Object oriented architecture for affective multimodal e-learning interfaces

Efthymios Alepis and Maria Virvou
Department of Informatics, University of Piraeus, 80 Karaoli & Dimitriou St., 18534 Piraeus, Greece

Abstract. This paper focuses on the architecture of affective multimodal user interfaces that exist in e-learning applications. The OO (Object Oriented) approach has been adopted in order to combine evidence from multiple modalities of interaction and data from emotion stereotypes and classify them into well structured objects with their own properties and methods. The proposed architecture can be adopted in future multi-purpose emotion recognition systems with multiple modalities and improved emotion detection algorithms. Furthermore, the resulting emotion detection server is capable of using and handling transmitted information from different sources of data during human-computer interaction. The evaluation of the OO architecture for affective interaction by programmers and teachers revealed that there are considerable improvements in the resulting e-learning system, including easiness in adding a modality of interaction, better structure for the available multi-modal information and user friendlier environment for the human-computer interaction. The evaluation of the resulting system provided quite satisfactory results considering the proposed affective multimodal architecture.

1. Introduction

The exploration of how humans interact with their environment and with each other constitutes one of the major scientific challenges of our century. Perceiving, learning and adapting to the world around us are commonly labeled as intelligent behavior [1]. In many situations human-computer interaction may be improved by multimodal emotional interaction in real time [2,3]. Affective computing has recently become a very important field of research because it focuses on recognizing and reproducing human feelings within human computer interaction. Human feelings are considered very important but only recently have started being taken into account in human-computer interaction. Thus, the area of affective computing is not yet well understood and needs a lot more research to reach maturity.

As Picard claims in [4], one of the major challenges in affective computing is to try to improve the accuracy of recognizing people’s emotions. During the last decade, the visual and the audio channel of human-computer interaction were considered as most important in human recognition of affective states [5]. Yet, research in psychophysiology has produced firm evidence that affective arousal has a range of somatic and physiological correlates, such as heart rate, skin clamminess, body temperature, etc. [6]. Correspondingly, a reasonable suggestion for the improvement of accuracy in affect recognition is the combination of more than one modes of interaction in user interfaces. It is hoped that the multimodal approach may provide not only better performance, but also more robustness [1].

Similar views about the benefits of the combination of modalities have been supported by many researchers in the field of human-computer interaction [7–11]. However, progress in emotion recognition based on multiple modalities has been rather slow. Although several approaches have been proposed to recognize human emotions based on facial expressions or speech uni-modally, relatively limited work has been done to fuse these two and other modalities to improve the accuracy and robustness of the emotion recognition systems [12].

The issue of combining multiple modalities raises the problem of how these modalities may be combined. Correspondingly, this problem consists of the determination of a general architecture of a multi-modal
emotion recognition system, as well as of the sophisticated mechanisms that will fuse this system’s available data in order to utilize the emotion recognition functions. In fact, the mathematical tools and theories that have been used for affect recognition can lead to a classification of affect recognizers. Such a classification has been made in [13] where affect recognizers have been classified into two groups on the basis of the mathematical tools that these recognizers have used: 1. The first group using traditional classification methods in pattern recognition, including rule-based systems, discriminate analysis, neural networks, Bayesian learning and other learning techniques. 2. The second group of approaches using Hidden Markov Models, Bayesian networks etc. Indeed, a recent piece of research that uses the above approaches for the integration of audio-visual evidence is reported in [14]. Specifically, for person-dependent recognition, Zeng and his colleagues [11] apply the voting method to combine the frame-based classification results from both audio and visual channels. For person-independent tests, they apply multi-stream hidden Markov models (HMM) to combine the information from multiple component streams.

The fore mentioned approaches indicate methodologies and algorithms that try to combine multi-modal information efficiently, and all of them present advantages in the field of affective computing. However, after a thorough investigation in the related scientific literature we found that there is a shortage of studies that aim at the theoretical data structures of multimodal information systems, aside from their algorithmic models. Furthermore, the well-known Object Oriented Method (OOM) has not been found to be incorporated in current emotion recognition systems as architecture for modeling them efficiently and reliably.

In view of the above, in this paper we present an object-oriented model for emotion recognition purposes. As a next step, we incorporate this model into an e-learning system in order to evaluate the resulting system. The object-oriented model is used to represent significantly the available information from the modalities of interaction, as well as information about the users and their actions. The proposed model incorporates all common object oriented concepts, which are described explicitly in the next section, trying to provide exploitable information, more robustness for the algorithms that are going to use the available data and easiness in the addition of potential new modalities for emotion recognition.

