may imply at least 'approval' or 'disapproval' towards the opinions/comments provided by the previous named speaking characters. Finally there are also (v) some other well known contextual indicators in Internet relay chat such as 'yeah/yes followed by a sentence ("yeah, we will see")', "I think so", 'no/nah followed by a sentence', "me too", "exactly", "thanks", "sorry", "grrrr", "hahahaha", etc. Such expressions are normally used to indicate affective responses to the previous input.

Since natural language is ambiguous and there are cases in which contextual information is required in order to appropriately interpret the affect conveyed in the input (e.g. "go on then"), our approach reported in the following integrates the above contextual linguistic indicators with cognitive contextual emotion prediction to uncover affect conveyed in emotionally ambiguous input.

4.3 Emotion Modeling in Communication Context

There are also other aspects which may influence the affect conveyed in the communication context. E.g. in our application, the affect con-

veyed by the speaking character himself/herself in the recent several turn-taking, the 'improvisational mood' that the speaking character is created, and emotions expressed by other characters, especially by the contradictory ones (e.g. the big bully), have great potential to influence the affect conveyed by the current speaking character (e.g. the bullied victim). Sometimes, the story themes or topics also have potential impact to emotions or feelings expressed by characters. For example, people tend to feel 'happy' when involved in discussions on positive topics such as harvest or raising salary, while people tend to feel 'sad' when engaged in the discussions on negative themes such as economy breakdown, tough examination etc.

In our application, although the hidden story sub-themes used in the scenarios are not that dramatic, they are still highly emotionally charged and used as the signals for potential changes of emotional context for each character. E.g. In the school bullying scenario (which is mainly about the bully, Mayid, is picking on the new comer to the school, Lisa. Lisa's friends, Elise and Dave, are trying to stop the bullying. The school teacher, Mrs Parton, also tries to find out what is going on), the director mainly provided interventions based on several main sub-themes of the story to push the improvisation forward, i.e. "Mayid starts bullying Lisa", "why Lisa is crying", "why Mayid is so nasty/a bully", "how Mayid feels when his uncle finds out about his behavior" etc. From the inspection of the recorded transcripts, when discussing the topic of "why Lisa is crying", we noticed that Mayid (the bully) tends to be really aggressive and rude, while Lisa (the bullied victim) tends to be upset and other characters (Lisa's friends and the school teacher) are inclined to show anger at Mayid. For the improvisation of the hidden story sub-theme "why Mayid is so nasty/a bully", the big bully changes from rude and aggressive to sad and embarrassed (e.g. because he is abused by his uncle), while Lisa and other characters become sympathetic (and sometimes caring) about Mayid. Usually all characters are trying to create the 'improvisational mood' according to the guidance of the hidden story sub-themes (provided via director's intervention). Therefore, the story sub-themes could be used as the indicators for potential emotional context change. The emotion patterns expressed by each

character within the improvisation of each story sub-theme could be very useful for the prediction of the affect shown in a similar topic context, although the improvisation of the characters is creative within the loose scenario. It will improve the performance of the emotional context prediction if we allow more emotional profiles for each story sub-theme to be added to the training data to reflect the creative improvisation (e.g. some improvisations went deeper for a particular topic).

Therefore, a Markov chain is used to learn from the emotional context shown in the recorded transcripts for each sub-theme and for each character, and generate other possible reasonable unseen emotional context similar to the training data for each character. Markov chains are usually used for word generation. In our application, they are used to record the frequencies of several emotions showing up after one particular emotion. A matrix has been constructed dynamically for neutral and the 12 most commonly used emotions in our application (caring, arguing, disapproving, approving, grateful, happy, sad, threatening, embarrassed, angry/rude, scared and sympathetic) with each row representing the previous emotion followed by the subsequent emotions in columns. The Markov chains employ roulette wheel selection to ensure to produce a greater probability to select emotions with higher frequencies than emotions with lower occurrences. This will allow the generation of emotional context to probabilistically follow the training data, which may reflect the creative nature of the improvisation.

