Affect Detection from Semantic Interpretation of Drama Improvisation

Li ZHANG[1] Ming JIANG[2]

(1) NORTHUMBRIA UNIVERSITY, Faculty of Engineering and Environment, Newcastle, UK
(2) UNIVERSITY OF LEEDS, School of Computer Science, UK
`li.zhang@northumbria.ac.uk`

ABSTRACT

We have developed an intelligent agent to engage with users in virtual drama improvisation previously. The intelligent agent was able to perform sentence-level affect detection from user inputs with strong emotional indicators. However, we noticed that many inputs with weak or no affect indicators also contain emotional implication but were regarded as neutral expressions by the previous interpretation. In this paper, we employ latent semantic analysis to go beyond linguistic restrictions and to perform topic theme detection and identify target audiences for those inputs with vague affect indicators and ambiguous target audiences. We also discuss how emotions embedded in such emotionally ambiguous inputs are detected with the consideration of interpersonal relationships, special sentence types and emotions experienced by the target audiences using a neural network based contextual affect detection. The work contributes to the conference themes on discourse and pragmatics, semantics and sentiment and text classification.

TITLE AND ABSTRACT IN CHINESE

对于戏剧即席创作语义解析的情感识别

我们曾开发了一个能跟用户进行虚拟戏剧即席创作交流的智能代理. 它能从有明显情感迹象的单句话中检测情感. 然而，很多不具有情感词的句子也有很强的情感寓意却被认为是中性. 在这里，我们使用 Latent Semantic Analysis, 摆脱语言特征的限制，用话题和目标听众的检测去识别情感隐含在那些只具有微弱情感信号的输入中. 并讨论怎样从这样的输入中，使用基于神经网络的上下文检测去识别情感.

KEYWORDS : Affect detection, semantic interpretation, drama improvisation
KEYWORDS IN CHINESE : 情感检测， 语义解析， 戏剧即席创作

1   Introduction

It is a long-term research goal to build a 'thinking' machine in the AI field. This endeavour has given rise to agent-based user interfaces (Endrass et al., 2011; Zhang et al., 2009). Moreover, we believe it will make intelligent agents possess human-like behaviour and narrow the communicative gap between machines and human-beings if they are equipped to interpret human emotions during social interaction. Thus in this research, we equip our AI agent with emotion and social intelligence. According to Kappas (2010), human emotions are psychological constructs with notoriously noisy, murky, and fuzzy boundaries. These natural features of emotion also make it difficult for a single modal recognition, such as via acoustic-prosodic features of speech or facial expressions. Since human being's reasoning process takes related context into consideration, in our research, we intend to make our agent take multi-channels of subtle emotional expressions embedded in social interaction contexts into consideration to draw reliable affect interpretation. The research presented here focuses on the production of intelligent agents with the abilities of interpreting dialogue contexts semantically to support affect detection as our first step of building an agent-based interface within this application domain.

The research presented here is conducted within a previously developed online multi-user role-play virtual drama framework, which allows school children aged 14 – 16 to perform drama performance training. In this platform young people could interact online in a 3D virtual drama stage with others under the guidance of a human director. In one session, up to five virtual characters are controlled on a virtual stage by human users ("actors"). The actors are given a loose scenario around which to improvise, but are at liberty to be creative. An intelligent agent is also involved in improvisation. It included an affect detection component, which detected affect from human characters' each individual text-based turn-taking input. This previous affect detection component was able to detect 15 emotions including basic and complex emotions, but the detection has not taken any context into consideration. The agent also made attempts to produce appropriate responses to help stimulate the improvisation based on the detected affect. The detected emotions are also used to generate emotional animations of the avatars.

