Building Corpora of Bodily Expressions of Affect

Marco Pasch, Monica Landoni

Faculty of Informatics, University of Lugano, Lugano, Switzerland. {marco.pasch, monica.landoni}@usi.ch

Abstract

Studies aimed at measuring users' engagement with interactive products mostly rely on collecting self-report data. However, there is an increasing body of research into systems able to assess the affective state of a user from objective data such as body movements. To train such systems we need corpora of motion capture data showing people in emotionally charged situations. Here, we report on our ongoing research into building corpora of movement data. In particular, we focus on two crucial areas in the building process. Firstly, challenges related to recording corpora "in-the-wild", i.e. in ecologically valid situations. Secondly, challenges related to obtaining ground truth labels, i.e. ensuring that individual sequences of movements are labeled correctly with affective states that can be seen in them.

Introduction

Studies aimed at measuring users' engagement with interactive products mostly rely on collecting self-report data 1. However, people can only report on what they are aware of. Self-report also requires that people can give an accurate description of what they think about a product. Children, for example, often have difficulties in reflecting and communicating their thoughts, beliefs, and wishes. Another potential bias in self-report data is that people not necessarily give us truthful answers. They may leave out items they feel embarrassed about or tell lies instead. Often these lies are innocent when people simply tell us what they believe we want to hear instead of their actual opinion.

An alternative to subjective self-report methods are those based on objective measurements. In recent years, the field of affective computing has come up with a number of approaches that provide more objective data. Typically, these approaches detect one or several modes of affective feedback. Intentionally or not, we continuously express affect. This can happen verbally, through the tone of our voice and our choice of words, as well as non-verbally, through our facial expressions, posture, gestures, amongst others. Physiological data like heart rate or skin conductance have also been used to infer the affective state of the user. When we talk to other people we intuitively pick up on cues to their affective states. As social beings this comes natural to us. However, having computers automatically recognize a person's affective state is still one of the big challenges in affective computing.

Our research aims at assessing engagement from bodily expressions of affect, in a way enabling a system to read the body language of users. This could be beneficial in a number of settings. It could provide researchers and practitioners evaluating interactive products with an initial assessment of how a user felt during interaction, which they can use as a starting point for conducting post-interaction interviews. It could also be an important step towards intelligent systems, which can adapt to their users 1. For instance, if a video game is able to detect the user is frustrated, it can adapt the gameplay and ensures to keep the user in the flow zone 3. Same holds for tutoring systems that ensure the learner is in a positive affective state.

The system we are working on is based on the Microsoft Kinect sensor. The Kinect sensor is a camera-based movement sensor, emitting an infrared grid and aggregating the reflected light into a skeletal representation of the body. Using the Kinect SDK we developed our own motion capture system which records the body movements of up to two players, represented in the form of skeletons. It is fairly robust in terms of ability to function in changing (indoor) environments and totally non-intrusive for the user.

To build systems able to assess affective states from body movements we must know which features within a movement carry cues to a person's affective state. To identify these features we need a corpus of movement data, i.e. must record movement data and label it with the according affective state. Existing corpora of affective behavior usually consist of acted data, which means they often show exaggerated movements done by people who were instructed to depict a particular affective state in a lab environment. However, for affective computing to advance it is important to move from acted data to data obtained from people actually feeling the expressed affective states in ecologically valid environments 4.

Apart from recording “in the wild”, we also must ensure that the data is labeled as accurate as possible.

Observing movement and affective behavior

We conducted initial observations of children playing active video games in a primary school in southern Switzerland. Observing children in a school is a perfect scenario for us, as the school is a very natural environment for its students. After establishing a contact, we conducted observations in the setting of the school's after-school program as well as physical education classes.

The after-school program is a playful setting, where children can spend their time doing homework or play games under supervision until their parents come to pick them up. The physical education class is a formal setting and comes with the great advantage that a central learning objective of this class is for children to learn how to deal with positive and negative emotions, which makes it an ideal scenario for our study.

In both settings we introduced a Nintendo Wii console and observed how children play the Wii Sports games, a set of games simulating tennis, bowling, baseball, and boxing. We also videotaped some of the observation sessions of the physical education classes and reviewed the videos with the teacher. This revealed that we can observe three classes of movements: task-related movements, affective expressions, and social behavior.

Task-related movements are all movements that have to do with playing the game. In the context of the Nintendo Wii this means mostly swings of the arms and flicks from the wrists. Obviously, these movements do also carry affective information. We can observe from the way a child executes, for example, the throw of a bowling ball whether he/she is joyous because the last throw was a strike, or whether he/she was just embarrassed by classmates prior to the throw.

