Uncertainty and Integration of Emotional States in e-Learning
Doctoral Consortium Paper

Grzegorz Brodny
Department of Software Engineering, Gdansk University of Technology, Narutowicza Str. 11/12, 80-233 Gdansk, Poland

Keywords: e-Learning, Emotion Recognition, Late Fusion, Early Fusion, Emotion Representation Models.

Abstract: One of the main applications of affective computing remains supporting e-learning process. Therefore, apart from human mentoring, automatic emotion recognition is also applied in monitoring learning activities. Specific context of e-learning, that happens at home desk or anywhere (mobile e-learning), adds additional challenge to emotion recognition, e.g. temporal unavailability and noise in input channels. Nowadays, affective computing has provided many solutions for emotion recognition. There are numerous emotion recognition algorithms which differ on input information channels, representation emotion model on output and classification method. The most common approach is to combine the emotion information channels. Using multiple input channels proved to be the most accurate and reliable, however there is no standard architecture proposed for this kind of solutions. This paper presents outline of the author's PhD thesis, which concentrates on integration of emotional states in educational applications with consideration of uncertainty. The paper presents state of art, the architecture of integration, performer experiments and planned simulations.

1 RESEARCH PROBLEM

One of the main applications of affective computing remains e-learning processes support. Based on many studies from the fields of pedagogy, it has been confirmed that emotions have a crucial impact on learning and e-learning e.g. (Binali et al. 2009) (Landowska 2013). Therefore, apart from human mentoring, automatic emotion recognition is also applied in monitoring learning activities. Although there are some ethical considerations regarding revealing affective states of a learner to a teacher, there are affective educational systems that were already built. The specific context of e-learning, that happens at home desk or practically anywhere (mobile e-learning), adds another challenge to emotion recognition, e.g. temporal unavailability or noisiness of input channels.

Nowadays, there are numerous emotion recognition algorithms that differ on input information channels, an emotion representation model as output and recognition method. The most important classification is based on input channels, as some are not always available in the e-learning environment. A recognition algorithm might use one or a combination of the following channels:

- visual information from cameras,
- body movements mattes,
- textual input of a user,
- voice signals,
- standard input devices usage,
- physiological measurements.

All of above listed input channels might be applied in the monitoring e-learning activities, but some of them are task- or user-dependent in e-learning context (Landowska et al. 2017) (Landowska et al. 2016).

As all emotion recognition channels are susceptible to some noise, the most common approach is to combine the channels (multimodal recognition) (Poria et al. 2017). This approach requires integration of data or results from different sources. There are two approaches to integration: early and late fusion methods. Both have some disadvantages and the challenge of multimodal integration constitutes the author's research problem. The challenge could be decomposed into the following subproblems:

(1) missing standard emotion representation model; There are many models of emotion representation and unfortunately, there is no standard nor the most frequently used one. As a

Brodny G.
Uncertainty and Integration of Emotional States in e-Learning - Doctoral Consortium Paper.
In *Doctoral Consortium (CSEDU 2017)*, pages 21-27

result, each emotion recognition algorithm and each solution uses another (and sometimes unique) emotion representation model (Gunes and Schuller 2013).

(2) discrepancies between results (recognized emotional states) obtained from different input channels and algorithms; Assuming the same investigation, time and a person, some experimenters observed huge discrepancies between recognized emotional states among different algorithms. Not only solutions based on diverse input channels exhibit the discrepancy, but also the same channel recorded twice (e.g. two camera locations) results in different results (Landowska et al. n.d.) (Landowska and Miler 2016).

(3) uncertainty of results, that differs among algorithms and contexts; Some solutions return a prediction of an emotional state even if conditions for prediction were suboptimal (large camera angle, insufficient lighting, other noise). At the same time, most of the emotion recognition tools do not report the reliability of the predicted state. Moreover, the reported emotional state might be provided using diverse scales and precision.

(4) disadvantages of integration methods; The early fusion method is non-resistant for periodically unavailable channels. The late fusion method is non-resistant for incompatible emotion representation models.

The author's research concentrate on the integration of emotional states in educational applications with consideration of uncertainty. These objectives are described in detail in section 2.

2 OUTLINE OF OBJECTIVES

Author's research concentrates on the following objective: prepare a method of integration of recognized emotional states, taking into account uncertainty. This objective might be decomposed to the following subproblems:

a) selection and application of appropriate methods of mapping between emotion representation models, especially to models that make sense in the e-learning context (the mapping increases measurement error);
b) method for calculation of uncertainty factor for different input channels;
c) integration method based on late fusion, including uncertainty;
d) post-hoc evaluation of emotion recognition, based on efficiency in a specific context;
e) architecture supporting the late fusion of emotion recognition results provided by algorithms from diverse vendors.

