TOWARDS AN INTERACTIVE E-LEARNING SYSTEM BASED ON EMOTIONS AND AFFECTIVE COGNITION

Panagiotis Vlamos
Department of Informatics

Andreas Floros
Department of Audiovisual Arts

Michail N. Giannakos
Department of Informatics

Konstantinos Ch. Drossos
Department of Audiovisual Arts
Ionian University
Greece

Abstract

In order to promote a more dynamic and flexible communication between the learner and the system, we present a structure of a new innovative and interactive e-learning system which implements emotion and level of cognition recognition. The system has as inputs the emotional and cognitive state of the user and re-organises the content and adjusts the flow of the course. Our concept aims to increase the learning efficiency of intelligent tutoring systems by using a combination of characteristics, such as content customization and user's emotion recognition, and adapting all these features into a learner-centered educational system.

Introduction

Intelligent tutoring systems (ITSs) are computer programs that are designed to incorporate techniques from the AI community in order to provide tutors which know what they teach, who they teach and how to teach it (Wenger, 1987).

Recent research (Weerasinghe & Mitrovic, 2005) supports that the interaction-reaction process between the instructor and the learner during a "traditional" lesson has an important, and positive, effect at the effectiveness of the teaching process. With the incremental infiltration of the distance-learning means in the teaching process and the deficiency of most distance-learning applications to provide the capability of real-time alteration of the teaching material (flow, organisation, difficulty) (Brusilovsky, 1999;

Rickel, 1989), the research for new, “clever” and interactive applications seems to be more that necessary. In favour for the above claims, stand also:

- the lack of direct communication of the instructor and the learner, in most distance-learning applications, which suggests a negative factor at the establishment of the aforementioned interaction-reaction process and

- the nature of most distance-learning applications, which, in most cases, is a static hypertext network (Brusilovsky, 1999; Keles, Ocak, Keles, & Gólcó, 2009), even if the development of new forms of distance-learning applications is in the scope of many researches (Angelides & Paul, 1993; Brusilovsky, 1999).

Thus, this paper proposes a new, “clever” and interactive application of distance-learning which will be capable of

- utilising the interaction-reaction process, which, in this case, consists of affective interaction-reaction between the instructor and the learner, in order to maximise the effectiveness of the teaching process, and real time decision of the adaptation of the teaching material, according to learner’s level of understanding,

- providing immediate and real time capability of reorganisation of the teaching material (flow, difficulty),

- providing the instructor’s interface by an intelligent virtual teacher with realistic affective characteristics.

It has to be mentioned that researches (Burleson & Picard, 2004; Chaffar & Frasson, 2004; D’mello, Craig, Gholson, & Franklin, 2005) point out the strengthen of effectiveness of the teaching process by the use of driven affective interaction–reaction by the instructor to the learner.

This paper is organized as follows. First, the section Recent Work examines related previous work after explaining the technology assist in education process. Following, the description of emotion recognition states and the definition of the level of cognition are in sections Emotions’ Recognition and Level of Cognition respectively. The section Modeling of Educational Content analyzes the modeling of educational content which is used in our system and the section Proposed System presents our proposed system end specifies you it could be worked. Last, but not least, Feature Work presents our conclusions and indicates future work.

Recent Work

The first introduction of the use of technology assisting means in the education process can be dated back in the 1960's, when a few educational applications used radio, T.V. and even cinema as assistive educational means (Kokos, Lionarakis, Matralis, & Panagiotopoulos, 1998). Thirty-five years passed and in 1995 there were the first implementations of the intelligent tutoring systems (ITS) (Brusilovsky, 1995). But even if ITS have great advantages they seem to lack compared to the traditional educational process, i.e. compared to the direct and real time communication and interaction of the tutor and the student in a classroom (Sarrafzadeh, Alexander, Dadgostar, Fan, & Bigdeli, 2008). Thus, the thought that the ITS would significantly improved if they would be able to recognise the emotional state of their user started to bloom among researches and, finally, gave birth to affective tutoring systems (Mao & Li, in press). As the term "affective tutoring systems" implies, ATS are tutoring systems capable of knowing and "understanding" their users' emotional state and responding accordingly (Aylett, Paiva, Dias, Hall, & Woods, 2009). Despite the fact that the term "affective tutoring system" can be traced back to 1997 (Sarrafzadeh et al., 2008), ATS first developed around 2003 (Alexander, Sarrafzadeh, & Fan, 2003) and recent work includes only a few implemented such systems.

