Proceedings of the Linguistic And Cognitive Approaches To Dialog Agents Symposium, Rafal Rzepka (Ed.), at the AISB 2010 convention, 29 March – 1 April 2010, De Montfort University, Leicester, UK

Metaphorical and Contextual Affect Detection in an Intelligent Agent

Li Zhang[1]

Abstract. Real-time 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 two particular types of metaphorical affective expressions (food and cooking metaphors) 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 symposium's themes on human-agent interaction based on affect sensing, affective computing and daily life dialog systems.

1 INTRODUCTION

We have reported an affect detection component, EMMA (emotion, metaphor and affect) on detecting simple and complex emotions, meta-emotions, value judgments etc from literal expressions in English in our previous work [1]. Recently, metaphorical language has drawn our attention since it has been widely used to provide effective vivid description. Fainsilber and Ortony [2] 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 our previous application, we have also identified several metaphorical affective expressions from the collected transcripts (such as animal metaphor ("X is a rat") and affects as external entities metaphor ("joy ran through me") [1, 3, 4]) and conducted initial development on affect sensing from these metaphorical phenomena.

The work presented here reports further developments on affect detection from another two particular English metaphorical expressions with affect implication, which include food metaphor ("X is walking meat", "X has a pizza face") and cooking metaphor ("the lawyer grilled the witness on the stand", "I knew I was cooked when the teacher showed up at the door") based on the syntactical analysis of metaphorical phenomena shown in previously collected transcripts in our study and another metaphor online dictionary[2]. The transcripts used in our work were previously created by the testing subjects (secondary school students) in previous user study [1]. Also, since context plays an important role in the interpretation of the affect conveyed by the user during the interaction, we have used Markov chain modelling (trained by the collected previous transcripts) and dynamic programming to recommend the most possible affect conveyed in the current input when the affect detection component interprets 'neutral' solely based on the analysis of the input itself during the proper improvisation (in our study, we use 'proper improvisation' to indicate that users stayed in their characters for creative role-playing after they became familiar with the chosen scenarios and each character).

The new developments have been incorporated into the affect detection component, which has been embedded in an intelligent conversational agent, engaged in a drama improvisation with human users under loose scenarios (e.g. school bullying and Crohn's disease designed for our application). The affective states detected by this component have been used to produce expressive behaviours for human-controlled characters. 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. It plays a close friend of the bullied victim (the leading role) in school bullying scenario, who tries to stop the bullying and a close friend of the sick leading character in Crohn's disease scenario who tries to give support to his friend with the decision on his life-changing operation.

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, human judges and the AI agent with and without the new development respectively in order to verify the efficiency of the metaphorical and contextual affect sensing.

The content is arranged in the following way. We report relevant work in section 2 and the new developments on affect detection from food and cooking metaphor in section 3. Affect interpretation based on context is discussed in section 4. Newly produced evaluation results of the affect detection component are reported in section 5. Finally we summarize our work and point out future directions for further developments in section 6.

2 RELATED WORK

Textual affect sensing is a rising research branch for natural language processing. ConceptNet [5] is a toolkit to provide practical textual reasoning for affect sensing for six basic emotions, text summarization and topic extraction. Shaikh et al. [6] 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* [7], a one act interactive drama, included shallow natural language processing for characters' open-ended

[1] School of Computing, Teesside University, TS1 3BA, UK. Email: {l.zhang}@tees.ac.uk.

[2] http://knowgramming.com/cooking_metaphors.htm

utterances, the detection of major emotions, rudeness and value judgements is not mentioned. Zhe and Boucouvalas [8] demonstrated an emotion extraction module embedded in an Internet chatting environment (see also Boucouvalas [9]). It used a part-of-speech tagger and a syntactic chunker to detect the emotional words and to analyse emotion intensity for the first person (e.g. 'I' or 'we'). Unfortunately the emotion detection focused only on emotional adjectives, and did not address deep issues such as figurative expression of emotion (discussed below). Also, the concentration purely on first-person emotions is narrow. There has been relevant work on general linguistic cues that could be used in practice for affect detection (e.g. Craggs and Wood [10]).