In Section 2 we make a sort overview of the OOP Method and the concepts that have been used for the generation and the analysis of object oriented models. In Section 3 we present the general architecture of the object oriented emotion recognition system, while in Section 4 we describe how emotion recognition data may be incorporated into the system. Finally in Section 5 we come up with the evaluation of the affective e-learning system, while in Section 6 conclusions and future work is presented.

2. The object oriented method

Object-oriented programming can trace its roots to the 1960s, but was not commonly used in mainstream software application development until the early 1990s [15]. Object-oriented programming provided researchers with ways to maintain software quality and to develop object oriented applications in part to address common problems by emphasizing discrete, reusable units of programming logic. An object-oriented program may be considered as a collection of cooperating objects, as opposed to the conventional model, in which a program is seen as a list of tasks (subroutines) to perform. In OOP, each object is capable of receiving messages, processing data, and sending messages to other objects and can be viewed as an independent mechanism with distinct roles or responsibilities.

Object oriented approaches have been already widely used in software development environments [16,17]. An improved object oriented model for Java is presented in [18]. In [19], an object-oriented analysis is adopted for the implementation of remote sensing imagery to GIS. The authors of this paper argue that there is a large gap between theoretically available information and used information to support decision making. As a proposed strategy to bridge this gap, these authors suggest the extension of their signal processing approach for image analysis by exploration of a hierarchical image object network to represent the strongly linked real-world objects.

Additionally to the OO method, the UML [20] approach has been developed to standardize the set of notations used by most well known object oriented methods. In order to support these approaches, CASE tools like Rational Rose [21] and Paradigm Plus [22] have been developed.

According to [23], a number of fundamental concepts are found in the strong majority of definitions of object oriented programming:
Fig. 1. Object model for the Emotion Recognition System.

- **Class**
  Classes define the abstract characteristics of objects, including their characteristics (attributes, fields and properties) and their functions.

- **Objects**
  Objects are patterns of a class.

- **Instances**
  An instance is the actual object created at runtime.

- **Methods**
  Methods illustrate objects’ abilities. In programming languages, methods are referred to as functions.

- **Message passing**
  Message passing represents the general process by which an object sends data to another object or asks the other object to invoke a method.

- **Inheritance**
  Subclasses are more specialized versions of a class, which inherit attributes and behaviours from their parent classes, and can introduce their own.

- **Abstraction**
  Abstraction is simplifying complex reality by modelling classes appropriate to the problem, and working at the most appropriate level of inheritance for a given aspect of the problem.

- **Encapsulation**
  Encapsulation conceals the functional details of a class from objects that send messages to it.

- **Polymorphism**
  Polymorphism allows the programmer to treat derived class members just like their parent class’ members. Polymorphism in object-oriented programming is the ability of objects belonging to different data types to respond to method calls of methods of the same name, each one according to an appropriate type-specific behaviour.

- **Decoupling**
  Decoupling allows for the separation of object interactions from classes and inheritance into distinct layers of abstraction.

### 3. General architecture of the system

The affective multimodal object model can be illustrated as a graphical model where system classes including attributes, services and class relationships are defined. Data that can be considered as relational to the general emotion recognition process are distinguished by two individual categories. The first category comprises of the multimodal data that may be collected by each individual modality. The emotion detection server
Fig. 2. Object model for the construction of the emotional stereotypes.

consists of one or more modalities and each of them are factually the emotion detection server’s properties. Correspondingly, emotional states constitute attributes and methods of each modality.

In each implementation of the system, there is a set of available modalities. The authors have presented in their previous works [24,25], emotion recognition systems that consist of two or three modalities. As it is illustrated in Fig. 1, the emotion detection server may read and exchange information with the available modalities. Each modality may provide the system with information about the recognition of one or more emotional states, with specific degrees of certainty. Additionally, each modality may provide supplementary information concerning the emotional states of users, which is associated with specific user actions during the human-computer interaction. Such actions include correct or wrong browsing, answers in tests in educational software environments, etc.

Stereotypic information is also very important for emotion recognition purposes and a complete study considering their incorporation into the emotion recognition system is shown in [26]. In this paper we suggest an object oriented structure of all the available stereotypic information. As it is illustrated in Fig. 2, the main emotion stereotype class consists of three subclasses. The first class stores and administers data that are associated with users’ characteristics. These characteristics derive from each user’s personality, such as the user’s educational level, the user’s sex and age, etc. and help the system improve its emotion recognition capabilities. The second subclass models stereotypic information concerning user actions during their interaction, while the third subclass represents pre-stored information about each modality’s ability of recognizing each one of the available emotional states.