One of them is "Easy with Eve", an affective tutoring system implemented for mathematics in primary school, developed in New Zealand and Australia (Sarrafzadeh et al., 2008). Easy with Eve included emotion recognition, based on gestures and face expressions tracking and recognition, interaction between the intelligent tutor and the student, and case-based method of adapting to student emotional states for the implementation of the educational process. Another recent ATS was developed in Tunisia and France, incorporating emotional recognition from video by tracking the facial expressions of the user (Ben Ammar, Neji, Alimi, & Gouarderes, 2009). As stated by the developers, their system is a multi-agent system who manages the cognitive and affective model of the user, while the teaching process is implemented by another embodied agent. Their ATS system is capable of recognising the affective state of the user and decides what teaching strategy to follow accordingly. Nevertheless, both of the aforementioned systems have little or no sound emotion recognition, e.g. from speech, implemented and according to the authors' knowledge, there are no other ATS implemented and/or published at the time being.

Emotions' Recognition

According to recent researches and their results, there is a suggestion that emotions have an important role in decision making, problem solving and intelligence in general (Cooper, Brna, & Martins, 2000). Moreover, and as stated previously in this paper, one of ATS's the key features is user's/student's emotional state recognition. It can be implemented with various ways, including tracking of human physiological reactions, e.g. skin inductance (Sunjung & Andre, 2004), or extracting useful features from audio and video channels, like facial expressions recognition (Ben Ammar et al., 2009), gestures recognition (Sarrafzadeh et al., 2008) and speech analysis (Benzeghiba et al., 2007).

Usually, in a traditional teaching environment, e.g. a classroom, the tutor can understand, or try to understand, a student's emotional state by compensating the facial expressions, gestures and voice characteristics of the student. Thus, this paper will focus on the audio and video channels, considering that sound and moving image are the key elements of emotional communication and interaction between the tutor and the student, even in a video call. Hence, this twofold approach to a student's emotional state recognition is proposed in order to try to simulate the traditional affective communication and interaction between the two principal participants of the teaching process, i.e. the tutor and the student.

Emotion Recognition from Speech

According to recent research on the field of emotions' recognition from voice/speech, there is the necessity for two processes in order to achieve the targeted recognition and these are a) speech signal processing, in order to extract the useful technical parameters for the emotions' recognition process and b) emotions' recognition. The results of the aforementioned researches are positive with 90% of distinct emotions' recognition (Borchert & Dusterhoft, 2005). In addition, is mentioned an: a) 85% success in emotions' recognition with the use of multiresolution analysis of wavelet packets method (Hyoun-Joo, Keun-Chang, Dae-Jong, & Myung-Geun, 2003); b) a percentage for success in recognition of 60-705 for joy-happiness, 70–80% for anger and 70–85% for sadness, with the use of neural networks for emotions' recognition and extracted technical features as acoustic energy of speech, phonemes and characteristics of frequency spectrum (Petrushin, 1999); and c) 68.59% total percentage of success, using linear prediction method for the extraction of the technical parameters and fuzzy model for emotions' recognitions (Razak, Yusof, & Komiya, 2003). Here has to bementioned that there are existing application which allow real time emotion recognition from speech (Vogt, André, & Bee, 2008).

Emotion Recognition from Moving Image

There are many researches and data bases regarding facial or face analysis through video, like Cohn, Zlochower, Lien, and Kanade (1999) and Lyons, Akamatsu, Kamachi, and Gyoba (1998). DeSilva, Miyasato, and Nakatsu (1998) studied the human ability to distinguish emotions through video and moving image. Results indicate than human can recognise anger, dislike, pleasure and surprise from video with a success rate of 47%. Moreover, recent ATS have implemented emotion recognition through video channel (Ben Ammar et al., 2009; Sarrafzadeh et al., 2008).

Level of Cognition

Cognition was found to be most amenable to study and they have been lot of principles developed for its study (Bloom, 1956). For example, people seem to learn better from words and pictures together than from words alone (Fletcher & Tobias, 2005; Mayer, Heiser, & Lonn, 2001). Most other multimedia principles entail finding the best way to organize and present visual and verbal information to learners. Multimedia design principles, are coherence, signaling, redundancy, modality, contiguity (split attention), segmenting, and pre-training principles (Nelson & Erlandson, 2008). Except from these, our proposed system uses the identification of the educational material according to learner needs. The aim of the proposed system is to increase learner's attention by changing the educational content dynamically. The dynamic change of the content based on static characteristics (performance in similar courses, general level, lesson difficulties etc) and dynamic characteristics, exported from learners camera and speaker (attention, mood, etc.).