There is also well-known research work on the development of emotional conversational agents. Egges et al. [11] have provided virtual characters with conversational emotional responsiveness. Elliott et al. [12] demonstrated tutoring systems that reason about users' emotions. They believe that motivation and emotion play very important roles in learning. Virtual tutors have been created in a way that not only having their own emotion appraisal and responsiveness, but also understanding users' emotional states according to their learning progress. Aylett et al. [13] also focused on the development of affective behaviour planning for the synthetic characters. Cavazza et al. [14] reported a conversational agent embodied in a wireless robot to provide suggestions for users on a healthy living life-style. Hierarchical Task Networks planner and semantic interpretation have been used in this work. The cognitive planner plays an important role in assisting with dialogue management, e.g. giving suggestions to the dialogue manager on what relevant questions should be raised to the user according to the healthy living plan currently generated. The user's response has also been adopted by the cognitive planner to influence the change of the current plan. The limitation of such planning systems is that they normally work reasonably well within the pre-defined domain knowledge, but they will strike when open-ended input going beyond the planner's knowledge has been used intensively during interaction. The system presented here intends to deal with some of such challenges.

Our work is distinctive in the following aspects: (1) affect detection from metaphorical expressions; (2) real-time affect sensing for basic and complex emotions, meta-emotions, value judgments etc in improvisational role-play situations from open-ended user input; (3) affect detection for second and third person cases ('you', 'she', 'they') (since English belongs to low context languages, it would allow us to interpret affect conveyed by the second and third person cases. However, in high context languages (e.g. Japanese), the interpretation of emotions conveyed by others rather than the experiencing subject will only be regarded as an opinion); and (4) affect interpretation based on context profiles.

3 FURTHER DEVELOPMENT ON METAPHORICAL AFFECT DETECTION

Before we introduce the new developments on affect detection from metaphor and context profiles, we introduce our previous work on affect detection and responding strategy for the AI agent briefly. Our original system has been developed for age 14-16 secondary school students to engage in role-play situations under loose scenarios 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 – school bullying and Crohn's disease. The AI agent could be activated to interact with human actors by playing a minor bit-part character in both scenarios as mentioned earlier. We have used responding regimes for the conversational AI agent in order to stir up the discussion and stimulate the improvisation, e.g. making responses when its confidence on interpreting affect from users' input is high. The language used by the secondary school students during their role-play is highly diverse with various online chatting features, such as abbreviations (e.g. 'den' (then), 'r' (are)), acronyms (e.g. 'lol' (laughing out loud)) and slang. Our previous work had pre-processing procedures to deal with abbreviations, acronyms, misspellings and slang [1]. Diverse metaphorical language phenomena have also been used to express emotions during the interaction, which challenge any NLP system greatly which intends to pursue logical semantic and syntactic analysis or affect/opinion mining and draw our attention for further study.

Affect detection from food metaphor

Food has been used extensively as metaphor for social position, group identity, religion, etc. E.g. food could be used as a metaphor for national identity. British have been called 'roastbeefs' by the French, while French have been referred to as 'frogs' by the British. It has also been used to indicate social hierarchy. E.g. in certain Andean countries, potatoes have been used to represent poor rural farmers of native American descent and white flour and bread have been used mainly to refer to wealthy European descent[3]. In our school bullying scenario, the big bully has called the bullied victim (Lisa) names, such as "u r a pizza", "Lisa has a pizza face" to exaggerate that fact that the victim has acne. Another most commonly used food metaphor is to use food to refer to a specific shape. E.g. body shape could be described as 'banana', 'pear' and 'apple'[4]. In our application, "Lisa has a pizza face" could also be interpreted as Lisa has a 'round (shape)' face. Therefore, insults could be conveyed in such food metaphorical expression. We especially focus on the statement of 'second-person/a singular proper noun + present-tense copular form + food term' to extract affect. A special semantic dictionary has been created by providing semantic tags to normal English lexicon. The semantic tags have been created by using Wmatrix [15], which facilitates the user to obtain corpus annotation with semantic and part-of-speech tags to compose dictionary. The semantic dictionary created consists mainly of food terms, animal names, measureable adjectives (such as size) etc with their corresponding semantic tags due to the fact they have the potential to convey affect and feelings.