This original affect detection processing was mainly built using pattern-matching rules that looked for simple grammatical patterns or templates. A syntactic parser, Rasp (Briscoe and Carroll, 2002), was also used to provide syntactical processing of each input. From the analysis of the collected transcripts, the original affect interpretation without any contextual inference proved to be effective enough for those inputs containing strong clear emotional indictors such as 'yes/no', 'haha', 'thanks' etc. There are also situations that users' inputs contain very weak or even no affect signals, thus contextual inference is needed to further derive the affect conveyed in such inputs. Moreover, it is noticed that in the collected transcripts the improvisational dialogues are often multi-threaded. This refers to the situation that conversational responses of different discussion themes to previous several speakers are mixed up due to the nature of the online chat setting. Therefore the detection of the most related discussion themes using semantic analysis is very crucial for the accurate interpretation of emotions implied in those inputs with ambiguous audiences and weak affect indicators.

2   Related work

There is much well-known research work in the field of intelligent conversational agents. Aylett et al. (2006) focused on the development of affective behaviour planning for their synthetic

characters. Endrass, Rehm and André (2011) carried out study on the culture-related differences in the domain of small talk behaviour. Their agents were equipped to generate culture specific dialogues. Recently textual affect sensing has also drawn researchers' attention. Neviarouskaya et al. (2010) provided a sentence-level rule-based textual affect sensing system to recognize judgments, appreciation and affective states. But the detection was still limited to the analysis of individual inputs. Ptaszynski et al. (2009) employed context-sensitive affect detection with the integration of a web-mining technique to detect affect from users' input and verify its contextual appropriateness. However, their system targeted interaction only between an agent and one human user, which reduced the complexity of the modelling of the interaction context.

There is also research related to building opinion-related lexical resources beneficial to opinion mining applications. E.g. Esuli (2008) employed a semi-supervised term classification model with quantitative analysis of definitions of terms provided by on-line dictionaries. The research generated a lexical resource, SentiWordNet. It provided positive, negative and objective orientations for a general category of terms and senses. Cambria and Hussain (2012) proposed a sentic computing framework for open-domain opinion mining and sentiment analysis based on the integration of common sense knowledge and graph mining and multi-dimensionality reduction techniques. Generally, they employed common sense computing techniques to bridge the semantic gap between word-level data and their corresponding concept-level opinions. Moreover, as mentioned earlier, naturalistic emotion expressions usually consist of a complex and continuously changed symphony of multimodal expressions. Kappas (2010) argued that it is inappropriate to conclude a smiling user is really happy. In fact, the same expression can be interpreted completely differently depending on the context that is given. Thus it also motivates us to use semantic interpretation of social contexts to inform affect detection in this research.

3 Semantic interpretation of social interaction contexts

In the collected transcripts, we noticed that the language used in our application domain is often complex, idiosyncratic and invariably ungrammatical. Most importantly, the language also 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. Therefore, in this work, we integrate contextual information to further derive the affect embedded in contexts and to provide affect interpretation for those without strong affect indicators.

In our original affect detection processing, we relied on keywords and partial phrases matching with simple semantic analysis using WordNet. However, we notice many concepts and emotional expressions can be described in various ways. Especially if the inputs contain no strong affect indicators, other approaches focusing on underlying semantic structures should be considered. Thus in this section we discuss the approaches of using latent semantic analysis (LSA) (Landauer and Dumais, 2008) and its related packages for terms and documents comparison to recover the most related discussion themes and target audiences to benefit affect detection.

LSA generally identifies relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. In order to compare the meanings behind the words, LSA maps both words and documents into a 'concept' space and performs comparison in this space. In detail, LSA assumes that there are some underlying latent semantic

structures in the data which are partially obscured by the randomness of the word choice. This random choice of words also introduces noise into the word-concept relationship. LSA aims to find the smallest set of concepts that spans all the documents. It employs singular value decomposition to estimate the hidden concept space and to remove the noise. This concept space associates syntactically different but semantically similar terms and documents. We use these transformed terms and documents in the concept space for retrieval rather than the original ones.

In our work, we employ the semantic vectors package (Widdows and Cohen, 2010) to perform LSA and analyze underlying relationships between documents and their similarities. This package provides APIs for concept space creation. It applies concept mapping algorithms to term-document matrices using Apache Lucene, a high-performance, full-featured text search engine library implemented in Java. We integrate this package with the AI agent's affect detection component to calculate semantic similarities between those inputs without strong affect signals and training documents with clear discussion themes. In this paper, we target the transcripts of the school bullying scenario[1] for context-based affect analysis.