Then a dynamic algorithm is used to find the most resembling emotional context for any given new situation from the Markov chain's training and generated emotional contexts. I.e. by using the algorithm, a particular series of emotions for a particular story sub-theme has been regarded as the most resembling context to the test emotional situation and an emotional state is recommended as the most probable emotion for the current user input. Since the most recent affect histories of other characters and relationships between characters may also have an impact on the affect conveyed by the speaking character, the recommended affect will be further evaluated (e.g. a most recent 'insulting' input from Mayid could make Lisa 'angry').

At the training stage, first of all, the school bullying transcripts collected from previous user testing have been divided into several topic sections with each of them belonging to one of the story sub-themes. The classification of the sub-themes is mainly based on the human director's intervention which was recorded in the transcripts. Then we used two human annotators to mark up the affect of every turn-taking input in the transcripts using context inference. Thus, for each character, we have summarized a series of emotions expressed throughout the improvisation of a particular story sub-theme. Since the improvisation is creative under the loose scenario, some of the sub-themes (e.g. "why Mayid is so nasty") have been suggested for improvisation for one than once in some transcripts and some of the topics (e.g. "why Lisa is crying") are only shown in a few of the collected transcripts. We made attempts to gather as many emotional contexts as possible for each character for the improvisation of each sub-theme in order to enrich the training data.

The following is a small portion of one recorded transcript used for the training of the Markov chain. The human annotators have marked up the affect expressed in each turn-taking input.

DIRECTOR: why is Lisa crying?

Elise Brown [caring]: lisa stop cryin

Lisa Murdoch [disagree]: lisa aint crying!!!!

Dave Simons [caring]: i dunno! y u cryin lisa?

Mayid Rahim [rude]: cuz she dnt realise she is lucky to b alive

Elise Brown [angry]: beat him up! itss onlii fat..he'll go down straight away

Mayid Rahim [insulting]: lisa, y u crying? u big baby!

Mrs Parton [caring]: lisa, r u ok?

For example, the emotional context for Mayid from the above example is: 'rude' and 'insulting' (we use one letter to represent each emotional label, thus in this example, i.e. 'R I'), and in the similar way, the emotional contexts for other characters have been derived from the above example, which are used as the training data for the Markov chain for the topic "why Lisa is crying". We have summarized the emotional context for each story sub-theme for each character from 4 school bullying transcripts and used them for the training of the Markov chain.

The topics considered at the training stage include: "Mayid starts bullying", "why is Lisa crying", "why is Mayid nasty/a bully" and "how does Mayid feel if his uncle knew about his behavior?"

At the test stage, our affect detection component, EMMA, is integrated with an AI agent and detects affect for each user input solely based on the analysis of individual turn-taking input itself. The above algorithms for context-based affect sensing will be activated when the affect detection component recognizes 'neutral' from the current input during the emotionally charged proper improvisation after all the characters have known each other and went on the virtual drama stage. First of all, the linguistic indicators are used to identify if the input with 'neutral' implication is a contextual-based input. E.g. we mainly focus on the checking of the five contextual implications we mentioned previously, including imperatives, prepositional phrases, conjunctions, simplified question sentences, character names, and other commonly used contextual indicators (e.g. "yeah", "I think so"). If any of the above contextual indicators exists, then we further analyze the affect embedded in the input with contextual emotion modeling reported here.

For example, we have collected the following transcript for testing. Normally the director intervened to suggest a topic change (e.g. "find out why Mayid is a bully"). Thus for a testing situation for a particular character, we use the emotion context attached with his/her user input starting right after the most recent director's intervention and ending at his/her last second input, since such a context may belong to one particular topic.

DIRECTOR: U R IN THE PLAYGROUND (indicating bullying starts)

1. Lisa Murdoch: leave me alone! [angry]
2. Mayid Rahim: WAT U GONNA DU? [neu] -> [angry]
3. Mayid Rahim: SHUT UR FAT MOUTH [angry]
4. Elise Brown: grrrrr [angry]
5. Elise Brown: im telin da dinna lady! [threatening]
6. Mayid Rahim: go on den [neutral] -> [angry]
7. Elise Brown: misssssssssssssss [neu]
8. Elise Brown: lol [happy]

9. Lisa Murdoch: mayid u gna gt banned [threatening]

10. Mayid Rahim: BY HU [neu] -> [angry]

The affect detection component detected that Lisa was ‘angry’ by saying “leave me alone!”. It also sensed that Mayid was ‘neutral’ by saying “WAT U GONNA DU (what are you going to do)?” without consideration of context. From Rasp, we obtained that the input is a simplified question sentence (a linguistic contextual indicator). Thus, it implies that it could be an emotional situation caused by the previous context (e.g. previous input from Lisa) and the further processing for emotion prediction is activated. Since we don’t have an emotional context yet at this stage for Mayid (the very first input from Mayid after the director’s intervention), we cannot resort to the Markov chain and the dynamic algorithm currently to predict the affect. However, we could use the emotional context of other characters to predict the affect for Mayid’s current input since we believe that an emotional input from a character, especially from an opponent character, has great potential to affect the emotions expressed by the current speaking character.

