

A Revisit to the Incorporation of Context - awareness in Affective Computing Systems

Aggeliki Vlachostergiou and Stefanos Kollias

Image Video and Multimedia Systems Laboratory
School of Electrical and Computer Engineering
National Technical University of Athens
Iroon Polytexneiou 9, 15780 Zografou, Greece
aggelikivl@image.ntua.gr, stefanos@cs.ntua.gr

Abstract. The research field of Human Computer Interaction (HCI) is moving towards more natural, sensitive and intelligent paradigms. Especially in domains such as Affective Computing (AC), incorporating interaction context has been identified as one of the most important requirements towards this concept. Nevertheless, research on modeling and utilizing context in affect-aware systems is quite scarce. In this paper, we revisit the definition of contexts in AC systems and propose a context incorporation framework based on semantic concept detection, extraction enrichment and representation in cognitive estimation, to further clarify and assist interpretation of contextual effects in Affective Computing systems.

Keywords: HCI, Affective Computing, Context, Context modeling, Context-aware systems, SEMAINE

1 Introduction

Affective Computing (AC) systems have been well developed in the past decades as an effective solution for recognizing, interpreting, processing, and simulating human affects. Additionally, context-aware Affective Computing systems [2] emerged as a novel way of shifting from context-aware Affective Computing systems to affective aware intelligent Human Computer Interaction systems to overcome the fact that contextual information can not be discounted in doing automatic analysis of human affective behavior. Thus, the fundamental assumption of context-aware Affective Computing systems is that context is able of shaping how people interpret the high and complex expressions of people and machines. The variation of these expressions in human behavior arise not only from a subject's internal psychological or cognitive state but also from other subjects or the environment. For instance, frowning behavior may be an indicator of anger or it may be due to concentration depending on the contextual interactional setting. Context-based affect aware analysis needs to clarify the preliminary selection of contextual variables in order to further assist interpretation of contextual effects in Affective Computing systems. As a result, how to incorporate contexts into

Affective Computing systems is always a research question in this domain. However, prior to that, which variables should be considered as contexts is still under question.

Currently, several context-aware Affective Computing systems [5] have been developed but very few research went back to discuss the definition of contexts. And researchers simply blend location, identities of people around the user, environment, social interaction etc. and other variables together to consider them as contexts, which further creates the confusion on which should be the first decisions to be made prior to creating context-aware automatic affect analysis systems. The definition and exploration of context are not only related to the selection of contextual variables in Affective Computing systems, but are also relevant to the interpretation of contextual effects based on the outcomes or findings in the experiments. It is obvious that the academic area focuses more on the development of effective context-aware Affective Computing systems, but ignores the identification of contexts and interpretation of contextual effects in recent years.

This paper is organized as follows: Section 2 provides an overview of how the term context has been studied so far in various disciplines. Section 3 presents an overview of my research and the preliminary findings in my previous work. Section 4 discusses my plan for achieving my overall objective for the remainder of my doctoral work and finally, Section 5 concludes by summarizing the impact and relevance of the proposed approaches to the field of context-aware Affective Computing.

2 Related Work

Actually contexts have been studied in various disciplines, such as ubiquitous computing, contextual advertising, social signal processing, HCI, gaming, mental health etc., where the definition differs resulting in a different understanding of contexts among those areas. In context-aware Affective Computing systems, the earliest research papers [5] may bring us to look back upon almost ten years ago; however, the field has yet to agree on the definition of context. Several researchers simply blend verbal content (semantics), knowledge of the general interactional setting, discourse and social situations and other variables together and consider all of them as contexts, which further creates confusion on which should be the most appropriate contextual parameter w.r.t context-aware Affective Computing systems (Figure 1).

The most commonly used definition is the one given by Abowd et al. in 1999 [1], “context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.” This definition hardly limits the set of variables that can be taken into account, and it is still ambiguous without clear guidelines to select appropriate variables in AC systems.