3 STATE OF THE ART

This section is divided into 3 parts. The first part presents emotion representation models and approaches to mapping between them. The second provides a review of research about applying affective computing methods in the e-learning, and the latter reviews existing methods of integration of emotional states.

3.1 Emotion Models and Approaches of Mappings

There are multiple emotion representation models and no standard model has been established so far (Valenza et al. 2012). Models fall into three categories: (1) categorical, (2) dimensional and (3) componential. (1) Categorical models are the most intuitive for human, but not for the computers (Gunes and Schuller 2013). They present each emotion as a combination of labelled emotional states. An example of those is a popular Ekman's model, that combines basic emotions: joy, fear, anger, surprise, sadness and disgust to represent complex emotional states (Scherer and Ekman 1984) (2) Dimensional models (usually two- or three-dimensional) represent emotions as compound of bipolar entity for example: valence (pleasant vs unpleasant), arousal (relaxed vs arousal) and dominance/power/control (submissiveness vs dominance) (Gunes and Schuller 2013). Emotions in these models are represented as a point in 2D, 3D or more dimensional space (Grandjean et al. 2008). These models are less intuitive for humans but more easy to be computed by applications. To be understandable for people, the points require some mapping to emotion labels. (3) Componential models of emotions are based on appraisal theory. The models are more complex and concentrate on how emotions are generated (Fontaine et al. 2007) (Grandjean et al. 2008) (Ortony et al. 1988).

Some authors claim that categorical models could be mapped to dimensional ones and vice versa. Some mappings are lossless (Gunes and Schuller 2013).

The researchers use a few mappings which are mainly derived from correlation coefficients. For example mapping between a big five model (categorical) and PAD (3D) was proposed as a function (Mehrabian 1996a) (Shi et al. 2012). The mapping was created by calculating a correlation between the factors from both models and the correlation coefficients were used as weights in the mapping functions. The next case was mapping between PAD and the Ekman model (categorical) (Shi et al. 2012), which was created in an analogical way. Mehrabian and Russel calculated correlation coefficients between PAD and models of personality (Mehrabian 1996b).

The next example is a mapping of the emotion labels to dimensional space proposed by (Hupont et al. 2011). The mapping provides weights that are derived from a database of coordinates from dimensional space to each label. The model can be used directly (e.g. in sentiment analysis). Another method of mapping was used by (Gebhard 2005) in OCC (componential model) to PAD (3D). In this mapping, the points from PAD space were created for each OCC parameters (24) as a label (e.g. anger, fear, distress). The PAD's coordinates were used as weights to mapping functions.

Based on the above review, universal models are dimensional ones, but the final models (which are used by applications) might require adjustments.

3.2 Affective Computing Methods in e-Learning

While applying affective computing methods in e-learning, one might not require the full spectrum of emotions. The most important emotional states, from educational perspective, include: frustration, boredom and flow/engagement (Binali et al. 2009) (Kołakowska et al. 2013) (Landowska 2013).

A few virtual systems have virtual characters, that deal with or visualize affect (Landowska 2008). One of those is the Virtual Human Project that is an educational platform for students with two avatars (a student and a teacher), each with a personality profile (different for each avatar) (Gebhard 2005). In this project OCC model of emotions was used, combined with PAD model for mood and the big five model for personality traits.

Another example of using affective computing in education is an Intelligent Tutoring System (ITS) Eve. Eve was an affect-aware tutoring system, which recognizes (in Ekman states) and expressed affect while teaching mathematics (Alexander et al. 2006).

3.3 Methods of Integration Emotional States

The methods of detecting emotional states could be categorized into four categories: (1) single algorithm (without integration), (2) early fusion, (3) late fusion, (4) hybrid fusion.

3.3.1 Single Algorithm

Nowadays, many affective solution use only one input channel and one emotion recognition algorithm based on that (Hupont et al. 2011). These solutions are very specific, dedicated for one problem and often reveal only one emotion e.g. a positive state, stress or lack-of-stress (Landowska 2013) (Chittaro and Sioni 2014).