In the most recent chat history, there is only one input from Lisa after the director’s intervention, which implied ‘anger’. Since Lisa and Mayid have a negative relationship (pre-defined by character profiles), then we predict Mayid currently experiences negative emotion. Since capitalizations have been used in Mayid’s input, we conclude that the affect implied in the input could be ‘angry’. However, EMMA could be fooled if the affect histories of other characters fail to provide any useful indication for prediction (e.g. if Lisa implied ‘neutral’ in the most recent input, the interpretation of the affect conveyed by Mayid would be still ‘neutral’).

EMMA also detected affect for the 3rd, 4th, and 5th user input in the above example (based on individual turn-taking) until it detected ‘neutral’ again from the 6th input “go on den (go on then)” from Mayid. Since it is an imperative mood sentence (a linguistic contextual indicator), the input may imply a potential (emotional) response to the previous speaking character. Since we couldn’t obtain the affect embedded in the imperative purely based on the analysis of the input itself, the contextual processing is required. Thus the emotional context profile for

Mayid is retrieved, i.e. [angry (the 2nd input) and angry (the 3rd input)]. The Markov chain is used to produce the possible emotional context based on the training data for each sub-theme for Mayid.

The following are generated example emotional profiles for the sub-theme “Mayid starts bullying” for the Mayid character:

1. T A A N A A [‘threatening, angry, angry, neutral, angry and angry’]

2. N A A A [‘neutral, angry, angry, and angry’]

3. D A I A A N A [‘disapproval, angry, insulting, angry, angry, angry, neutral, and angry’]

4. I A A N [‘insulting, angry, angry and neutral’]

The dynamic algorithm is used to find the smallest edit distance between the test emotional context [angry and angry] and the training and generated emotional context for the Mayid character for each sub-theme. In the above example, the second and fourth emotional sequences have the smallest edit distance (distance = 2) to the test emotional context and the former suggests ‘angry’ as the affect conveyed in the current input (“go on den”) while the latter implies ‘neutral’ expressed in the current input. Thus we need to resort to the emotional context of other characters to justify the recommended affects. From the chatting log, we find that Lisa was ‘angry’ in her most recent input (the 1st input) while Elise was ‘threatening’ in her most recent input (the 5th input). Since the bully, Mayid, has a negative relationships with Lisa (being ‘angry’) and Elise (being ‘threatening’), the imperative input (“go on den”) may indicate ‘angry’ rather than ‘neutral’. Therefore our processing adjusts the affect from ‘neutral’ to ‘angry’ for the 6th input.

In this way, by considering the linguistic contextual indicators, the potential emotional context one character was in, relationships with others and recent emotional profiles of other characters, our affect detection component has been able to inference emotion based on context to mark up the rest of the above test example transcript (e.g. Mayid being ‘angry’ for the 10th input). However our processing could be fooled easily by various diverse ways for affective expressions and creative improvisation (test emotional patterns not shown in the training and

generated sets). We intend to adopt better emotion simulation tools, more linguistic hints, psychological (context-based) emotional theories for further improvements. Also, our processing currently only focused on the school bullying scenario. We are on our way to extend the context-based affect sensing to the Crohn’s disease scenario to further evaluate its efficiency.

5 Evaluation and Conclusion

We 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.

Therefore, we produce a new set of results for the evaluation of the updated affect detection component with metaphorical and context-based affect detection based on the analysis of some recorded transcripts of school bullying scenario. Generally two human judges (not engaged in any development stage) marked up the affect of 150 turn-taking user input from the recorded another 4 transcripts from school bullying scenario (different from those used for the training of Markov chains). In order to verify the efficiency of the new developments, we provide Cohen’s Kappa inter-agreements for EMMA’s performance with and without the new developments for the detection of the most commonly used 12 affective states. In the school bullying scenario, EMMA played a minor bit-part character (Lisa’s friend: Dave). The agreement for human judge A/B is 0.45. The inter-agreements between human judge A/B and

EMMA with and without the new developments are presented in Table 1.