The notion of cognitive parameters in learning is a key issue. It is also well known for the difficulty of integration into e-learning procedures. According to Bloom (1956) "[the cognitive domain] includes those objectives which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills." We focus to locate the abilities and characteristics of the learner, and then to create an education material according to their needs. Through this identification our system scopes to increase the cognition level of knowledge.

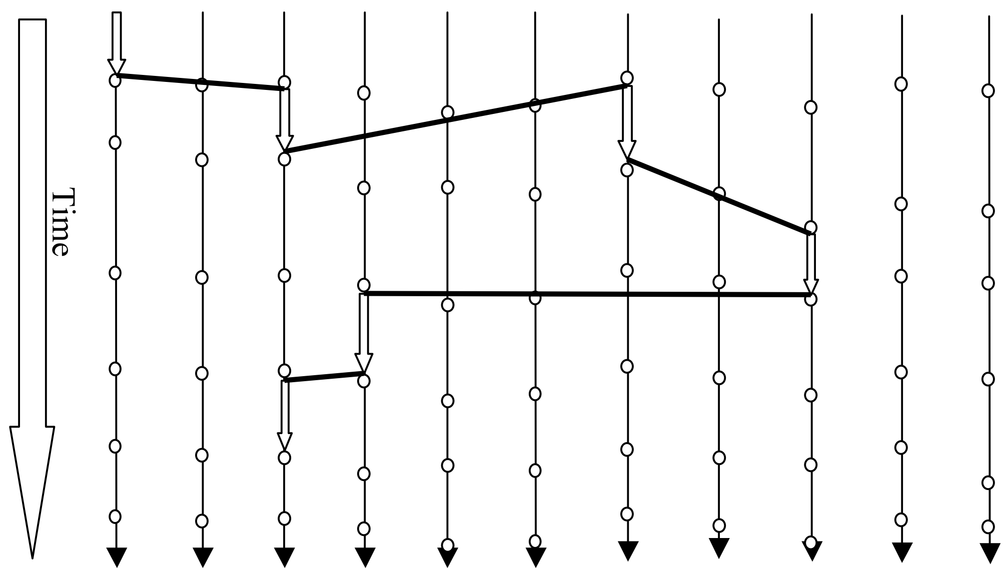
Modelling of Educational Content

Learning technology standards are being developed by a variety of institutions and regular standardization forums. Some international initiatives such as the IMS project have developed relevant e-learning specifications to model learning content, sequencing, quizzes and repositories while the ADL initiative in the U.S. has developed its own proposal named SCORM, which has a learning-content packaging format with additional sequencing extensions. Reference models such as SCORM have been proven effective in providing interoperability of content and course materials across delivery platforms. Metadata standards such as IEEE LOM (IEEE, 2002) and the Dublin Core (Dublin, 2010) provide an effective way to describe and catalog individual content objects. But content and system interoperability combined with content tagging and management are insufficient.

From a technical point of view, there is a distinction between a) passive content, which can be simply delivered as it is, using any technology, and b) more complex learning material, such as learning processes or instructional functionalities, which require more meaningful and complex LMS services. Our proposal corresponds in the last case, we hereby propose a multi adaptive method to organize and customize the educational content into learner needs.

More specific, our approach is based on the design of a large number of courses. These courses will differ, mainly, in content, pedagogical methodology, velocity of knowledge transmission and they are expected to have as a result the learning effectiveness. According to **Figure 1** every thin parallel line is the different courses, the nodes are the points where we can change the course according to the camera and microphone feedback from the learner. The bold lines are an example route of a specific course. There is no limitation in content or time that we have to spend. Each node can be independent from the other nodes but also be necessary for the whole system. The combination of content fragmentation and the dynamic option based on learner characteristics and behavior will help us to make the content adaptive.