3.3.2 Early Fusion (Called Also Feature-Level Fusion)

The early fusion method uses data from multiple input channels that are combined during the data collection step into one input vector (before classification). All data types are processed at the same time. This method usually provides high accuracy (Hupont et al. 2011), but becomes more challenging as the number of input channels increases. The main challenges in this method include:

a) Time synchronization for data from each channel (resulting in incomplete feature vectors);
b) Learning a classifier with vectors containing missing values (when channels are inaccessible);
c) Large feature vectors when fusing many channels, (feature selection techniques are used to maximize the performance of the classifier) (Gunes and Piccardi 2005);
d) Adding a new channel/module often requires retraining and/or rebuilding all solutions (low scalability).

3.3.3 Late Fusion (Also Called Decision-Level Fusion)

In late fusion, in contrast to early fusion, integration of data is performed during decision step. This method is based on an independent processing of data from each input channel and training multiple classifiers. Each of the classifiers provides one hypothesis on emotional state. The integration function provides a final estimate of emotional state based on partial results. This method provides more scalability than the early fusion because a new module is just one more result to integrate. The main challenges in the method include:

a) Time synchronization for data from diverse modules – integrate a subset of results or wait for all modules to provide a hypothesis?
b) Mapping output from modules to one, final output model.

3.3.4 Hybrid Fusion

The hybrid methods are a combination of late and early fusion. Each module has a separate classifier as in the late fusion but also has access to input data from all input channels. The main advantage of this method is preservation of algorithms independence, while still using combined information from multiple channels. However, the challenges remain more less the same as in the late fusion method.

3.3.5 Summary of Fusion Method

The approach used in this research is a late fusion method, with a potential extension to a hybrid fusion. The early fusion method is difficult to maintain and extend with new observation channels. Moreover, the early fusion approach is not possible with the use of existing off-the-shelf solutions, including commercial software. Late or the hybrid fusion method supports integration, exchange and modifiability of modules for emotion recognition ("black box" approach).

4 METHODOLOGY

The research methodology in the presented PhD work is in general based on experiments and simulations. To compare the algorithms accuracy experiments were carried with different input channels. Experiments were used also to collect the data for simulations. Simulations can be carried offline, using the real data from experiments and data available in emotional databases (Cowie et al. 2005).

4.1 Experiments

Three experiments were performed so far.

4.1.1 The Experiment 1. Learning via Playing a Educational Game

The goal of the experiment was to investigate emotional states while learning using an educational game (Landowska and Miler 2016). The game was about managing IT projects and participants were computer science students. The participants were asked to play a game several times and both their emotional state as well as educational outputs were measured.

The emotion recognition channels in these experiments were: facial expressions, self-report, physiological signals. Details of the experiment were described in (Landowska and Miler 2016). This PhD work will use the data from the experiment to perform off-line simulations of the proposed integration methods.

4.1.2 The Experiment 2. Learning with a Moodle Course

The aim of the experiment was to investigate emotional states while using a Moodle course with diverse activities. Three Moodle activity types were employed: watching a lecture, solving a quiz and adding a forum entry on a subject pre-defined by a teacher. In this experiment, the simultaneous 4 cameras recordings were used for facial expression analysis. Self-report, physiological measurements and sentiment analysis of textual inputs were also employed (Landowska et al., 2017).

4.1.3 The Experiment 3. Learning with on-Line Tutorials

The aim of the experiment was to investigate emotional states while learning using video tutorials of Inkscape tool (Landowska et al. n.d.).

The emotion recognition channels in these experiments were: facial expressions recording (2 cameras), keystroke dynamics, mouse movements patterns, opinion-like text and self-report.

4.1.4 Summary of Experiments

After carrying the three experiments some general observations were made:

- a location of the camera is one of the crucial factors influencing recognized emotional states.
- availability of some emotion observation channels is task-dependent (e.g. sentiment analysis depends on writing tasks) and/or

user-dependent,

- physiological signals provide information only on arousal and not on the valence of an emotional state,
- self-report is the most dependent on human will and should be confirmed with another observation channel;
- peripherals (mouse/keyboard) usage patterns reveal information on affect with relatively low granularity and accuracy and should be combined with other observation channels.

These observations confirmed the assumption on multichannel observation having a potential for improving accuracy of emotion recognition.

4.2 Simulations

After collecting data from experiments and from available databases a set of simulations will be carried out. This section was divided into a few parts. The first one presents the method of integration. The second part provides preliminary simulation plans.

