Early Predicting of Student Struggles Using Body Language

Mr. Matthew L Dering, Penn State University

Matthew Dering is a PhD student at Penn State University studying computer vision and deep learning.

Dr. Conrad Tucker, Pennsylvania State University, University Park

Dr. Tucker holds a joint appointment as Assistant Professor in Engineering Design and Industrial Engineering at The Pennsylvania State University. He is also affiliate faculty in Computer Science and Engineering. He teaches Introduction to Engineering Design (EDSGN 100) at the undergraduate level and developed and taught a graduate-level course titled Data Mining–Driven Design (EDSGN 561). As part of the Engineering Design Program's "Summers by Design" (SBD) program, Dr. Tucker supervises students from Penn State during the summer semester in a two-week engineering design program at the École Centrale de Nantes in Nantes, France.

Dr. Tucker is the director of the Design Analysis Technology Advancement (D.A.T.A) Laboratory. His research interests are in formalizing system design processes under the paradigm of knowledge discovery, optimization, data mining, and informatics. His research interests include applications in complex systems design and operation, product portfolio/family design, and sustainable system design optimization in the areas of engineering education, energy generation systems, consumer electronics, environment, and national security.

Early Predicting of Student Struggles Using Body Language

Abstract

The accuracy of RGB-D sensing has enabled many technical achievements in applications such as gamification, task recognition, as well as pedagogical applications. The ability of these sensors to track many body parts simultaneously has introduced a new data modality for analysis. By analyzing body language, this work can predict if a student will struggle in the future, and if an instructor should intervene. To accomplish this, a study is performed to determine how early (after how many seconds) does it become possible to determine if a student will struggle. A simple neural network is proposed which is used to jointly classify body language and predict task performance. By modeling the input as both instances and sequences, a peak F Score of 0.459 was obtained, after observing a student for just two seconds. Finally, an unsupervised method yielded a model which could determine if a student would struggle after just 1 second with 59.9% accuracy.

1 Introduction

In this work, the role of machine learning for planning student intervention is investigated. Specifically, this work asks two questions: (i) Can a student's struggles be predicted based on body language? (ii) How soon can these struggles be predicted? Identifying early signs of student struggles with a task has many benefits. Often, teachers are in charge of many students, so predicting future struggles would allow them to direct their attention where it is most needed.

In advanced settings, students are occasionally asked to perform tasks that carry risks (e.g., chemicals in scientific labs, machinery and equipment in engineering labs, etc.). Students struggling may inadvertently injure themselves or others, and this may not always be obvious to the instructor, given that instructors are often in charge of dozens of students and cannot focus on all simultaneously. Studies have shown that students who struggle for too long become more difficult to teach, correct, and can encounter difficulties when gaining competency[2,25].

Off the shelf depth-enabled sensing devices (RGB-D) such as the Microsoft Kinect have enabled the automated tracking of several people in a frame at once. By leveraging this technology as well as machine learning, a model is created which can predict if an instructor should intervene with a student before the student begins to struggle.

Predicting if a student might struggle with a task is an ambiguous goal. For example, what defines struggling (e.g., bottom half of participants, slower than a target determined by an expert, etc.)? An additional concern is how quickly should a prediction be made. If a student shows signs of struggle early, intervention may prevent them from learning on their own.

Similarly, a student who struggles for too long may grow frustrated and become resistant to intervention. This work considers several values for each of these variables, in order to assess the proper values for planning intervention.

In short, this work aims to predict a binary variable indicating if the student ultimately struggles with the task. Inside this problem are several subproblems. First, how should this binary variable be set? Second, at what point is the highest accuracy achieved in making this prediction? Finally, is this accuracy best achieved by a system which views student activity as a sequence or as independent observations?

This work is organized as follows: Section 2 discusses the related work. Section 3 and Section 4 detail the experimental method used. Section 5 discusses the findings and Section 6 outlines future work.