In order to perform semantic comparison between user inputs and documents belonging to different topic categories, sample documents with strong topic themes are collected. Personal articles from the Experience project (www.experienceproject.com) are used for this purpose. These articles belong to 12 categories including Education, Family & Friends, Health & Wellness, etc. Since we intend to perform discussion theme detection for the transcripts of those employed testing scenarios (including school bullying and Crohn's disease), we extracted documents close enough to these scenarios including articles of Crohn's disease (five articles), school bullying (five), family care for children (five), food choice (three), school life including school uniform (10) and school lunch (10) etc. Phrase and sentence level expressions implying 'disagreement' and 'suggestion' were also gathered from several other articles published on the website. Thus we have training documents with eight discussion themes including 'Crohn's disease', 'bullying', 'family care', 'food choice', 'school lunch', 'school uniform', 'suggestions' and 'disagreement'. The first six themes are sensitive and crucial discussion topics to the employed scenarios, while the last two themes are intended to capture arguments expressed in multiple ways. Affect detection from metaphorical expressions often poses great challenges to automatic linguistic processing systems. In order to detect a few metaphorical phenomena, we include four types of metaphorical examples published on the following website: http://knowgramming.com, in our training corpus. These include cooking, family, weather, and farm metaphors. We also borrowed a group of 'Ideas as External Entities' metaphor examples from the ATT-Meta databank (http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/) to enrich the metaphor categories. Individual files are used to store each type of the metaphorical expressions. All the sample documents of the above 13 categories are regarded as training files.

We also added some training documents with broader topic themes as noise training data in order to evaluate the robustness of topic theme detection. Five articles of each of the following themes are employed: 'alcoholism', 'voluntary work', 'self-employment', 'politics', and 'hobbies'. These are also added to the training corpus for topic theme detection. The following example interaction of the school bullying scenario is used to demonstrate how we detect the discussion themes for those inputs with weak affect indicators and ambiguous target audiences.

---

[1] 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.

1. Mrs Parton: children, stop arguing! [disapproval]
2. Mayid: u shut up, how the hell does that sound like a gal, u twat!!! [angry]
3. Elise: Stop it, Mayid. Lisa how r u? [disapproval]
4. Mayid: do ya even have any brain to think about that one! [Topic: bullying and disease, Target audience: Elise, Emotion: angry]
5. Lisa: hi, elise, *I'm alright*. [neutral]
6. Elise: cuz it jus does. Actually I'm cleverer than u think, u wus. [angry]
7. Mayid: ur da most ugly wus face! [angry]
8. Dave: could u please all tune ur voice down. [Played by the AI agent]
9. Elise: look at ur face u twat. [angry]
10. Mayid: my face is beautiful and wot, u jealous!! [angry]
11. Elise: I think the mirror breaks all da time u look in it. [Topic: bullying, Target audience: Mayid, Emotion: angry]
12. Mayid: hahaha. [happy]
13. Dave: Are these all desperate people? [Played by the AI agent]
14. Mayid: u looking in da mirror rite now, but u probably can't see urself with all the cracks. [Topic: bullying, family care and suggestion, Target audience: Elise, Emotion: angry]

The original affect detection focuses on inputs with strong emotion signals and provides affect annotation for such inputs in the above example. The emotion indicators are also illustrated in italics in the above interaction. The inputs without an affect label followed straightaway are those with weak affect indicators (4th, 11th and 14th inputs). Therefore further processing is needed to recover their discussion themes and identify their most likely audiences in order to identify implied emotions more accurately. The general idea for the detection of discussion themes is to use LSA to calculate semantic distances between each test input and all the training files with clear topic themes. Semantic distances between the test input and the 13 valid topic terms (e.g. 'disease') are also calculated. The detected topics are derived from the integration of these semantic similarity outputs. We start with the 4th input to demonstrate the theme detection.