4.2.1 Method of Integration

The method of emotional states integration used in this research was a part of Emotion Monitor, which was described in (Landowska 2015). The concept of the stand assumed combining multiple modalities used in emotion recognition in order to improve accuracy of affect classification. A model of Emotion Monitor architecture was presented in Figure 1. An area covered by this PhD research is mark with a dotted line. Integrated algorithms are treated as in "black-box" approach and can use the early fusion mechanism on algorithm level (as in hybrid fusion method). Algorithms get input data from emotion observation channels and provide the hypothesis using some emotion representation model. In the next step, each of the algorithms' hypothesis must be mapped to one common emotion model (if needed). Next, the integration function combines partial hypotheses to an integrated state, which could be sent to the application as a final decision on recognized emotional state.

The proposed methods and architecture aims at addressing the problem 2e defined in the objectives section.

4.2.2 Preparation of Simulations

Before simulations could be started the following prerequisites must be met:

a) Collect the algorithms to integrate, which differ input channels and using different emotion representation models.
b) Prepare an integration function that can be tested in simulations.
c) Prepare method of calculating uncertainty factor for different input channels.
d) Collect the data with labels in different emotion models (preferably labels in two or more models in one database).

4.2.3 Simulation 1 – Compare Accuracy of the Algorithms

The first simulation helps to constitute a ranking of algorithms with accuracy rates.

A plan is to compare the algorithms, using exactly the same input data, from the same experiment and the same channels. Input data should be labelled with emotions in emotion models being outputs of tested algorithms (without mapping).

The results from these simulations are required for the next simulations.

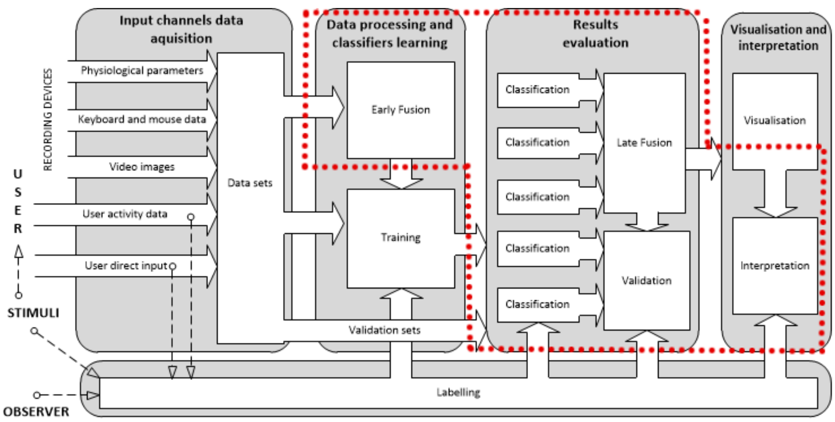

Figure 1: Conceptual model of multimodal emotion recognition fusion and the scope of integration solution.

4.2.4 Simulation 2 – Evaluate a Measurement Error of Mapping

The aim of this simulation is a selection of the optimal mapping models and evaluation of error imposed by the transformation.

Data: Can be different for each algorithm. Preferably using two or more emotion model labels in one input set. The same data as Simulation 1 might be used if additionally labelled.

Plan: For each algorithm, for each set from datasets: (1) Provide input data to an algorithm, (2) get results (an emotional state estimate), (3) perform mappings. (4) Calculate the accuracy of results after mapping. (5) Calculate the mapping error – difference between accuracy obtained from simulation 1 and after mapping.

The simulation might allow choosing the best mapping method.

4.2.5 Simulation 3 – Evaluate the Uncertainty

The aim of this simulation is verification of a method for calculating uncertainty. This simulation should address the problem 2b.

Data: Input data from one of the experiments, but using different settings e.g. different camera locations.

Plan: For each algorithm or at least one for each channel: (1) Provide an algorithm with data from different sources. (2) Calculate the accuracy for each source independently with uncertainty factor. (3) Compare accuracy and uncertainty factor for each source and integrated result.

If a function of calculating uncertainty is correct, an integrated result is expected to be less uncertain and more accurate.

4.2.6 Simulation 4 – Evaluate the Function of Integration

The aim of the simulation is to verify the integration function using uncertainty factor. This simulation should answer to the problem 2c.

Data: The same data as in simulation 3.

Plan: Integrate an emotional states from simulation 3. Compare accuracy of integrated states with states from partial algorithms.

5 EXPECTED OUTCOME

The main expected outcome of Author's PhD thesis is preparing a method of integration of recognized emotional states, taking into account uncertainty. This method should improve the accuracy of emotion recognition. It should also allow applying some of the off-the-shelf software to different contexts, especially concentrating on e-learning. As an expected long-term result, the integration method should be applicable to e-learning platforms and educational games, which aim at supporting learners in maintaining attention and positive attitude in educational processes.