	Human Judge A	Human Judge B
EMMA (previous version)	0.38	0.30
EMMA (new version)	0.40	0.32

Table 1: Inter-agreements between human judges and EMMA with and without the new developments

Although further work is needed, the new developments on metaphorical and contextual affect sensing have improved EMMA’s performance of affect detection in the test transcripts comparing with the previous version.

The evaluation results indicated that most of the improvements (approximately 80%) are obtained for negative affect detection based on the inference of context information. But there are still some cases: when the two human judges both believed that user inputs carried negative affective states (such as angry, threatening, disapproval etc), EMMA regarded them as neutral. One most obvious reason is that some of the previous pipeline processing (such as dealing with mis-spelling, acronyms etc, and syntactic processing from Rasp etc) failed to recover the standard user input or recognize the complex structure of the input which led to less interesting and less emotional context and may affect the performance of contextual affect sensing. (The work of Sproat et al. (2001) can point out helpful directions on this aspect.) Currently we achieved 69% average accuracy rate for the contextual affect sensing for the emotion interpretation of all the human controlled characters in school bullying scenario. We also aim to extend the evaluation of the context-based affect detection using transcripts from other scenarios. Moreover, some of the improvements (nearly 20%) in the updated affect sensing component are made by the metaphorical processing. However, since the test transcripts contained a very small number of metaphorical language phenomena comparatively, we intend to use other resources (e.g. The Wall Street Journal and other metaphorical databases (such as ATT-Meta, 2008)) to further evaluate the new development on metaphorical affect sensing.

References

- ATT-Meta Project Databank: Examples of Usage of Metaphors of Mind. 2008. <http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/>.
- 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.
- 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.
- Cavazza, M., Smith, C., Charlton, D., Zhang, L., Turunen, M. and Hakulinen, J. 2008. A 'Companion' ECA with Planning and Activity Modelling. In *Proceedings of the 7th International Conference on Autonomous Agents and Multi-Agent Systems*. Portugal, 1281-1284.
- Craggs, R. & Wood, M. 2004. A Two Dimensional Annotation Scheme for Emotion in Dialogue. In *Proceedings of AAAI Spring Symposium: Exploring Attitude and Affect in Text*.
- 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. Springer-Verlag: Berlin, 453-461.
- Esuli, A. and Sebastiani, F. 2006. Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics*, 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.
- 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.
- Mateas, M. 2002. Interactive Drama, Art and Artificial Intelligence. Ph.D. Thesis. School of Computer Science, Carnegie Mellon University.
- Rayson, P. 2003. Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. Ph.D. thesis, Lancaster University.
- Shaikh, M.A.M., Prendinger, H. & Mitsuru, I. 2007. Assessing sentiment of text by semantic dependency and contextual valence analysis. In *Proceeding of AII 2007*, 191-202.
- Sproat, R., Black, A., Chen, S., Kumar, S., Ostendorf, M. and Richards, C. 2001. Normalization of Non-standard Words. *Computer Speech and Language*, 15(3), 287-333.
- Wallington, A.M., Agerri, R., Barnden, J.A., Lee, M.G. & Rumbell, T. 2008. Affect Transfer by Metaphor for an Intelligent Conversational Agent. In *Procs of LREC 2008 Workshop on Sentiment Analysis: Emotion, Metaphor, Ontology and Terminology*, pp.107-113. Morocco.
- Zhang, L., Barnden, J.A. Hendley, R.J. Lee, M.G. Wallington, A.M. and Wen, Z. 2008a. Affect Detection and Metaphor in E-drama. *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 18, No. 2, 234-252.
- Zhang, L., Gillies, M. & Barnden, J.A. 2008b. EMMA: an Automated Intelligent Actor in E-drama. In *Proceedings of International Conference on Intelligent User Interfaces*. 13th -16th Jan 2008. Canary Islands, Spain. pp. 409-412.
- 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), pp.5-38.
- 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.