2 Related Work

2.1 Human Data Analysis

New technology in human sensing has lead to an explosion in work in activity recognition, anticipation, and other related fields. One major motivation for this interest is in facilitating a better working relationship between people and robots. Koppula and Saxena created a model which allowed a robot to anticipate what a human might do next, and was able to move to assist[13]. Mainprice and Berenson suggested a similar idea which allowed robots to plan manipulation of objects with a human[17].

Humans as a subject of machine learning is a very active area of research. Some straightforward human tendencies have been effectively modeled with machine learning such as predicting if a human will continue to walk in the same direction[21]. Other advancements have enabled the study of human affect and emotion as it pertains to tasks. The major focus of this work has been in using computer vision to ascertain emotion (such as happy or frustrated), and studying performance. Many approaches to ascertaining affective states have been proposed[6,22]. For example, the CERT toolbox allows real time estimates of emotional state[16].

These approaches mostly leverage 2D (video) data, and hence are limited by that medium. The advent of depth enabled sensors has allowed real time pose tracking for humans, which has been used for body language studies[11,28]. This has been used to quantify emotional states during lectures[5], provide feedback to public speakers[19], and to study human robot interaction[18]. These previous works each demonstrate that enough information is contained in human pose and body language to predict a broader factor, such as emotion or satisfaction with an interaction. This method proposed in this work takes this one step further and investigates at what point can a computer automatically predict struggles.

2.2 Educational Machine Learning

Machine learning in educational settings considers many data modalities, to help students and instructors accomplish their educational goals[24]. Education settings generate many different types of data, including text and video. A particularly active area of research is the massive online open course (MOOC)[12]. These are courses with large numbers of people enrolled, which take place via a computer. One of the features of MOOCs which has been extensively studied is the textual data of students discussing the course in forums. Text data is highly structured and thus lends itself well to being mined for sentiment[26], though the utility of this is not always clear[27]. Other work leverages student interaction with a computer tutor to serve as an automated assessment tool[23]. Modeling student progress as a distribution of skills has enabled educators to determine underlying patterns which indicate the abilities needed to succeed in a subject matter[15].

Predicting student performance is an active area of research in the educational community. A highly accurate prediction has many benefits, including early intervention, and appropriate allocation of instructor resources. Predicting student performance given student data is a two pronged problem. Measuring student performance is constantly shifting, even as new models for mining student data are being proposed. Baker and Inventado give an overview in their review[4]. Many methods rely on student surveys[8], grades[10], or propensity to drop out[7]. At the same time, many advances have been made in both input and data modeling. Baker et al. detected students affect based solely on their interaction with an online tutor[3]. San Pedro et al. proposed a similar system for determining student affect[20].

The use of commodity off the shelf computer hardware such as depth enabled sensors has seen some use in educational settings. For instance, Alexiadis et al. proposed an application of the Kinect for assessing dance performances[1]. Kyan et al. further studied the use of this technology for ballet training[14]. D'Mello et al. used Kinects for capturing student-teacher interaction[9]. These works demonstrate the utility of low cost depth enabled sensors for easily capturing physical tasks, as well as human dynamics. The method proposed in this work builds on that by determining if student body language can be used to predict student performance in physical tasks, and determining at what point can an optimal determination be made.

3 Method

This method will determine if there are early body language predictors of poor task performance. For example, consider a student in an engineering laboratory, attempting to perform a task. This method supposes an environment where the student is being monitored by a RGB-D sensor (such as a Microsoft Kinect). This sensor and attached system is tasked with monitoring the student for signs that they may struggle, so that an instructor can aid them. Since it is an engineering laboratory, they may be working with dangerous equipment, further motivating this work. This method will determine in detail how much time is needed before an accurate prediction of student struggle can be made.