6 STAGE OF THE RESEARCH

This paper presented the outline and selected details from Author's PhD thesis. This section summarizes, which parts of research were already done, which are in progress and which are not started yet.

The first version method of integration based on late fusion, including uncertainty, was already implemented. It was implemented in C# and passed the basic tests. It's waiting for validation in Simulation 4.

All of three planned experiments were completed. Some data from external emotion databases are obtained, but author is still looking for more data in the next steps.

The architecture supporting the late fusion of emotion recognition algorithm provided by diverse vendors was developed and the first implementation was already tested, revealing some potential for improvement. The second implementation of the architecture is in progress. A paper about the architecture is prepared.

Some algorithms were collected and prepared (implementing wrappers) to integrate with the architecture. The method for calculation of uncertainty factor for different input channels is designed and currently under development.

Some mappings between emotion models were collected, but the list is not closed yet. Simulations regarding mapping accuracy are planned as the first step to follow.

ACKNOWLEDGEMENTS

This work was supported in part by Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no Pol-Nor/209260/108/2015 as well as by DS Funds of ETI Faculty, Gdansk University of Technology.

REFERENCES

Alexander, S., Sarrafzadeh, A. & Hill, S., 2006. Easy with eve: A functional affective tutoring system. *and Affective Issues in ITS. 8th.*

Binali, H.H., Wu, C. & Potdar, V., 2009. A new significant area: Emotion detection in E-learning using opinion mining techniques. In *2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, DEST '09*. IEEE, pp. 259–264.

Chittaro, L. & Sioni, R., 2014. Affective computing vs. affective placebo: Study of a biofeedback-controlled game for relaxation training. *International Journal of Human-Computer Studies*.

Cowie, R., Douglas-Cowie, E. & Cox, C., 2005. Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural Networks*, 18(4), pp.371–388.

Fontaine, J.R.J. et al., 2007. The world of emotions is not two-dimensional. *Psychological Science*, 18(12), pp.1050–1057.

Gebhard, P., 2005. ALMA – A Layered Model of Affect. *AAMAS '05: Proceedings of the 4th international joint conference on Autonomous agents and multiagent systemsent systems*, pp.29–36.

Grandjean, D., Sander, D. & Scherer, K.R., 2008. Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*.

Gunes, H. & Piccardi, M., 2005. Affect recognition from face and body: early fusion vs. late fusion. *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, 4, pp.3437–3443.

Gunes, H. & Schuller, B., 2013. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2), pp.120–136.

Hupont, I. et al., 2011. Scalable multimodal fusion for continuous affect sensing. In IEEE.

Kołakowska, A., Landowska, A. & Szwoch, M., 2013. Emotion recognition and its application in software engineering. *(HSI), 2013 The 6th.*

Landowska, A., 2013. Affective computing and affective learning--methods, tools and prospects. *EduAction. Electronic education magazine*, 1(5), pp.16–31.

Landowska, A., 2015. Emotion monitor - Concept, construction and lessons learned. In *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015*. pp. 75–80.

Landowska, A. Brodny, G., 2017. Postrzeganie inwazyjności automatycznego rozpoznawania emocji w kontekście edukacyjnym. *EduAction. Electronic education magazine*, pp.1–18., Submitted.

Landowska, A., 2008. The role and construction of educational agents in distance learning environments. In *Proceedings of the 2008 1st International Conference on Information Technology, IT 2008*.

Landowska, A., Brodny, G. & Wrobel, M.R., Limitations of emotion recognition from facial expressions in e-learning context. Submitted.

Landowska, A. & Miler, J., 2016. Limitations of Emotion Recognition in Software User Experience Evaluation Context. In *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*. pp. 1631–1640.

Mehrabian, A., 1996a. Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model. *Australian Journal of Psychology*, 48(2), pp.86–92.

Mehrabian, A., 1996b. Pleasure-Arousal . Dominance : A General Framework for Describing and Measuring Individual Differences in Temperament. *Current Psychology*, 14(4), pp.261–292.

Ortony, A., Clore, G.L. & Collins, A., 1988. *The Cognitive Structure of Emotions*,

Poria, S. et al., 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*.

Scherer, K.R. & Ekman, P., 1984. *Approaches to emotion*, L. Erlbaum Associates.

Shi, Z. et al., 2012. Affective transfer computing model based on attenuation emotion mechanism. *Journal on Multimodal User Interfaces*, 5(1–2), pp.3–18.

Valenza, G., Lanata, A. & Scilingo, E., 2012. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE transactions on*.