In our application, Rasp, a statistical-based domain-independent robust parsing system for English, informs the system the user input with the desired structure - 'second-person/a singular proper noun + present-tense copular form + noun phrases' (e.g. "Lisa is a pizza", "u r a hard working man", "u r a peach"). The noun phrases are examined in order to recover the main noun term. Then its corresponding semantic

[3] http://www.answers.com/topic/food-as-metaphor
[4] http://jsgfood.blogspot.com/2008/02/food-metaphors.html

tag is derived from the composed semantic dictionary if it is a food term, or an animal-name etc. E.g. "u r a peach" has been regarded as "second-person + present-tense copular form + [food-term]". WordNet [16] has been employed in order to get the synset of the food term. If among the synset, the food term has been explained as a certain type of human being, such as 'beauty', 'sweetheart' etc. Then another small slang-semantic dictionary collected in our previous study containing terms for special person types (such as 'freak', 'angle') and their corresponding evaluation values (negative or positive) has been adopted in order to obtain the evaluation values of such synonyms of the food term. If the synonyms are positive (e.g. 'beauty'), then we conclude that the input is an affectionate expression with a food metaphor (e.g. "u r a peach"). The processing procedures are listed in the following for the input "u r a peach":

1. Rasp recognizes the input has a structure of 'second-person (you) + present-tense copular form (are) + noun phrases (a peach)';
2. The newly composed semantic dictionary is used to obtain the semantic tag for the main noun term ('peach') in the noun phrase, 'peach' -> 'food term';
3. Thus the input is regarded as 'second-person (you) + present-tense copular form (are) + [food-term: peach]'. The system interprets that the input is a food metaphor;
4. The food term, 'peach', is sent to WordNet to get its synonyms. The synonyms include special person types, such as 'beauty' and 'sweetheart';
5. The special slang-semantic dictionary containing special person types and their corresponding evaluation values is used to obtain the evaluation values (positive) of special person types 'beauty' and 'sweetheart';
6. Since the evaluation values are positive, the system concludes the input expresses affectionate with a food metaphor.

However, in most of the cases, WordNet doesn't provide any description of types of human beings when explaining a food term (e.g. 'pizza', 'meat' etc). According to the nature of the scenarios (e.g. bullying) we used, we simply conclude that the input implies insulting with a food metaphor when calling someone food terms ("u r walking meat", "Lisa is a pizza").

Another interesting phenomenon drawing our attention is food as shape metaphor. As mentioned earlier, food is often used as a metaphor to refer to body shapes (e.g. "you have a pear body shape", "Lisa has a garlic nose", "Lisa has a pizza face"). They might indicate literal truth, but most of which are potentially used to indicate very unpleasant truth. Thus they could be regarded as insulting. We extend our semantic dictionary created with the assistance of Wmatrix [15] by adding terms of physiological human body parts, such as face, nose, body etc. For the user's input with a structure of 'second-person/a singular proper noun + have/has + noun phrases' informed by Rasp, the system provides a semantic tag for each word in the object noun phrase. If the semantic tag sequence of the noun phrase indicates that it consists of a food term followed by a physiological term ('pizza face'), the system interprets that the input implies insulting with a food metaphor. Our processing could also identify input such as "she is a cake lover/eater" as literal expressions since the main objective terms are not food terms but persons/consumers. In our study, we have also made attempts to classify several metaphorical phenomena (including food metaphor) from literal expressions no matter if such metaphorical expressions conveyed any affect or not.

However, examples, such as "you have a banana body shape" and "you are a meat and potatoes man", haven't been used to express insults, but instead the former used to indicate a slim body and the latter to indicate a hearty appetite and robust character. Other examples such as "you are what you eat" could be very challenging theoretically and practically. They also indicate the direction of the future extension of our current system.

Affect detection from cooking metaphor

Another type of metaphor that has great potential to carry affect implication is cooking metaphor. Very often, the agent himself/herself would become the victim of slow or intensive cooking (e.g. grilled, cooked, etc). Or one agent can perform cooking like actions towards another agent to realize (at least most of the time) punishment or torture. Examples are as follows, "the lawyer grilled the witness on the stand", "he knew he was cooked when he saw his boss standing at the door", "he basted her with flattery to get the job", "She knew she was fried when the teacher handed back her paper" etc.

In these examples, the suffering agents have been regarded 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 examples also indicate the submissive of the suffering agents. E.g. in the examples, passive sentences ("he knew he was going to be grilled when he got home") have been very often used to imply unwillingness and victimised 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 to recognize user input with such syntactic structures ('A + copular form + VVN', 'A + VV0/VVD/VVZ + B'). Many sentences could possess such syntactic information (e.g. "Lisa was bullied", "he grills Lisa", "I was hit by a car", "Lisa was given the task to play the victim role", "he knew he was cooked", "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' labelled as 'past participle form of lexical verbs (VVN)', 'likes' and 'grills' labelled as '-s form of lexical verbs (VVZ)') and the semantic interpretation for their base forms is discovered from WordNet [16]. 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 'CREATE', or 'CHEAT' etc. One verb may recover several hypernyms and in our application, we collect all of them. Another evaluation profile [17] 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 a singular proper noun ('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 hypernym (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. Two example implementations are provided in the following.

