A data mining driven approach to real-time classroom feedback through multimodal sensing

Weiling Li¹, David Muñoz², Alex Richert³, Conrad S. Tucker²

The Pennsylvania State University

ABSTRACT

The communication of complex concepts is a challenge for teachers, not only in primary and secondary schools but in university classrooms. In order to adapt and maximize the impact of the lectures on the class, it is necessary that teachers receive frequent and reliable feedback from their students. Currently, such feedback is typically gathered through written quizzes or surveys, which cannot accurately measure levels of student engagement in real time over the course of a full lecture. In this paper, we present a novel data mining driven approach for gathering classroom feedback by digitally analyzing student body language. We use a multimodal sensor to precisely record students’ body language in real time, and use classification rule mining to compare the body language data with feedback obtained with written surveys. From the methodology proposed, insights will be gained into the feasibility of using digital body language analysis to accurately assess levels of student engagement.

1. Introduction: the need for feedback

One of the main challenges faced by instructors in the classroom environment is to measure levels of student engagement, as well as to identify disruptive behaviors that could potentially affect performance. Some aspects of student attitude can be measured through evaluations and surveys. However, these approaches are limited, and usually do not allow the teachers to gather information in real time for the duration of an entire lecture. To properly understand and measure levels of engagement, decoding body language becomes crucial. In general, nonverbal communication represents more than 50% in the communication process (Birdwhistell, 1970), and invaluable insights

¹Dept. of Computer Science and Engineering, College of Engineering
²Harold and Inge Marcus Dept. of Industrial and Manufacturing Engineering, College of Engineering
³Dept. of Astronomy & Astrophysics, Eberly College of Science
into mental states can be obtained if properly understood. However, it is difficult for teacher to accurately and quantitatively monitor the body language of each student, especially in large classes. Therefore, new methods for properly assessing student attitude in real time are needed.

In this paper, a novel data mining driven approach is proposed to address the issues of analyzing body language in real time. We observe student body language through multimodal sensing, and use written surveys to quantify levels of student engagement. We then perform classification rule mining on the data obtained to determine which, if any, aspects of body language most strongly predict student responses. If student attitudes can be measured through multimodal sensing, then teachers will potentially have the tools needed to identify those concepts that their students especially struggle with and to evaluate the effectiveness of different teaching styles, as well as identify individual students who are particularly disinterested or struggling with the material. In principle, the applications of such an approach in the classroom setting alone are diverse, and could potentially aid in not only identifying levels of student engagement, but various aspects of academic performance, disruptive behaviors (student chatter, cell phone usage), and academic dishonesty (such as cheating during exams).

In §2, a review of relevant literature in body language, technology in the classroom and use of data mining in education areas is presented. In §3, the methodology proposed is described including considerations for the experimental setup, collection of data, parameterization of data and the classification algorithms used to derive the rules for predicting student feedback. Finally in §4, initial results based on preliminary experiments are shown.

## 2. Literature review

### 2.1. Body language and nonverbal communication

It was long believed that words were the key drivers of communication. This idea started to disappear in 1872 when Charles Darwin published “The Expression of the Emotions in Man and Animals,” which suggested that mammals consistently reveal their emotional states through their facial expressions. However, it was not until the early 1950s when the term “nonverbal communication” became popular and widely studied (Ruesch and Kees, 1956). Nowadays, many types of nonverbal communication are well understood, with gestures and body language being the areas which have been studied the most (Davis 1973; Key 1975; Knapp 1980; Leathers 2001; Knapp et al. 2009; Wharton 2009; Maslow et al. 2011).

Body language plays a crucial role in the communication process. It may provide cues to detect various aspects of an individual’s mental state. According to Birdwhistell (1970), words represent only 7% of the communication process, while paralinguistic cues represent about 38%
and nonverbal communication represents 55%. As stated by Peter Drucker, considered the father of management theory, “The most important thing in communication is to hear what isn’t being said.” This reflects the idea that the primary driver in communication is not words, but non-verbal cues.

Understanding nonverbal communication is important in many fields. Some areas in which it has been studied the most are business (Cooper 1979, Chaney and Martin 1995, Bovée et al. 2003), politics (Mehrabian 1971, Geis 1987) and education (Banbury and Hebert, 1992). In addition, according to Watzlawick et al. (1967), the first axiom of communication states that “one cannot not communicate” and therefore every behavior forms an integral part of the communication process. As a consequence, to maximize student understanding in the classroom, teachers must know how to effectively decode the signals sent by their students through posture and movement.