Affective expressions are movements that have no utilitarian function and that only serve the communication of affective states. An example for this is a child throwing his/her hands in the air after scoring a strike in bowling. Another example is a child walking spiritlessly away from the console with hanging shoulders after loosing a game. There is obviously a social component present in that the expression of affect has a social function, such as eliciting sympathy when feeling sad. In fact, most affective expressions happen in a social context. When there is no recipient present, we feel less need to communicate our affective states. Our third class of movements is social behavior. Here, movements have no utilitarian function and are also not primarily communications of affect. We see for instance one girl hugging another girl in a gesture of empathy or a boy putting an arm around another boy to express their friendship.

Recording movement data “in-the-wild”

An important requirement for this is ensuring that the data is obtained from a naturalistic setting. This means that the data comes from people expressing real, i.e., actually felt, affect rather than simply portraying affective states. It also means that we capture the data in an environment that is familiar to people from whom we obtain it. Ideally, it also means they are doing things they are familiar with. In essence, we try to reduce a potential novelty effect and a feeling of insecurity due to alien circumstances as much as possible.

We are currently piloting a study aimed at capturing movement data for which we return to the aforementioned school. Apart from the school being a familiar environment, this setting comes with the advantage that the children there are already used to our presence. Similar to our previous observations, we want the children to engage in playing active video games as the Nintendo Wii. From our observations we already know that the Wii games are known to most of the children. So also the use case is familiar for them.

Children will be playing in couples with Wii Sports games of their choice, playing in a competitive fashion and asked a few questions at the end of a session. Here we want to know how they feel during playing so that we can match their subjective experience against direct observations and the teacher's analysis of the whole experience. In addition to recording movements with the Kinect based recognition system we will video tape the session for later analysis.

Having the teacher on board and enthusiastic to take part in the study is invaluable for us. Not only his professional knowledge on student behavior, but also the fact he has known most of the students since they enrolled in the school has already provided us with many insights during our initial observational study. One of the main learning goals of the physical education syllabus is for students to learn to identify and control their emotions. So our study does not interfere with the class' activities, it may actually provide the teacher with new insights and aid achieving the learning goals.

Labeling data

In order to be able to use the motion capture data for training and testing an affect recognition system, we need to label individual sequences with what affective state can be seen in them. In affective computing, this process is known as ground truth labeling.

As the labels are the benchmark that a system is tested against, it is an important step for the creation of affect recognition systems. Usually, labels are obtained from letting observers rate the sequences and for instance by majority vote assign a label to each sequence. Our own labeling procedure will involve two researchers assigning labels and discussing those with the expert (teacher) in order to validate these. We will develop an appropriate coding protocol to be followed by all human judges involved.

However, we argue that this common practice of ground truth labeling can introduce biases. Often judges are recruited from colleagues of the researchers. This can introduce a population bias into the ratings and makes a generalization to how people in general rate expressions of affect difficult.

Because of this we want to explore how emerging methods such as crowdsourcing can help in improving ground truth labeling. Crowdsourcing here means letting a large population of observers rate the sequences online. This way we can include a bigger population from a wider background into the rating process. However, this also means letting go of a controlled environment where we have no control over the circumstances in which the observers rate the sequences. Still, we believe that exploring different rating schemes is an undertaking worthwhile. Ideally we can find more robust rating schemes. For the very least, we hope this can inform researchers in developing future corpora.

Conclusion

We described our ongoing research into building an automatic affect recognition system based on body movements. In particular, we discussed the necessity of corpora of annotated movement data that shows affective behavior of humans in naturalistic settings for training such systems as well as requirements towards building them. Once we have recording and labeled the data, we plan to make it available to other researchers. Existing corpora of affective behavior usually consist of acted data, which means they often show exaggerated movements done by people who were instructed to depict a particular affective state in a lab environment. However, for affective computing to advance it is important to move from acted data to data obtained from people actually feeling the expressed affective states in ecologically valid environments. We believe our corpus can help other researchers to test their systems with more valid data than what is available today.

References

1. Bargas-Avila, J. A., Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI'11)*, ACM Press, pp. 2689–2698.

2. Pantic, M., Nijholt, A., Pentland, A., Huang, T. (2008). Human-Centred Intelligent Human-Computer Interaction (HCI²): how far are we from attaining it? *Int. J. of Autonomous and Adaptive Communicative Systems* **1**(2), 168-187.

3. Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM* **50**(4), 31–34.

4. Picard, R.W. (2011). Measuring Affect in the Wild. *Proceedings of the 4th Int. Conf. on Affective Computing and Intelligent Interaction* (ACII'11), Springer, p. 3.