3.1 Data Description

The primary data modality of this work is joint coordinates of the human body. The advent of RGB-D sensors has also brought extremely accurate joint locations sensed extremely quickly. For a student in a laboratory environment, these sensors can track the student as they move through space, and provide joint data to the system, the sum of which is the students overall body language and pose. This data is defined by some number of joints J (this can vary by software implementation or hardware limitations), each of which has several pieces of data associated with it. In particular, three coordinates in 3D space, two coordinates in 2D space,

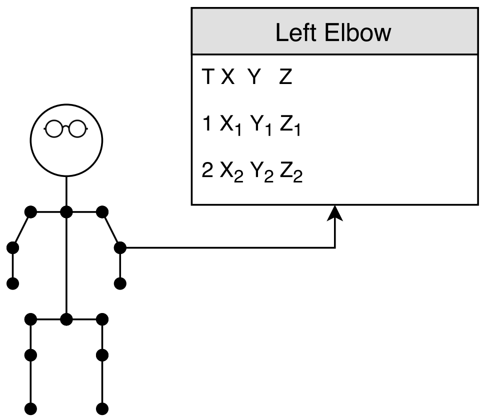


Figure 1: Body data of a student gathered performing an action

and a quaternion of values which give its orientation with respect to the parent joint (i.e. the angle). A visual description of this can be seen in Figure 1. The target data is the variable which this model will predict. In this case, student performance will be measured by how long it takes them to perform a task.

This model considers this input in two formats: individual input (i.e. snapshots in time), and sequential input (i.e. short clips of time). These two inputs will clarify whether it is more predictive to consider sequenced information or as independent variables. It is also important to determine not only if low task performance can be predicted, but at what point can this performance be predicted most accurately. Therefore, different time periods are considered as the threshold for determining failure (in one second increments, first second, first two seconds, and so forth). This addresses the trade-off between waiting too long to intervene and making a highly accurate decision to intervene.

3.2 Model Definition

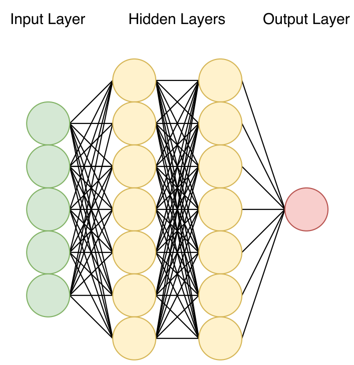


Figure 2: A neural Network with two hidden layers.

Two primary types of models are used in this work. The first is a multilayer perceptron, similar to the generic network seen in Figure 2. This is a type of neural network which consists of several layers of fully connected neurons, interspersed with a non-linearity. Briefly, a neural network consists of *neurons*, connected to one another in *layers*. These neurons are simple mathematical machines which perform a mathematical operation given by Equation 1:

$$y = f(w \times x + b) \tag{1}$$

where y is the output, x is the input, f is a (typically non-linear) function, w is a learned weight, and b is a learned bias. By combining matrix multiplication and non-linear functions, neural networks have proven very effective at studying more complex categorization problems, such as those discussed in this work. This class of networks has been shown to perform extremely well at multi-dimensional classification tasks, such as the one described in this work.

The other type of model employed in this work is a recurrent neural network. These networks have been extremely successful in modeling temporal data, that is, data which is gathered over time. One of the most common forms of recurrent networks is the Long Short Term Memory (LSTM), which allows networks to remember and eventually forget past input, using a concept called a forget gate. LSTMs are commonly used for sequence classification of multi-dimensional data, for tasks such as sentiment classification of textual data. In a way, this work also classifies sequences by sentiment, where the sentiment is determining if the student will struggle and needs attention.

3.3 Model Testing

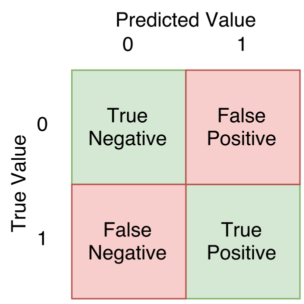


Figure 3: Confusion matrix showing the two types of error.