One of the major contributions of the object oriented architecture of the emotional recognition system is the great easiness in adapting the system to new or more roles that may lead to better emotion recognition capabilities (such as the incorporation of new modalities). Moreover, this architecture provides a framework for emotion recognition systems in different computerized environments and laboratory installations. As it is illustrated in Fig. 3, the resulting emotion detection server can exchange emotional interaction data from multiple sources, such as personal computers, mobile devices, or even integrated laboratory installations with multiple devices as input modalities.

The affective module is incorporated into a sophisticated e-learning environment. The general architecture of the e-learning system consists of the main educational application with the presentation of theory and tests, a programmable human-like animated agent, a monitoring user modeling component and a database
that is structured according to the pre-mentioned object oriented model.

While using the educational application from a desktop computer, users are being taught a particular course (in our example, anatomy). The information is given in text form while at the same time the animated agent reads it out loud using an incorporated speech engine. Users can choose specific parts of the human body and all the available information is retrieved from the systems’ database. In particular, the main application is installed either on a public computer where all users may have access, or alternatively each student may have a copy on his/her own personal computer. An example of using the main application is illustrated in Fig. 4. The animated agent is present in these modes to make the interaction more human-like.
While the users interact with the main educational application and for the needs of emotion recognition, a monitoring component records the actions of users from the keyboard and the microphone. These actions are then processed in conjunction with the multi-criteria model and interpreted in terms of emotions. The basic function of the monitoring component is to capture all the data inserted by the user either orally or by using the keyboard and the mouse of the computer. The data is recorded to the system's database and the results are returned to the basic application the user interacts with. Figure 5 illustrates the "monitoring" component that records the user's input and the exact time of each event.

4. Emotion recognition data into objects

In Section 3 we have shown how the object-oriented method may be used in order to provide a reliable model structuring the information that is used in emotion recognition systems. Furthermore, this approach can be adopted in future emotion recognition systems with more modalities and improved emotion detection mechanisms. In this section we present actual emotion recognition data that are classified according to the aforementioned object-oriented architecture. The aim of this section is to illustrate the variety of different data, in order to indicate the necessity of a well-structured approach which can classify and manipulate them.

4.1. Data from multiple modalities

As is shown in [27], different computer modalities have distinguishable capabilities in recognizing humans' emotional states. For example, some emotions may give significant visual evidence of their existence to any observer, while in other situations aural data may be preferable. Furthermore, in human-computer interaction, specific categories of users may use different modalities for their interaction. In an em-
prical study that we conducted in previous research work we have noticed that a very high percentage (85%) of young people (below 30 years old) who are also inexperienced with computers reported to have preferred expressing themselves through the oral mode rather than the keyboard during their interaction with a computer. On the contrary participants who were computer experts did not give us considerable data for the affect perception during the oral communication with their computer.

Another important issue concerning the combination of data from multiple modalities is the fact that different modalities may give evidence for different emotions or emotional states. A modality may be able to provide a system with information about six discrete emotions, while another modality may only decide whether positive or negative feelings are detected.

4.2. Data from user input actions

The emotion recognition system incorporated a user monitoring component that captured all users’ actions concerning the personal computer’s keyboard and the computer’s microphone. After processing the collected data we also came up with statistical results that associated user input actions with emotional states. More specifically, considering the keyboard we have the following categories of user actions: k1) user types normally k2) user types quickly (speed higher than the usual speed of the particular user) k3) user types slowly (speed lower than the usual speed of the particular user) k4) user uses the “delete” key of his/her personal computer often k5) user presses unrelated keys on the keyboard k6) user does not use the keyboard. These actions were also considered as criteria for the evaluation of emotion with respect to the user’s action in the keyboard.

Considering the users’ basic input actions through the microphone we come up with 7 cases: m1) user speaks using strong language m2) users uses exclamations m3) user speaks with a high voice volume (higher than the average recorded level) m4) user speaks with a low voice volume (low than the average recorded level) m5) user speaks in a normal voice volume m6) user speaks words from a specific list of words showing an emotion m7) user does not say anything. These seven actions were also considered as criteria for the evaluation of emotion with respect to what the user says.