E.g. for the user input "I was fried by the head teacher", the step-by-step 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 recognised as a cooking metaphor;

4. The three hypernyms are sent to the evaluation profile to obtain their evaluation values, thus 'KILL' is labelled as negative while others can't be returned with 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 from common-sense knowledge, the input indicates the user who is speaking suffered from a negative action and may lead to a negative emotional state (e.g. 'sad' or 'angry').

Another example input is "she grilled him in front of the class". Since the hypernyms of 'grill' only contain 'COOK' (recognized as a cooking metaphor), we cannot trace it from the evaluation profile to return an evaluation value. However, since the input has the following syntactic and semantic structure: "PPHS1 (she) + COOK + PPHO1 (him)', it implies that a third-person subject performed a cooking action towards another human object. The human object ('him') may experience 'sad' (or 'angry') emotional state.

Sometimes, our processing may not be able to recognize the metaphorical phenomena (the cooking metaphor). However it is able to interpret the affect conveyed in the user input. E.g. if we have input "he basted her with flattery to get the job", the recognition process is as follows:

1. Rasp recognises the input has the main syntactic structure of "PPHS1 (he) + VVD (basted) + PPHO1 (her)';

2. 'Baste' (verb base form) is sent to WordNet to obtain its hypernyms, which include 'MOISTEN', 'BEAT', and 'SEW';

3. The hypernyms do not include 'COOK', thus it is not recognised as a cooking metaphor;

4. The hypernyms are sent to the evaluation profile to obtain their evaluation values. We obtain negative evaluation value for the hypernym 'BEAT' from the profile.

5. The input has the following syntactic and semantic structure: 'PPHS1 (he) + VVD (BEAT: negative) + PPHO1 (her)'. It implies that a third-person subject performed a negative action towards another human object.

6. The human object ('her') may experience 'sad' or 'angry' emotional state.

Although the processing sometimes is not ideal, our method still manages to retrieve the affect embedded in the user input. Our processing is indeed at a very initial stage to sense affect from cooking metaphor and only can successfully detect approximate 30% of the cooking metaphorical phenomena shown in the online metaphor dictionary mentioned earlier. However it has pointed out promising directions for affect detection from figurative language. After our efforts to improve the performance of affect sensing from individual turn-taking user input, we focus on affect adjustification using context profiles. In the future work, we also intend to use metaphor ontology to recognise and generate metaphors.

4 AFFECT SENSING FROM CONTEXT PROFILES

Our previous affect detection has been performed solely based on individual turn-taking user input. Thus the context information has been ignored. However, such information plays a very important role in the interpretation of the affect conveyed during interaction. In this section, we are going to discuss the linguistic contextual indicators, cognitive emotion simulation from communication context and our approach developed based on the above features.

Linguistic indicators for contextual communication

We have noticed some linguistic indicators for contextual communication in the recorded transcripts. One useful indicator is (i) imperatives (such as "shut up", "go on then", and "bring it on"), which are normally used to imply negative or positive responses to the previous speaking characters. Other useful contextual indicators are (ii) prepositional phrases (e.g. "by who?", "in the green room?"), 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 from the previous speakers. The following is part of an example transcript from school bullying scenario, which is mainly about the bully, Mayid, 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 italic input from the bully, Mayid, is the improvisation based on context. The linguistic contextual indicators (such as 'too' and 'and') are used to indicate the intention of the big bully of strengthening his threatening towards other characters (especially to Lisa, the bullied victim). In this example, Dave was played by EMMA.

Mayid: im am the bully ha ha ha

(AI) Dave: Hey guys! Lisa, r u all right?

Lisa: i am ok

Mayid: i need Elise to hit her ha ha
Mayid: Lisa is mine
Elise: lisa, why r u upset??
Elise: u look ok yesterday
(AI) Dave: Hey m8, r u all right? u look terrible.
Mayid: i run Lisa over and kill her ha ha
Lisa: i am ok
Mayid: *my mum is a bully too*
Elise: i love lisa
(AI) Dave: That sounds great
Mayid: *and my dad*

In addition, (iii) short phrases for questions are also used frequently in the recorded transcripts to convey contextual communication, e.g. "where?", "who?", "who is Dave" or "what". However, except for the potential implication of 'confusion', there is rarely any other affective state expressed in such short question style responses.