Documents	Similarity scores for document vectors closest to the vector for the topic theme, 'bullying'
bullied1.txt	0.733
bullied2.txt	0.472
bullied3.txt	0.285
family_care4.txt	0.232
school_uniform.txt	0.231
crohn2.txt	0.230
test_corpus1.txt (the 4th input)	0.220

TABLE 1 – Partial scores for document vectors closest to the vector of the theme 'bullying'

In order to produce a concept space, the corresponding semantic vector APIs are used to create a Lucene index for all the training samples and the test file ('test_corpus1.txt' contains the 4th input). This generated index is then used to create term and document vectors, i.e. the concept space. First of all, we provide rankings for all the training files and the test input based on their semantic distances to a topic theme by searching for document vectors closest to that of a specific term (e.g. 'bullying'). The 4th input thus semantically relates to the topic theme, 'bullying', the most among all the 13 topics. Table 1 shows the partial outputs of such semantic calculation. Moreover, another effective approach for topic detection is to find the semantic similarity

between documents. If the semantic distances between training files and the test file are calculated, then it provides another source of information for topic detection. Therefore we use the CompareTerms semantic vector API to calculate semantic similarities between documents.

The similarity results show there are three training files (bullied3.txt, bullied2.txt and crohn3.txt) semantically most similar to the test file. These three files respectively recommend the following two themes: 'bullying' and 'disease'. In the processing of finding documents closest to a topic theme vector (see Table 1), the test input also achieves the best ranking for the 'bullying' theme. With the integration of the semantic similarity results between document vectors, the processing concludes that the 4th input relates most closely to topics of 'bullying' and 'disease'. In order to identify its target audiences, we start from the 3rd input to derive topic themes until retrieving the input with at least partially the same themes as those of the 4th input. The original affect processing detects the 3rd input is most likely to indicate 'bullying' with a rude attitude. It shares one of the themes embedded in the 4th input. The 3rd input from Elise also mentions Mayid as its audience. Thus the target audience of the 4th input is Elise, who started the conversation in the first place.

In a similar way, the topic detection processing also identifies the 11th input from Elise indicates a theme of 'bullying'. In order to find its target audience, the theme detection starts from the 10th input from Mayid. The original affect processing identifies the 10th input shows an 'angry' emotion indicated by a strong affect indicator, thus it contains a 'bullying' theme. Moreover, the 9th input is the last round input from the same speaker, Elise. The original affect detection also identifies it as an 'angry' aggressive input. Based on the above reasoning, Elise showed aggressive behaviour in the last round input, followed by Mayid's angry response. Therefore this new round input from Elise with a strong 'bullying' theme most likely continues the previous bullying discussion. Thus the 11th input from Elise regards Mayid as the most intended audience.

By searching for document vectors closest to those of the topics 'family care' and 'bullying', the 14th input from Mayid shows high semantic closeness to these two topics. The similarity calculation between document vectors indicates that it is also most closely related to 'bullied3.txt (0.813)' and 'suggestion1.txt (0.788)'. Thus the 14th input is most likely to indicate topics of 'bullying', 'family care' and 'suggestion'. Since the 13th input from Dave, played by the AI agent, indicates 'disapproval', it is regarded to indicate 'bullying'. Thus Dave is one of the audiences of this 14th input. Moreover, as discussed earlier, the 11th input from Elise contains a 'bullying' theme with Mayid as the audience. Thus the 14th input from Mayid is unlikely to indicate topics of 'family care' or 'suggestion', but more likely to indicate 'bullying' with Elise and Dave as the intended audiences. In general, the semantic-based theme detection is able to help the AI agent derive the most related discussion themes and identify the most intended audiences for those inputs without strong affect indicators. We believe these are very important aspects for the accurate interpretation of the emotion contexts.

4   A neural network-based contextual affect detection

The research of Wang et al. (2011) discussed that feedback of artificial listeners can be influenced by relationships, personalities and culture. The research of Hareli and Rafaeli (2008) also pointed out that "one person's emotion is a factor that can shape the behaviours, thoughts and emotions of other people". Thus in this work such interpersonal (positive (friendly) or negative (hostile)) relationships are also employed to advise affect detection in social contexts.