2.2. Body language in the classroom setting

Non-verbal communication, especially body language, can potentially play an important role in the classroom environment. Teachers who effectively communicate through their own body language may be able to positively impact student performance. According to Chaudhry and Arif (2012), there is a strong association between nonverbal behavior of the teachers and student achievements. The importance of nonverbal communication in classroom management is studied in Zeki (2009), which concludes that eye contact, mimicry, and gestures on the part of the instructor are crucial to managing the classroom and promoting student engagement. As a consequence, a friendlier, motivating, and comfortable environment can be promoted by properly using the codes of nonverbal communication. According to Hogan and Stubbs (2008), unclear or misleading body language is one of the main barriers to effective communication. Bass et al. (2011) confirm the importance of body language in the professor-student interaction. Many other studies have been conducted to help teachers to better communicate by using their body language and other non-verbal cues (Minskoff 1980; Kellerman 1992; Caswell 1993; Corts and Pollio 1999; Miller 2005; Minskoff 2008; Hood 2011; Tabensky 2012).

Although research has been proposed for providing ways for teachers to better communicate using body language, only a few studies have focused on providing methods for understanding student body language. One of the main barriers in identifying students’ nonverbal cues is the difficulty of properly paying attention to large groups. According to Sparks (2010). The average class size in the U.S. is 25 students. Therefore, interpreting each student’s facial expressions, posture, and movement patterns is almost impossible. Thus, more sophisticated mechanisms must be used in the future to overcome this challenge.
2.3. Technology in the classroom

Since the last decade, the use of technology in the classroom to improve student performance has increased (Demuelle et al. 1998, Bitter and Legacy 2002, Bitner and Bitner 2002). Debavec et al. (2006) examine students’ use of technology for learning and compare their performance against more traditional learning methods. It was determined that to improve achievement, an interaction between traditional methods and technology-oriented methods is needed. A comparison of an access grid technology versus traditional videoconferencing in classroom and research environments was conducted by Suarez (2007). According to a survey presented in Frey and Birnbaum (2002), most students agreed that computer-assisted lessons had a positive impact on their learning.

Although it has been proved that the use of technology can effectively improve student performance (Schacter 1999, Schacter and Fagnano 1999, Harter and Harter 2004, Clotfelter et al. 2008), not many technology approaches have focused on collecting and providing timely lecture feedback. Gligoric et al (2012) is one of the few researches discussing how to use technology to provide real-time feedback to teachers. Through the use of multimodal sensors, feedback is provided to achieve an end-to-end communication. As a consequence, the teacher can rapidly adapt their presentation to the audience to achieve a maximum understanding impact. According to their research, by using a multimodal sensor device, valuable insights and inferences about audience attitude and lecture quality can be made.

2.4. Data mining techniques employed in the classroom setting

Data mining refers to the extraction of information from a large data set and deriving new knowledge through the identification of patterns, either through unsupervised learning (clustering) or supervised learning (classification; Witten et al, 2011). Data mining in educational systems is an iterative cycle of gathering and analyzing educational information from students, educators, and academic administrators, then transforming data into knowledge, and analyzing this knowledge to aid in decision making (Romero & Ventura, 2007). Data in this context comes from databases of student/instructor information, class and schedule information, and information from distance learning courses (student access records, etc.), and the information generated can be used to identify various problems and make policy decisions that will improve learning outcomes.

To extract important factors from the vast amounts of educational data, feature selection techniques help to reduce the computation time and enhance the predictive accuracy of the model (Witten, 2011). Feature selection techniques are divided into two categories. The first one is the filter method, which ranks feature according to some univariate metric (e.g., Fisher scores, information gain) and selects the highest ranking features. The other one is the wrapper method, which
searches for the “best” subset of features based on the classification accuracy (e.g., feature subset search). Ramaswami & Bhaskaran (2010) make use of the filtered feature selection technique to select the best subset of variables on the basis of the values of $\chi^2$ measure (Witten, 2011) that measures divergence from the distribution expected if one assumes the feature occurrence and the class value are assumed independent.

The main task of data mining driven educational systems is to predict student performance and identify those students who may be at risk of poor performance. Pandey and Pal (2011) predict new student performance by means of Bayesian classification, based on analyzing past data such as language skills and background qualifications. Shirwaikar and Rajadhyax (2012) employ a decision tree to classify student performance into different categories so that academic help can be provided to at-risk students. They also identify clusters of students in order to help to improve the performance of the students who are weak in particular subjects. Parack et al. (2012) perform association mining with the Apriori algorithm to profile students based on academic records, such as exam scores, term work grades, attendance and practical exams. They also use $k$-means clustering to identify patterns in student performance.