(iv) The character names are also normally used in the user input to indicate that the current input is intended for a particular character, 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 characters and the named characters. 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 (i) some other well known contextual indicators such as 'yeah/yes followed by a sentence ("yeah, yeah, we are good", "yeah, we will see", "yeah, we cool Lisa")', "I think so", 'no/nah followed by a sentence', "me too", "exactly", "thanks", "sorry", "lol", "grrrr", "hahahaha", "ohhhh" etc. Such expressions are normally used to indicate affective responses to the previous user input.

Since natural language is ambiguous and there are cases that contextual information is required in order to appropriately interpret the affect conveyed in the input (e.g. "go on then", "and my dad"), 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.

Emotion modelling 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 conveyed 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 or funeral. 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, 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 his behaviour" 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. Then a dynamic algorithm is used to find the most resembling emotional context for any given new situation from Markov chain's training and generated emotional contexts. The most resembling emotional context is used to adjust/predict the affect conveyed in the current user input. The recommended affect will also be further justified by the most recent affect conveyed by other characters (e.g. a most recent 'insulting/rude' input from Mayid may cause Lisa and her friends to be 'angry').

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. In our previous work, we detected 25 affective states including basic (e.g. 'happiness') and complex emotions (e.g. 'embarrassment' and other affective states described by OCC model), meta-emotions, value judgments etc. We have employed 12 most commonly used emotions (caring, arguing, disapproving, approving, grateful, happy, sad, threatening, embarrassed, angry/rude, scared and sympathetic) out of the 25 affective states in our present work on contextual emotion analysis and prediction. A matrix has been constructed dynamically for the 12 most commonly used emotions in our application 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.

From the inspection of the previous testing transcripts, we noticed that the human director normally 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. By using the dynamic 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 possible emotion for the current user input. The detailed explanation of the approach taken and example runs are in the following. Since the most recent affect histories of other characters may also have impact on the affect conveyed by the current speaking character, the recommended affect by the dynamic algorithm will also be further evaluated by these affect histories.

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 (including "Mayid starts bullying Lisa", "why Lisa is crying", "why Mayid is so nasty/a bully", "how Mayid feels when his uncle finds out his behaviour" etc). The classification of the sub-theme is mainly based on the human director's intervention which was recorded in the transcripts. Then we have 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 emotional contexts as many as possible (at least 4 emotional contexts) 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 user 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 [rude]: lisa, y u crying? u big baby!

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

E.g. the emotional context for Mayid from the above example is: 'rude' and 'rude' (we use one letter to represent each emotional label, e.g. the emotional context here is represented as 'R R'), 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 summarised the emotional contexts 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 we have taken into consideration at the training stage are the following: "Mayid starts bullying.", "why is Lisa crying?", "why is Mayid nasty/a bully?" and "how does mayid feel if his uncle knew about his behaviour?"

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 as usual. 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 communication contextual input. E.g. we mainly focus on the checking of the five contextual implications we mentioned previously, including imperatives, prepositional phrases and conjunctions, simplified question sentences, character names, and other commonly used contextual indicators (such as 'yea/yes/yeah' or 'yea/yes/yeah followed by a sentence'). If any of the above contextual indicators exists, then we further analyse the affect embedded in the input with contextual emotion modelling reported here.

E.g. we have collected the following transcript for testing.

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: miss[sssssssssssssss] [neu]
8. Elise Brown: lol [happy]
9. Lisa Murdoch: mayid u gna gt banned [threatening]
10. Mayid Rahim: LOL [happy]
11. Mayid Rahim: BY HU [neu] -> [angry]
12. Elise Brown: mayid got a tongue yo! [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. Therefore, it implies that it could be a situation raised by the previous context (e.g. previous user input from Lisa) and the further processing for emotion prediction is activated. Since we also 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 dynamic programming algorithms at this stage 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 chatting and affect history, there is only one user input from Lisa after the director's intervention, which implied 'anger'. Since Lisa and Mayid have a negative

relationship, we predict that Mayid currently experiences negative emotion. Since capitalisations have been used in Mayid's input, then 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 reasoning of the emotion conveyed in the current user input (e.g. if Lisa implied 'neutral' in the most recent input, then the interpretation of the affect conveyed by Mayid would still be '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, the input may imply a potential (emotional) response to the previous speaking character. Sometimes affect expressed in the imperatives are very explicit, e.g. "shut up/it (rude/angry)". However, there are imperatives which convey affect very implicitly, e.g. the current input, "go on then". Further processing is required to uncover the affect embedded in it. Thus the emotional context profile for the Mayid character 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 the Mayid character.