The models trained for this purpose will be tested for precision (Equation 2) and recall (Equation 3).

$$\text{precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{3}$$

where TP is the number of true positives (that is, correctly identified cases of students struggling), FP is the number of false positives (those incorrectly identified as struggling), and FN is the number of false negatives (those incorrectly identified as not struggling). An illustration of this is shown in Figure 3. For example, in this case, a false positive would result in an intervention where one was not needed, and a false negative would result in no intervention where one was needed. The F-Score metric is simply the harmonic mean of precision and recall.

4 Case Study

This case study considers students who were asked to perform three tasks in an engineering design setting. These students were drawn from an introductory engineering design course.

Figure 4: A student performing one of the tasks in this work.

They were tasked with hammering two nails and a screw into a block of wood. The purpose of the screw task was to induce the student to struggle (as hammering a screw into wood is difficult). Having a task designed to be struggled during can be helpful when exploring how to best define the concept of struggling (e.g. the threshold discussed above). An image of a student working on one of the tasks is shown in Figure 4.

Before beginning, the raw X, Y, Z data (taken in meters), is normalized to the distance between shoulder and opposite hip, as well as centered about the head. This helps account for the variation in subject size and location in the frame. For example, this will avoid any confusion about students who are on the right side of the space vs those on the left. Additionally, this makes the model robust to changes in subject height, so that a tall subject and a short subject with similar body language are compared fairly.

5 Results

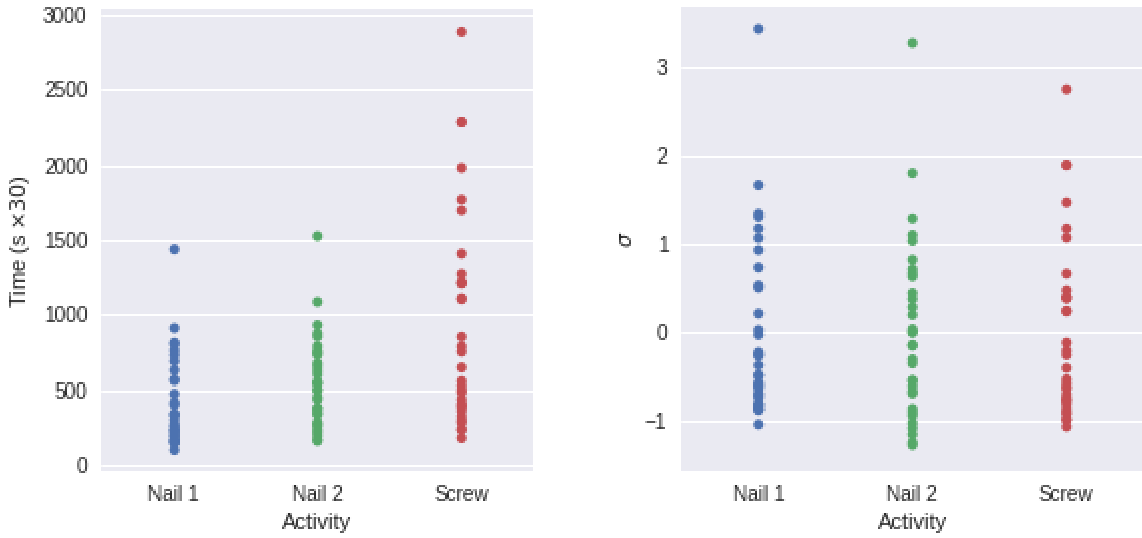
5.1 Thresholds

Before determining the predictive ability of body language for task struggle, the identification of struggle must be first recognized in the data. Figure 5(a) shows the breakdown of each of the three activities and how long it took for individual students (note that the values are counted in frames of a 30 FPS video). Normalizing these distributions for standard deviation and mean, can be seen in Figure 5(b).

From this plot, it is clear that student performance varied very little in the first two tasks. There are also a few clear outliers in these tasks. However, the third task, hammering a screw into a board, has a much wider distribution. Finding the appropriate threshold is a critical part of this task. First, students who took more than 1.644 (95% coverage on a normal distribution) times the standard deviation are considered to be struggling. In practice, this highlighted 7 students who struggled, two each on the nail tasks and three on the screw task. One of the students who struggled with Nail 2 also struggled with Screw, otherwise the other 5 instances are 5 distinct students.