Visualization tools are also provided to instructors so that the results of mining can be better understood in a visual manner. Concept mapping is a visual representation of the structure of knowledge about a particular topic developed by Professor Joseph D. Novak in the 1960s. Chiu and Lin (2012) analyze concept mapping data by the sequential pattern analysis method. It has been found that the mapping sequences produced by higher-score students are similar, while ones produced by low-score students appear very diverse. Social networking-based methods have been developed to study individuals and their relationships (Scott, 2000). Fire et al. (2012) visualize the social interactions among the students in a course. The result shows a high correlation between the grade of a student and that of his closest friend in the social network.

2.5. Multimodal sensors and data mining

Capturing and decoding body language is a challenge due to the dynamic nature of these raw data. New technologies such as multimodal (three-dimensional) sensors have helped researchers to capture and analyze body language effectively. Patsadu et al. (2012) used the Microsoft Kinect to recognize three categories of motion: standing up, sitting down, and lying down. According to their study, a 100% rate of accuracy was obtained when using neural networks trained through backpropagation. For the other three classification methods used (support vector machines, decision trees, and naive Bayes classifiers), the average accuracy was 93.72%. Raptis et al. (2011) used the Kinect to identify dance movements based on sixteen joints. The accuracy obtained was 96.9% based on 4-second recordings. Preis et al. (2012) investigated gait recognition using three different
classifiers: 1R, C4.5 and naive Bayes. Out of the three, the naive Bayes classifier obtained the best results when using all features (including the coordinates of the 20 joints and some biometric features such as length of legs, torso, etc.); a success rate of 91% was achieved. Popa et al. (2012) classified shopping-related behaviors, such as browsing and seeking assistance, using distribution moments—variance, skewness, kurtosis—among image pixels occupied by customers. In Stone and Skubic (2011), a set of temporal and special gait parameters were proposed as a way to detect the early stages of illnesses and functional declinations in an in-home risk assessment monitoring system. Then, parameters such as walking speed and body rotation were estimated. As can be seen, synergies can be obtained from the combination of multimodal sensors and data mining concepts. However, a detailed parameterization of the raw data obtained is required to ensure a proper characterization of body language and movement.

3. Methodology

In the following, we describe our methodology for the collection and analysis of data. In §3.1, we discuss our approach for collecting data through experiments involving real students. In §3.2, we address the technical aspects of the Kinect device and interpretation of raw multimodal data. Finally, in §3.4, we outline the classification algorithms to be compared for predicting student attitudes based on body language.

3.1. Experimental setup

In order to provide teachers with computational tools that will allow them to measure levels of student engagement based on body language, we must derive classification rules that connect observed body language to metrics of engagement that are obtained separately. To this end, we conduct experiments wherein student subjects are placed in a realistic classroom setting and presented with several short lectures. Student body posture is recorded during each lecture using the Kinect, and student engagement is measured separately through written surveys. Survey responses serve as the class variables used for the classification algorithms discussed in §3.4.

For identifying experimental subjects, students are selected from diverse academic and demographic backgrounds in order to account for any aspects of body language which may vary across culture or other demographic categories. Subjects are placed in a classroom setting where a single Kinect device is set up covertly in order to avoid anomalous behavior stemming from the students’ awareness of being recorded. To further enhance realism and de-emphasize the experimental nature of the situation, we use deception by informing students that, rather than the students themselves
being the target of the study, the goal of the study is to compare the different teaching methods of the lecturers (students are given these instructions from a script read out by the administrator of the experiment). This suggests to the students that they will find some of the lectures presented more interesting than others, and allows them to act more as they would in an ordinary classroom setting. As a final step in creating a nominal classroom environment, 10-15 extra students are seated in the room to produce a typical class size.

In addition to identifying students from diverse backgrounds, we wish to, as much as possible, observe each student at a range of engagement levels over the course of the experiment. Observing each student in their maximally and minimally engaged states, for instance, makes the digital body language data maximally informative and aids our classification algorithms in clearly distinguishing each type of behavior (classification of body language data cannot be performed if, for example, all students report being disinterested in each and every lecture). To ensure such a range of student attitudes, the three lectures presented span a range of STEM subjects, as well as a range of intended levels of difficulty and interestingness.