Figure 1: Schematic representation of adaptive content



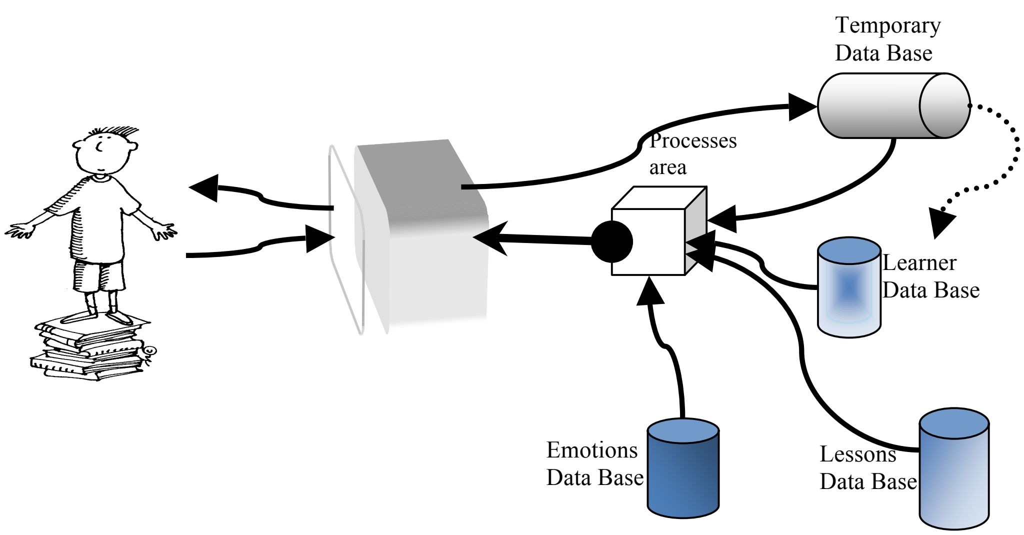
The Proposed System

Tutoring systems are computer-based learning systems that seek to reflect new methods of teaching and learning based on one-to-one interaction. To be classified as “intelligent”, they must present “human-like” tutoring capabilities (Wenger, 1987). Hence they should be able to adjust the content and delivery to students’ characteristics and needs by analysing and/or anticipating theirs affective responses and behaviors.

Our proposed affective tutoring system consists of a temporary data base, learner data base, lessons data base, emotions data base, processes area and a learner’s interactive system. In order to accomplish better comprehension all the parts of the system would be analyzed separately. Firstly the temporary data base would be empty at the beginning of ever session. Then data from the situation (e.g. emotional, mood, high or low comprehension) of the learner at this session will fill this data base, these data will be used from processes area during the educational procedure and at the end of the session some of the data will enrich learners data base. The learner’s data base will store learners file, that file includes all learner’s courses and performance archive. The lessons data base will be a static base consisting of the content of the courses at every version according to content model, quirks of every course and pedagogical techniques for each different science. Emotions data base will be the responsible for classifying and recognising user's emotions, based on two, or more, dimensions emotional model, having as dimensions key features for emotion recognition, like the valence–arousal model (Russell, 2003). Emotions' recognition will be held according to geometric areas on the dimensions of the emotion model, in which has been identified that this particular area corresponds in a particular emotion, according to the analysis of the technical features extracted from the channels of moving image and voice/speech.

Learner's interactive system transfers the course into the learner and learner's feedback into the temporary data base according to emotion recognition, based on emotions' recognition from moving image and voice/speech. The processes area feeds about learner's condition from the temporary data base, about learner's archive and characteristics from learner's data base, about lessons characteristics and content from lessons data base, until this point processes area will decide the next content then emotions data base will formulate the content according to emotional needs. This algorithm will execute at each node of the content. A schematic representation of the proposed system can be seen in **Figure 2**.

Figure 2: Proposed system's schematic representation



Future Work

Concluding, there is more to be done regarding the proposed system. One important factor, following a, possibly, low fidelity implementation of the proposed system, will be an assessment of the resulting enhancement in the educational process induced by the usage of the proposed system. The feedback returned from the assessment will point the necessary changes for the enhancement of the system itself which will lead to a cross assessment of the proposed system with the traditional educational procedure and/or other ATS/ITS. Moreover, it is hoped that the resulting system will help in the development of the education in areas where there is lack in education personnel, by contributing to the implementation of courses that cannot be carried out due to the aforementioned lack.

References

ADL. *Advanced Distributed Learning*. Retrieved February 15, 2010, from http://www.adlnet.gov

Alexander, S., Sarrafzadeh, A., & Fan, C. (2003). *Pay attention! The computer is watching: Affective tutoring systems*. Paper presented at the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003.

Angelides, M. C., & Paul, R. J. (1993). Developing an intelligent tutoring system for a business simulation game. *Simulation Practice and Theory, 1*(3), 109–135.

Aylett, R., Paiva, A., Dias, J., Hall, L., & Woods, S. (2009). Affective agents for education against bullying. *Affective Information Processing* (pp. 75–90).

Ben Ammar, M., Neji, M., Alimi, A. M., & Gouarderes, G. (2009). The affective tutoring system. *Expert Systems with Applications.*

Benzeghiba, M., Mori, R. D., Deroo, O., Dupont, S., Erbes, T., Jouvet, D. et al. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication, 49*(10–11), 763–786.

Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook 1: Cognitive domain* (2nd ed.). New York: Longman, Green & Company.