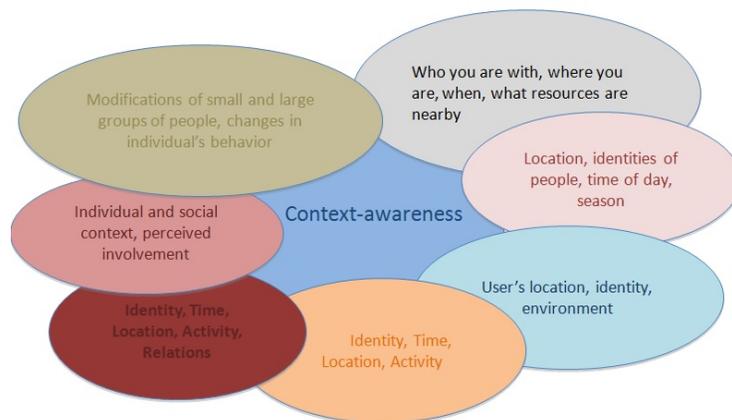


Fig. 1: Number of variables that have been considered so far as contextual parameters in the area of Affective Computing.

Apparently, the definition and selection of contexts is a domain-specific problem, where classification of contextual variables is a typical way to put different variables in categories but it is still not general enough and not flexible in interpreting the contextual effects. The debate or discussion may be finally ended by the idea proposed by Duric et al. [3], which is known as W5+ formalization and incorporates context corresponding to the answering of the following important aspects of context: Who you are with (e.g. dyadic/multiparty interactions), What is communicated (e.g., (non)-linguistic message), How the information is communicated (the persons affective cues), Why, i.e., in which context the information is passed on, Where the user is, What his current task is, How he/she feels (has his mood been polarized changing from negative to positive?) and which (re)action should be taken to satisfy humans needs).

Unfortunately, so far the efforts on human affective behavior understanding are usually context independent [6]. In light of these observations, understanding the process of a natural progression of context-related questions when people interact in a social environment could provide new insights into the mechanisms of their interaction context and affectivity. The Who, What, Where context-related questions have been mainly answered either separately or in groups of two or three using the information extracted from multimodal input streams [6]. Thus, as of date, no general W5+ formalization exist, due to the fact that current systems which answer to most of the W questions are founded on different psychological theories of emotion. Recent research on progressing to the questions of “Why” and “How” has led to the emerging field of sentiment analysis, through mining opinions and sentiments from natural language, which involves a deep understanding of semantic rules proper of a language.

3 Preliminary Work

The key objective of my PhD research is to computationally identify, automatically extract and incorporate contextual information into affect aware recognition frameworks, with the aim of identifying context-aware emotional-specific patterns.

3.1 Context identification Framework

To fulfill the need of understanding whether and how context is incorporated in automatic analysis of human affective behavior, we propose a novel context-aware incorporation framework (Fig. 2) which (i): includes detection and extraction of semantic context concepts, (ii): better enriches a number of Psychological Foundations with sentiment values and (iii): enhances emotional models with context information and context concept representation in appraisal estimation [11].

As a first step, our preliminary results are focused on bridging the gap at concept level by exploiting semantics, cognitive and affective information, associated with the image verbal content (semantics), which for the needs of our research is the contextual interactional information between the user and the operator of the SEMAINE database [8], keeping fixed the “Where” context-related question. This context concept-based annotation method, that we are examining, allows the system to go beyond a mere syntactic analysis of the semantics associated with fixed window sizes¹. In most of traditional annotation methods, emotions and contextual information are not always inferred by appraisals and thus contextual information about the causes of that emotion is not taken into account [6].

3.2 Approach

In this section, our preliminary results of our proposed semantic context concept extractor, are described in details in [11], and their application to indicative examples validate our proposed approach are presented:

A. Data Corpus: The model here is confronted with the SEMAINE corpus [8]. This audiovisual corpus comprises manually-transcribed sessions with natural emotional displays. These sessions are recorded from two individuals, an operator and a user, interacting through teleprompter screens from two different rooms. The emotions were elicited with the sensitive artificial listener (SAL) framework, where the operator assumes four personalities aiming to elicit positive and negative emotional reactions from the user. Agent’s utterances are constrained by a script, however, some deviations to the script occur in the database.

¹ The window length corresponds to 16 conversational turns and is displayed on figures for future visualization purposes.