In the example mentioned in section 3, the topic detection identifies the most likely audience of the 4th input from Mayid is Elise. That is, the most related social context of the 4th input is the 3rd input indicating a 'bullying' negative theme contributed by Elise. Especially, the speaker, Mayid (the bully) and the audience, Elise (the bullied victim's best friend) have a tense relationship, thus the 4th input from Mayid with the themes of 'bullying' and 'disease' will be most likely to show 'sad' or 'outrageous/angry' indication. Moreover, the processing also reveals that the 11th input from Elise is mainly related to the 'bullying' topic and its target audience is Mayid. Since Mayid and Elise share a tense relationship and the bully, Mayid, has expressed an 'angry' emotion in the most related context (i.e. the 10th input), this 11th bullying input from Elise most probably indicates 'anger'. In a similar way, the 14th input from Mayid is also embedded in a negative context contributed by the 11th and 13th inputs with strong bullying themes. Thus this last input is more likely to continue the 'bullying' discussion theme rather than focusing on any other topics such as 'family care' and 'suggestion'. Therefore it most probably indicates 'anger'. Moreover, in this work, we also employ sentence types as another dimension for context-based affect detection. Especially we detect rhetorical questions using LSA. E.g., the semantic vector API is used to perform semantic similarity comparison between rhetorical & normal training document vectors and the 4th input from Mayid. The processing recognizes the 4th input as a rhetorical question with a high confidence score.

Moreover, we implement the above reasoning of emotional influences between characters using a supervised neural network algorithm, Backpropagation. The neural network we used employs a three-layer topology: one input, one hidden and one output layer, with six nodes in the input layer and 10 nodes respectively in the hidden and output layers. The six nodes in the input layer indicate the most recent emotions expressed by potential up to four target audiences, a sentence type and an averaged relationship value between the speaker and audiences. The 10 nodes in the output layer represent the 10 output detected affective states ('neutral', 'approval', 'disapproval', 'angry', 'grateful', 'regretful', 'happy', 'sad', 'worried' and 'caring'). They are chosen because of their high occurrences in the annotation of the training set. These emotion labels are mainly borrowed from Ekman (1992) and the OCC emotion model (Ortony et al., 1988). We also notice that the semantic boundaries between some of the emotions are rather fuzzy, e.g., 'regret' overlapping with 'sadness'. However, although these two emotions both belong to the appraising of events (consequences for self), 'sadness' reflects more generally on one's well-being while 'regret' is a specific kind of distress involving more specific events about which the experiencing person is displeased. In this application, 'sadness' is used for context-based general emotion appraisal while 'regret' is used only when the input contains specific strong affective indicators such as 'sorry' and 'I should*n't have done that*'. Moreover, the output emotion with the highest weighting is regarded as the most probable emotion implied in the current input.

500 example inputs with agreed annotations from the bullying scenario are used to train the neural network. After it is trained to reach a reasonable error rate (< 0.05 with an average training time: 3.5s), it is used for testing to predict emotional influence of other participant characters towards the speaking character. In the example discussed in section 3, for the 4th input, the neural net considers the following as inputs: the implied 'angry' emotion by the audience, Elise, 'a negative relationship' and a rhetorical question input. The algorithm detects 'anger' implied in the 4th input. Similarly, it interprets both the 11th and 14th inputs indicating 'angry' emotions.

In order to improve the system's robustness, we use semantic orientations of words/phrases embedded in sentences and min-margin based active learning to detect emotions from open-

ended inputs without the constraints of pre-defined scenarios. Especially it helps to interpret emotions when daily-life discussion outside of the scenarios is less heated with diverse number of audiences, or emotion contexts of audiences or relationships between characters are not available.

5   Evaluation and conclusion

User testing was conducted previously with 200 British secondary school students to evaluate the affect detection and the AI agent's performance. We use previously collected transcripts to evaluate the efficiency of the updated affect detection with contextual inference. In order to evaluate the performances of the topic theme detection and the neural network based affect detection, three transcripts of another scenario, Crohn's disease, are used. Two human judges are employed to annotate the topic themes of the extracted 300 inputs from the test transcripts using the 13 topics. We used Cohen's Kappa to measure the agreement level between human judges for the topic annotation and obtained 0.813. Then the 250 inputs with agreed annotations are used as the gold standards to test the performance of the theme detection. A pattern matching baseline system is used to compare the performance with that of the LSA. We obtain an averaged precision, 0.783, and an averaged recall, 0.753, using the LSA while the baseline system achieves an averaged precision of 0.609 and an averaged recall of 0.587 for the 13 topic theme detection. Generally the semantic-based interpretation achieves better performances than the baseline system.