The final component of the classroom procedures is the student survey administered after each lecture. In the case of Bidwell & Fuchs (2011), the attention levels of elementary-level students were evaluated by experts in child behavior who were present during each session. We use written surveys in order to reduce subjective judgments regarding apparent student attitudes, and also to provide quantitative feedback which readily lends itself to evaluation through classification algorithms. The National Survey of Student Engagement (NSSE) is a widely used, well validated survey of engagement for undergraduates\(^1\). The survey, however, focuses on student engagement over the course of several semesters, and does not immediately lend itself to evaluating student engagement for a single class, much less a single lecture. Other student surveys, such as CLASSE (Ouimet & Smallwood, 1995), are derived from the NSSE, and are appropriate for measuring levels of student engagement for a single, semester-long course. For the current research, we derive a new survey from the NSSE and CLASSE that is appropriate for measuring student engagement over the course of a single short lecture. Like the NSSE and CLASSE, our survey uses the Likert scale for student responses—multiple choice answers in the form of, “good, neutral, bad,” etc., typically with four or five options per question. Our survey contains questions about the format of the lecture, the overall performance of the instructor, and the difficulty and interestingness of the material itself. This allows us to determine in general the aspects of a lecture to which students are the most responsive; for instance, it is possible that student body language reveals student attitudes toward the instructor and his or her teaching style rather than toward the material itself. In addition to the administration of this survey after each lecture, student subjects complete a general

\(^1\)http://nsse.iub.edu/
questionnaire based on the academic and demographic questions found on the NSSE. This gives us the opportunity to further study the causes and significance of student body language and attitudes (for instance, students with different academic and demographic backgrounds may demonstrate interest or disinterest in different ways).

3.2. **Kinect technology to capture body language**

For collection of body language data, we use the *Microsoft Kinect*, a low cost, commercially available multimodal sensor system originally developed for interactive video games. This device is composed of an RGB camera that provides two-dimensional position information, and an infrared emitter and camera which provide depth information. Through these sensors, body motion can be tracked in real-time. The *Kinect* is capable of capturing the *xyz*-coordinates for each of twenty tracking point distributed across the whole body, however since students in our classroom experiment are seated, only twelve points in the upper body are recorded (shoulders, elbows, hands, wrists, base of the neck, and head).

Capturing and decoding body language is by nature dynamic; being able to sequentially track skeletal movement is therefore critical for measuring student behavior over the course of each lecture. The *Kinect* is capable of capturing coordinate data in real time at 30 frames per second, or once every 33 milliseconds. This rapid recording of body motion allows us to calculate the instantaneous velocities of each tracking point.

3.3. **Parametrization of raw multimodal data**

The data recorded with the *Kinect* device are the *xyz*-coordinates for each tracking point at each time step. In order to turn this raw data into information capable of revealing student attitudes, we implement several parametrization schemes to generate the data used by each classification algorithm. At each time step, angles at each joint are calculated (say, wrist-elbow-shoulder), as well as higher distribution moments for all twelve tracking points (variance, skewness, kurtosis). The instantaneous velocities of each tracking point are calculated, as well as the dispersion of each of these velocities over the course of each lecture, indicating overall level of change in body posture during the lesson (a lot of motion might, for instance, indicate boredom and impatience). In addition, we calculate ratios among all point-to-point distances. We have chosen to calculate an extensive array of parameters to account for possible subtle distinctions in student body language; for instance, it may be discovered that the posture of a student who is engaged somewhat resembles the posture of a student who is disinterested, and we therefore wish to use all available information,
such as tracking point velocities, to account for such distinctions as much as possible.

3.4. Classification algorithms

In the current research, we use classification algorithms to determine the relationship between body language and levels of student engagement. The values of the predicting attributes are derived from observed student body language as observed with multimodal sensors, and are compared with the quantitative student feedback gathered with a written survey which includes questions regarding interaction level with the instructor, interest level, etc. As more than 500 features are generated after parametrization of multimodal data, feature selection is an essential step to select the most relevant features for classification. We use the wrapper-based feature selection to search for the most informative subset of features. In the experiment we apply naive Bayes, SVM, Logistic regression, C4.5, CART, AdaBoost, and Random Forest with 5-fold cross-validation using Weka software (Waikato Environment for Knowledge Analysis)\(^2\). We select the classifier with the highest score of F-measure.