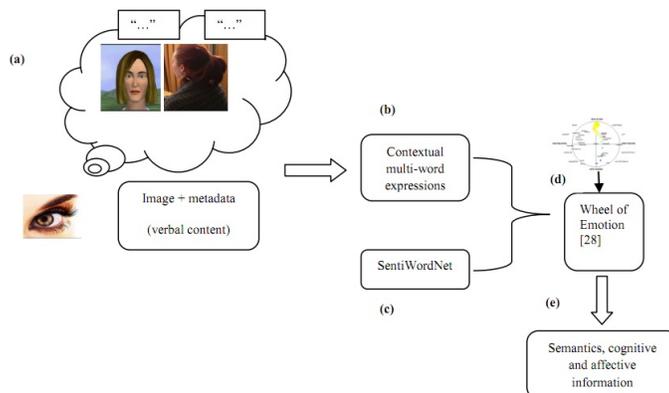


Fig. 2: System’s Overview: **(a)** We discover semantic context concepts from verbal content (semantics) associated with SEMAINE dataset and **(b)** represent each one with multi-word expressions, enhanced with sentiment values **(c)** to enrich a number of Psychological Foundations **(d)**. We finally show that this proposed approach could show a clear connection between semantics, cognitive and affective information prediction **(e)**.

B. Pre-Processing: The pre-processing submodule firstly interprets all the affective valence indicators usually contained in the verbal content of transcriptions, such as special punctuation, complete upper-case words, exclamation words and negations. Handling negation is an important concern in such scenario, as it can reverse the meaning of the examined sentence. Secondly, it converts text to lower-case and, after lemmatizing it, splits the sentence into single clauses according to grammatical conjunctions and punctuation.

These n-grams are not used blindly as fixed word patterns but are exploited as reference for the module, in order to extract multiple-word concepts from information-rich sentences. So, differently from other shallow parsers, the module can recognize complex concepts also when irregular verbs are used or when these are interspersed with adjectives and adverbs, for example, the concept “buy easter present” in the sentence “I bought a lot of very nice Easter presents”.

C. Semantic context concept parser: The aim of the semantic parser is to break sentences into clauses and, hence, deconstruct such clauses into concepts. This deconstruction uses lexicons which are based on sequences of lexemes that represent multiple-word concepts extracted from ConceptNet, WordNet [9] and other linguistic resources.

Under this view, the Stanford Parser ² has been used according to Python NLTK ³; a general assumption during clause separation is that, if a piece of text contains a preposition or subordinating conjunction, the words preceding this function or are interpreted not as events but as objects. Secondly, dependency

² <http://nlp.stanford.edu:8080/parser/>

³ <http://nltk.org>

structure elements are processed by means of Stanford Lemmatizer for each sentence. Each potential noun chunk associated with individual verb chunks is paired with the stemmed verb in order to detect multi-word expressions of the form “verb plus object”. The pos-based bigram algorithm extracts concepts, but in order to capture event concepts, matches between the object concepts and the normalized verb chunks are searched. It is important to build the dependency tree before lemmatization as swapping the two steps result in several imprecisions caused by the lower grammatical accuracy of lemmatized sentences. Each verb and its associated noun phrase are considered in turn, and of more concepts is extracted from these.

D. Opinion and Sentiment Lexicon: Current approaches to concept-level sentiment analysis mainly leverage on existing affective knowledge bases such as ANEW, WordNet-Affect and SentiWordNet [4]. However, for the needs of our current work, we use the SentiWordNet, which is a concept-level opinion lexicon and contains multi-word expressions labeled by their polarity scores.⁴

4 Future Research

My future work will focus on the integration of previous preliminary findings and insights on context affect aware methodologies w.r.t. the production, interpretation and analysis level respectively. To achieve this goal, we have to deal with the following specific challenges:

- 1) *Extraction and Representation:* How can we extract and computationally measure contextually rich features w.r.t. the verbal content?
- 2) *Learning:* What are the proper learning methods needed to build models?
- 3) *Incorporation:* At what time unit should we make an emotion inference (e.g. at the frame or utterance level) and how should we measure the performance of our system?

A. Extraction and Representation

In Affective Computing, all of the state-of-the-art algorithms perform well on individual sentences without considering any context information, but their accuracy is dramatically lowered because they fail to consider context and the syntactic structure of the verbal content (transcriptions) at the same time. Based on the assumption that the context around a sentence or pair of sentences also plays an important role in determining sentiment, I plan to employ in my experiments a conditional random field (CRF) model [7] to capture syntactic, structural and contextual features of sentences. To this aim, I propose to utilize the SEMAINE and RECOLA datasets [8,10] for both human-agent and human-human social

⁴ To avoid the Sentiwordnet’s multi-interpretations a combination of the following methods have been examined: a) Pos tagging to reduce the number of candidate senses, b) Cosine similarity between the sentence and the gloss of each sense of the word in WordNet and c) the “SenseRelate” method to measure the “WordNet similarity” between different senses of the target word and its surrounding words.

and situational interactions⁵, as they provide continuous-time dimensional labels (valence-activation dimensional space).