Borchert, M., & Dusterhoft, A. (2005). *Emotions in speech — Experiments with prosody and quality features in speech for use in categorical and dimensional emotion recognition environments*. Paper presented at the 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE '5.,

Brusilovsky, P. (1995). *Intelligent tutoring systems for World-Wide Web*. Paper presented at the Third International WWW Conference (Posters).

Brusilovsky, P. (1999). Adaptive and intelligent technologies for web-based education. *Special Issue on Intelligent Systems and Teleteaching, Künstliche Intelligenz, 4*, 19–25.

Burleson, W., & Picard, R. W. (2004). *Affective agents: Sustaining motivation to learn through failure and a state of stuck*. Paper presented at the Social and Emotional Intelligence in Learning Environments Workshop in conjunction with the 7th International Conference on Intelligent Tutoring Systems.

Chaffar, S., & Frasson, C. (2004). *Inducing optimal emotional state for learning in intelligent tutoring systems*. Paper presented at the Intelligent Tutoring Systems, 7th International Conference ITS 2004.

Cohn, J. F., Zlochower, A. J., Lien, J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology, 36*(1), 35–43.

Cooper, B., Brna, P., & Martins, A. (2000). Effective affective in intelligent systems - building on evidence of empathy in teaching and learning. *Affective interactions: towards a new generation of computer interfaces* (pp. 21–34): New York : Springer-Verlag.

D'mello, S. K., Craig, S. D., Gholson, B., & Franklin, S. (2005). *Integrating affect sensors in an intelligent tutoring system*. Paper presented at the In Affective Interactions: The Computer in the Affective Loop Workshop at 2005 Intl. Conf. on Intelligent User Interfaces.

DeSilva, L. C., Miyasato, T., & Nakatsu, R. (1998). Use of multimodal information in facial emotion recognition. *IEICE transactions on information and systems, E81*(D(1)), 105–114.

Dublin. (2010). *Dublin Core Metadata Initiative.* Retrieved February 20, 2010, from http://www.dublincore.org

Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 117–134). New York: Cambridge University Press.

Hyoun-Joo, G., Keun-Chang, K., Dae-Jong, L., & Myung-Geun, C. (2003). *Emotion recognition from the facial image and speech signal*. Paper presented at the SICE 2003 Annual Conference

IEEE. (2002). *IEEE Standard for Learning Object Metadata, IEEE-SA* Standard 1484.12.1-2002. New York.

Keles, A., Ocak, R., Keles, A., & Gólcó, A. (2009). ZOSMAT: Web-based intelligent tutoring system for teaching-learning process. *Expert Systems with Applications, 36*(2, Part 1), 1229–1239.

Kokos, A., Lionarakis, A., Matralis, C., & Panagiotopoulos, C. (1998). *Open and distance learning education, education material and new technologies* (Vol. C): Hellenic Open University.

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). *Coding facial expressions with Gabor wavelets*. Paper presented at the 3rd International Conference on Face\& Gesture Recognition.

Mao, X., & Li, Z. (in press).Agent based affective tutoring systems: A pilot study. *Computers & Education*.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198.

Nelson, B., & Erlandson, B. (2008). Managing cognitive load in educational multi-user virtual environments: Reflection on design practice. *Educational Technology Research and Development, 56*(5), 619–641.

Petrushin, A. V. (1999). *Emotion in speech: Recognition and application to call centers*. Paper presented at the Conference on Artificial Neural Networks in Engineering (ANNIE '99).

Razak, A. A., Yusof, M. H. M., & Komiya, R. (2003). *Towards automatic recognition of emotion in speech*. Paper presented at the 3rd IEEE International Symposium on Signal Processing and Information Technology.

Rickel, J. W. (1989). Intelligent computer-aided instruction: A survey organized around system components. *IEEE Transactions on Systems, Man, and Cybernetics, 19*(1), 40–57.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172.

Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., & Bigdeli, A. (2008). “How do you know that I don't understand?” A look at the future of intelligent tutoring systems. *Computers in Human Behavior, 24*(4), 1342–1363.

Sunjung, K., & Andre, E. (2004). *Composing affective music with a generate and sense approach*. Paper presented at the Seventeenth International Florida Artificial Intelligence Research Society Conference.

Vogt, T., André, E., & Bee, N. (2008). EmoVoice — A framework for online recognition of emotions from voice. *Perception in Multimodal Dialogue Systems* (pp. 188–199).

Weerasinghe, A., & Mitrovic, A. (2005). *Using affective leaner states to enhance learning*. Paper presented at the 9th International Conference, KES 2005.

Wenger, E. (1987). *Artificial intelligence and tutoring systems*. Los Altos, CA: Morgan Kaufmann.