The following are examples of the generated 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. A A A ['angry, angry, and angry']
3. N A N A N A ['neutral, angry, neutral, angry, neutral, and angry']
4. A A N ['angry, angry and neutral']
5. D A A A A A A N A ['disapproval, angry, angry, angry, angry, angry, neutral, and angry']

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 = 1) 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 the affect detection component 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 (especially the leading opponent character), 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 11th input). However our processing could be fooled easily by various diverse ways for affective expressions and creative improvisation (emotional patterns not shown in the training and generated sets). We intend to adopt more linguistic contextual hints, psychological (context-based) emotional theories and better simulation tools (e.g. hidden Markov modelling) for further improvements. Our processing currently only focused on school bullying scenario. We are on our way to extend the context-based affect sensing to other scenarios to further evaluate its efficiency.

5 EVALUATION

We carried out user testing with 220 secondary school students from Birmingham schools and Education Village in Darlington 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 had an experience of both scenarios, one including the AI minor character only and the other including the human-controlled minor character only. Such arrangement could not only enable us to measure any statistically significant difference to users' engagement and enjoyment due to the involvement of the AI minor character, but also provide us the opportunity to compare the performance of the AI minor character with that of the human-controlled one. 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 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 the 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 the transcripts used for the training of Markov chain). 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 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 for some of the characters and 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. 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. Wallstreet Journal and other metaphorical databases (e.g. ATT-Meta)) to further evaluate the new development on metaphorical affect sensing.

6 CONCLUSIONS & FUTURE WORK

We have made new developments on automatic affect sensing in metaphorical figurative language and employed context profiles for affect interpretation. Although EMMA could be challenged by the rich diverse variations of the metaphorical language phenomena we focused on and other improvisational complex context situations, we believe these areas are very crucial for development of effective intelligent user interfaces and our processing has made promising initial steps towards these areas. Also, further testing is needed to further evaluate the efficiency of metaphorical and context-based affect detection. We also intend to compare our system's affect sensing performance with that of a well-known affect sensing tool, ConceptNet [5], as another way for future evaluation.

REFERENCES

[1] L. Zhang, J. A. Barnden, R. J. Hendley, M. G. Lee, A. M. Wallington and Z. Wen. Affect Detection and Metaphor in E-drama. *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 18, No. 2, 234-252. (2008).

[2] L. Fainsilber and A. Ortony. Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity*, 2(4), 239-250. (1987).

[3] ATT-Meta Project Databank: Examples of Usage of Metaphors of Mind.http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/. (2008).

[4] Z. Kövecses. 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. (1998).

[5] H. Liu & P. Singh. ConceptNet: A practical commonsense reasoning toolkit. *BT Technology Journal*, Volume 22, Kluwer Academic Publishers. (2004).

[6] M. A. M. Shaikh, H. Prendinger & I. Mitsuru. Assessing sentiment of text by semantic dependency and contextual valence analysis. In *Proceeding of ACII 2007*, 191-202. (2007).

[7] M. Mateas. Ph.D. Thesis. Interactive Drama, Art and Artificial Intelligence. School of Computer Science, Carnegie Mellon University. (2002).

[8] X. Zhe & A. C. Boucouvalas. 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. (2002).

[9] A. C. Boucouvalas. Real Time Text-to-Emotion Engine for Expressive Internet Communications. In *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments*. G. Riva, F. Davide and W. IJsselsteijn (eds.), 305-318. (2002).

[10] R. Craggs & M. Wood. A Two Dimensional Annotation Scheme for Emotion in Dialogue. In *Proceedings of AAAI Spring Symposium: Exploring Attitude and Affect in Text.* (2004).

[11] A. Egges, S. Kshirsagar & N. Magnenat-Thalmann. 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. (2003).

[12] C. Elliott, J. Rickel & J. Lester. Integrating Affective Computing into Animated Tutoring Agents. In *Proceedings of IJCAI'97 Workshop on Intelligent Interface Agents*, 113-121. (1997).

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

[14] M. Cavazza, C. Smith, D. Charlton, L. Zhang, M. Turunen and J. Hakulinen. 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. (2008).

[15] P. Rayson. Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. Ph.D. thesis, Lancaster University. (2003).

[16] C. Fellbaum. WordNet, an Electronic Lexical Database. The MIT press. (1998).

[17] A. Esuli and F. Sebastiani. Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of EACL-06*, Trento, IT. 193-200. (2006).