The human judges also annotated these 250 inputs with the output 10 emotions. The inter-annotator agreement between human judge A/B is 0.65. While the previous version of the affect detection achieves 0.43 in good cases, the new version achieves agreement levels with human judge A/B respectively 0.55 and 0.58. The new version achieves inter-annotator agreements generally fairly close to the agreement level between human annotators themselves.

Moreover, in order to provide evaluation results for the neural network-based affect detection, the human judges' previous annotations are converted into positive, negative and neutral. Then 203 inputs with agreed annotations are used as the gold standards. The annotations achieved by the neural net are also converted into solely positive and negative. A baseline system is built using simple Bayesian networks in order to further measure the neural network-based detection. The Bayesian network used emotions implied in the last two inputs as its inputs. The output is the predicted affect implied in the current input. The neural network inference with the consideration of relationships, sentence types and audiences' emotions achieved an average precision of 0.833 and an average recall of 0.827 while the baseline system achieved a precision of 0.609 and a recall of 0.633. Especially our approach coped well with the sudden change of emotions due to unexpected topic change, while such situations challenged the baseline system greatly.

We also noticed that the training and test transcripts contained imbalanced class categories, e.g. more negative inputs presented than positive and neutral ones. In order to deal with such imbalanced classifications, we employ min-margin based active learning. It proved to be efficient in dealing with open-ended and imbalanced affect classifications in our application. In future work, we aim to equip the AI agent with culturally related small talk behaviour in order to ease the interaction. The presented semantic analysis also shows great potential to automatically recognize emotional metaphorical expressions and contribute to the responding regimes for the AI agent's development. Other uncertainty sampling techniques will also be employed. We believe these are crucial aspects for the development of effective agent-based interfaces.

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.

Briscoe, E. and 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.

Cambria, E. and Hussain, A. (2012). Sentic Computing: Techniques, Tools, and Applications. Springer, 2012 Edition. ISBN-10: 9400750692.

Ekman, P. (1992). An Argument for Basic Emotions. In Cognition and Emotion, 6, 169-200.

Endrass, B., Rehm, M. & André, E. (2011). Planning Small Talk Behavior with Cultural Influences for Multiagent Systems. Computer Speech and Language. 25(2):158-174.

Esuli, A. (2008). Automatic Generation of Lexical Resources for Opinion Mining: Models, Algorithms and Applications. PhD thesis. PhD School "Leonardo da Vinci", University of Pisa.

Hareli, S. and Rafaeli, A. (2008). Emotion cycles: On the social influence of emotion in organizations. Research in Organizational Behavior, 28, 35-59.

Kappas, A. (2010). Smile when you read this, whether you like it or not: Conceptual challenges to affect detection. IEEE Transactions on Affective Computing, 1 (1), 38-41.

Landauer, T.K. and Dumais, S. (2008). Latent semantic analysis. Scholarpedia, 3(11):4356.

Neviarouskaya, A., Prendinger, H. and Ishizuka, M. (2010). Recognition of Affect, Judgment, and Appreciation in Text. In Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, pp. 806-814.

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 Proceeding of IJCAI.

Wang, Z., Lee, J. and Marsella, S. (2011). Towards More Comprehensive Listening Behavior: Beyond the Bobble Head. In Proceedings of International Conference on Intelligent Virtual Agents.

Widdows, D. and Cohen, T. (2010). The Semantic Vectors Package: New Algorithms and Public Tools for Distributional Semantics. In Proceedings of IEEE International Conference on Semantic Computing.

Zhang, L., Gillies, M., Dhaliwal, K., Gower, A., Robertson, D. and 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.