4.3. Stereotypic information

Considering the important problem of which mode gives better results or in what extent should the evidence from each mode should be taken into account, the authors have proposed a novel approach for calculating the weight of significance of each mode based on stereotypes and a multi-criteria theory [26,27]. Stereotype-based reasoning takes an initial impression of the user and uses this to build a user model based on default assumptions [28]. Stereotypes constitute a powerful mechanism for building user models [28]. This is due to the fact that stereotypes represent information that enables the system to make a large number of plausible inferences on the basis of a substantially smaller number of observations [29]. The stereotype inferences are used in combination with a decision theory, namely Simple Additive Weighting (SAW) [30,31] for estimating weight of significance of each mode in the affective reasoning of the system for a particular user.

In previous research work of the authors [26], we have classified our users into stereotypes concerning their age, their educational level, their computer knowledge level and their gender. For each user there is a val-
ue that corresponds to a four-dimensional stereotypic vector of the form:

(User Name, Stereotypic Characteristic 1, Stereotypic Characteristic 2, Stereotypic Characteristic 3, Stereotypic Characteristic 4).

Stereotypic Characteristic 1 refers to the user’s age and is an element of the following set concerning ages: [(10–16), (16–24), (24–36), (36–50), (over 50)]. Stereotypic Characteristic 2 refers to the user’s computer knowledge level and is an element of the following set concerning the user’s computer experience in months (using a personal computer): [(less than 1), (1–6), (6–12), (over 12)]. Similarly, we have defined stereotypic characteristics 3 and 4 that refer to the user’s educational level and to the user’s gender respectively. The inferences of this stereotype vector provide information about the weights of importance of each mode for the users belonging to that stereotype. For example, if a user belongs to the stereotype [(16–24)] is inferred to have a tendency to express his/her feelings through the oral mode of interaction.

Stereotypes can provide inferences concerning hypotheses about users’ feelings and which modality should be more important for providing evidence about users’ feelings. More specifically, in many cases, data from the vocal interaction or the interaction through the keyboard gives evidence of different emotions with quite similar degrees of certainty. For example, the system may have evidence that a user is either angry while saying or typing something or stressed or even confused. The incorporation of stereotypes in the system provides inferences concerning people belonging to the same category with the user that may help in recognizing an emotion that is more common for the users of this category among others or even distinguishing emotions. Evidence for the character or the personality of a particular user may raise the degree of certainty for a particular emotion recognized.

5. Evaluation

For the evaluation of the affective e-learning system, 15 computer programmers, as well as 10 instructors participated. All of the instructors who participated in the experiment were familiar with the use of computers. In addition, they had been trained for the use of the affective e-learning system before the experiment.

When interviewed, all of the instructors confirmed that the resulting system had a user-friendly interface and that the affective facilities were either useful or very useful. More specifically, 9 of them stated that they found the emotional interaction facilities either useful or very useful both for the creation and maintenance of their courses whereas only 1 of them said that s/he would not use the emotional interaction features during the creation and maintenance of an e-learning course. The exact answers of instructors to questions about the sophisticated emotional interaction features of the e-learning system are illustrated in Figs 6 and 7.

As expected, all of the 9 instructors who found useful the emotional interaction features of the e-learning system, argued that the ability to add a modality of interaction with the system (for example a web camera or a microphone) was also very useful especially in situations where users as students had different computer installations with different modalities for interaction.

Additionally, the 15 computer programmers were asked to score the degree of usefulness of the resulting systems’ Object Oriented architecture after having used...
it and examined its structure. As a result of this evaluation study, computer programmers have appreciated the improvements the OO architecture had offered. The OO affective facilities were considered very friendly usable and useful by computer programmers who had previous computing experience with human emotion detection systems. Figure 8 illustrates the computer programmers’ scores, taking their values from 0 indicating that the OO architecture was not useful at all, up to 5 indicating that the resulting structure was found to be very useful to them.

6. Conclusions

In this paper, we described a multimodal emotion recognition system that is structured according to the object oriented method. The system uses the OO approach that combines evidence from multiple modalities of interaction and data from emotion stereotypes and classifies them into well structured objects with their own properties and methods. Advantages of the proposed approach include the well-known conveniences and capabilities of object oriented structures, such as easiness in the system’s maintenance, great extensibility, better communication through different modalities, good cooperation with different object oriented programming languages, easiness in code debugging, as well as code reusability.

The system’s architecture can be adopted in future emotion recognition systems with multiple modalities and improved emotion detection algorithms. Furthermore the resulting emotion detection server is capable of using and handling transmitted information from different sources of human-computer interaction. Independent user interfaces may send wirelessly or wired information about users’ interaction to the emotion detection server and the server can respond with information about possibly recognized emotional states of the users.
As for future work we plan to improve our system and test the system’s structure by exploiting additional modalities of interaction.

References