4. Results

In order to explore the suitability of the Kinect for digitally observing body language, we recorded test data and applied various parametrizations. A single subject was seated at a desk and assumed several poses corresponding with stereotypical student postures: neutral, bored/sleepy, confused, and raised hand (shown in Fig. 1). We processed the test data by taking the median for each joint value for every three frames; this removes the jitter frequently found in Kinect data, and removes non-values due to device errors and out-of-frame tracking points.

We tested several parametrization schemes based on our sample data. First, we calculate the higher moments of the distributions of points along each axis for each frame (variance, skewness, kurtosis), similar to the parametrization scheme validated in customer behavior analysis by Popa et al. (2011). We also calculated several angles for triplets of consecutive joints, providing a more detailed representation of student posture (in principle, the students exact body position can be derived from the complete ensemble of joint angle information). We selected one representative frame from each of the four test postures and calculated the above quantities. The variance, skewness, and kurtosis data for each axis for each test posture are given in Table 1, along with the angle between the abdomen, neck, and head tracking points. We find that skewness and kurtosis along

\(^2\)http://www.cs.waikato.ac.nz/ml/weka/
Fig. 1.— Four test postures, recorded with the *Kinect*: neutral (upper left), raised hand (lower left), bored/sleepy (upper right), and confused/head to the side (lower right).
all three axes strongly distinguish between each posture.

It is important to note the differences between the case study presented here and the complete methodology that we propose. In the current case study, we have only used a small ensemble of parameters. Also, we have not examined the time-domain aspect, i.e. we have looked only at single frames and have not calculated joint velocities. Both of these considerations will be incorporated into our future work, and we will also use student surveys as the basis of body language assessment, rather than using pre-assigned, sterotypical student poses.

5. Conclusions & future work

There is a widespread need among university-level instructors for accurate, real-time feedback so that they can effectively teach difficult topics and manage large classrooms. Given the recent rise of commercially available imaging and computing power, there is a great deal of promise in devices such as the Microsoft Kinect in collecting data on student body language and behavior which can then be analyzed in real time on a personal computer. We have demonstrated the capability of the Kinect to observe student body language, and determined that there are several parametrizations that can be used to effectively distinguish several important postures. Further results on the connection between body language and student attitude may eventually provide teachers with the tools for measuring the individual levels of engagement for each student, as well as the class as a whole. This and forthcoming research on student behavior will establish a new role for technology in the classroom, as well as ultimately serve to provide university-level instructors with tools that will allow them to effectively plan and deliver their lectures.
REFERENCES


This preprint was prepared with the AAS LaTeX macros v5.2.
Table 1. Comparison of test parametrizations for four postures: neutral, confused, bored, and raised hand. \( V, S, \) and \( K \) are the computed variance, skewness, and kurtosis along each axis for each posture; \( \angle ANH \) is the calculated angle between the \emph{Kinect} tracking points corresponding with the abdomen, neck, and head.

<table>
<thead>
<tr>
<th>Posture</th>
<th>( V_x )</th>
<th>( V_y )</th>
<th>( V_z )</th>
<th>( S_x )</th>
<th>( S_y )</th>
<th>( S_z )</th>
<th>( K_x )</th>
<th>( K_y )</th>
<th>( K_z )</th>
<th>( \angle ANH )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neut.</td>
<td>0.027</td>
<td>0.021</td>
<td>0.023</td>
<td>0.255</td>
<td>1.061</td>
<td>-0.782</td>
<td>-1.370</td>
<td>0.081</td>
<td>-0.905</td>
<td>176°</td>
</tr>
<tr>
<td>Conf.</td>
<td>0.026</td>
<td>0.010</td>
<td>0.025</td>
<td>0.306</td>
<td>1.671</td>
<td>-0.672</td>
<td>-1.245</td>
<td>2.030</td>
<td>-0.907</td>
<td>149°</td>
</tr>
<tr>
<td>Bored</td>
<td>0.010</td>
<td>0.006</td>
<td>0.007</td>
<td>-0.092</td>
<td>1.595</td>
<td>0.137</td>
<td>-0.803</td>
<td>1.722</td>
<td>-1.433</td>
<td>120°</td>
</tr>
<tr>
<td>Hand</td>
<td>0.023</td>
<td>0.026</td>
<td>0.027</td>
<td>0.472</td>
<td>0.282</td>
<td>-0.465</td>
<td>-1.092</td>
<td>-1.340</td>
<td>-1.444</td>
<td>170°</td>
</tr>
</tbody>
</table>