B. Learning

In the learning stage, I plan to use discriminative learning methods and learn patterns of contextual emotion-specific segments in a supervised manner. My hypothesis is that training contextual rich classification systems using segments lacking clarity may lead to lower analysis rates, and using only segments with high clarity will lead to improved performance as well as more efficient training due to the decrease in training data. I will first extract contextual rich audiovisual descriptors of the above mentioned emotional corpora [8, 10] and then I will utilize these “socially contextually rich” segments with consistent emotional cues and their emotional labels to train a discriminative model. In the case of the RECOLA dataset, in which only the first 5 minutes have been annotated, I will use the active learning method to build classifiers using less training data for expanding labeled data pool.

C. Incorporation

Finally, the goal at the interpretation stage is to show that incorporating context can both improve the system’s performance and disambiguate multimodal behaviors. To this aim, I propose to first identify “socially contextually rich” segments of the test data and incorporate emotion at the segment level using the learned weights of discriminative models. I plan to conduct both classification and regression tasks. I hypothesize that the classification of the “socially contextually rich” segments can increase the regression of all the data, since the captured and longer-range information may be more useful. I will predict continuous-valued labels in dimensional space using regression models such as Support Vector Regression. Also, I will cluster and define new emotion classes from the dimensional labels and identify classes using Support Vector Machines (SVM). The SVM outputs will be combined to infer dimensional space outputs.

5 Conclusion

In this paper, we point out the motivation and importance of identifying or defining the incorporation of context in Affective Computing systems, especially when the disambiguation of a multimodal behavior depends on the contextual interactional setting. Afterwards, we propose the context incorporation framework based on semantic concept detection, extraction enrichment and representation in cognitive estimation. And finally, we provide relevant analysis and conclude to future work that needs to be developed. In future work, we would like to explore the interpretation of contextual effects on affect production, interpretation and analysis respectively.

⁵ In the SEMAINE corpus, situation is determined by the user’s response while in RECOLA corpus the situational context is defined by the roles during the collaborative, intensive and interactional task.

Acknowledgements

This research has been co-financed by the European Union (European Social Fund - EFS) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

References

1. Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P.: Towards a better understanding of context and context-awareness. In: *Handheld and Ubiquitous Computing*. pp. 304–307. Springer (1999)
2. Dey, A.K.: Context-aware computing: The cyberdesk project. *aaai 1998 spring symposium on intelligent environments* (1998)
3. Duric, Z., Gray, W.D., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M.J., Schunn, C., Wechsler, H.: Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. *Proceedings of the IEEE* 90(7), 1272–1289 (2002)
4. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06)*. vol. 6, pp. 417–422 (2006)
5. Graves, A., Fernández, S., Schmidhuber, J.: Bidirectional lstm networks for improved phoneme classification and recognition. In: *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II*, pp. 799–804. ICANN’05, Springer (2005)
6. Gunes, H., Schuller, B.: Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing* 31(2), 120 – 136 (2013)
7. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning*. pp. 282–289. ICML ’01, Morgan Kaufmann Publishers Inc. (2001)
8. McKeown, G., Valstar, M., Cowie, R., Pantic, M., Schroder, M.: The semaine database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing* 3(1), 5–17 (2012)
9. Miller, G.A.: Wordnet: a lexical database for english. *Communications of the ACM* 38(11), 39–41 (1995)
10. Ringeval, F., Sonderegger, A., Sauer, J., Lalanne, D.: Introducing the recola multimodal corpus of remote collaborative and affective interactions. In: *10th International Conference on Automatic Face and Gesture Recognition (FG)*. pp. 1–8. IEEE (2013)
11. Vlachostergiou, A., Caridakis, G., Raouzaoui, A., Kollias, S.: Hci and natural progression of context-related questions. In: *Human-Computer Interaction: Design and Evaluation*, pp. 530–541. Springer (2015)