5.2 Time Cutoffs

One of the points of interest for this work is at what point does the system have enough information to make an accurate decision. To begin, the observations (of both position and orientation), are considered as independent observations. While it seems logical that sequential modeling



(a) Time taken by each student, per activity

(b) Unit norm distribution of activity time

Figure 5: Time distribution of activity data.

	Individual			Recurrent		
	Precision	Recall	F Score	Precision	Recall	F Score
Seconds						
1	0.199	0.367	0.258	0.080	0.133	0.100
2	0.161	0.222	0.186	0.300	0.545	0.387
3	0.190	0.211	0.200	0.092	0.176	0.121
4	0.101	0.142	0.118	0.628	0.391	0.482
5	0.049	0.087	0.062	0.359	0.264	0.305
6	0.000	0.000	0.000	0.351	0.257	0.297
7	0.167	0.213	0.187	0.219	0.260	0.238

Table 1: Results for 95th percentile struggle threshold.

would yield more information, this has a few advantages. For one, it is faster, allowing for this to run on a portable system. It also allows for a cascading model to be employed, which can analyze these independent predictions to make a final more accurate determination.

In order to analyze this time cutoff, several values were proposed. The first n observations for each student performing an activity are used to train this model, where n is a variable corresponding to the amount of time needed. The results of this can be seen in Table 1. While there is a general trend of increasing accuracy over time, these results are not accurate enough to make confident decisions.

5.3 Sequential Analysis

The sequential relationship of human data is clear. At time t, the location of a persons joints, and their configuration, is very similar to time $t - 1$. To determine if this could better predict future struggles, a similar experiment is performed. To perform this analysis, the recurrent network is presented with body data in the form of 10 frame increments, every 5 frames. For example, frames 0–10, 5–15, and so on. The results of this can be seen in Table 1. These results represent a substantial increase over the independent observation based method. This is encouraging as it signifies that the system is learning short periods of body language which predict struggle. Still, these models do not perform well enough to be actionable. This is most likely due to difficulties training for such a particular dataset. Future work may consider augmenting this dataset with additional users, of varying experience levels.

5.4 Struggle Threshold Analysis

One problem common to each of these results is the heavy class imbalance among training data. The threshold was chosen empirically based on apparent distributions in the data, and while these distributions seemed clear, the threshold selected left just seven instances of struggle. After dividing these into training and test sets, it becomes very difficult to train and validate a model. This can have a substantial effect on model training, and while techniques exist to combat these class imbalances (for instance, weighted penalties), they are not generally intended to handle cases of single digit positive samples. This section explores some methods of overcoming these limitations.

First, rather than choosing struggle threshold using standard deviation, instances were chosen by hand, to include more from the third task, hammering a screw. Since this task is designed to induce struggle, it makes sense to consider more from this class, as its mean and standard deviation will be larger on the whole. Put another way, the cutoff considered by the previous two sections only considered subjects who were struggling at a task which was designed to cause struggle.

Including the next three highest struggling users, the results are shown in Table 2. These show a marked improvement in performance of predicting if a student struggles. Notably, the independent experiments were much more successful, and the recurrent experiment improved for nearly every time quanta. The best performance achieved an F Score of 0.459, which indicates that overall this performs nearly as well as a coin flip, and far outperforms random chance.

Overall the model performance may be best explained by the peculiarities of human subjects. For example, many students begin a task in a neutral position, regardless of whether they will struggle with the task or not (increasing the likelihood of a false negative). Some students may show signs of struggle early but recover and complete the task without ultimately struggling, quantitatively (increasing the likelihood of a false positive). Even the differences between two students who both struggle may be clear, as one student begins at second 2, while the other does not until second 4. This latter instance especially merits further study, as finding similarities between a new student and a past student can aid in providing customized feedback.

5.5 Unsupervised Approach

One of the interesting aspects of this method is that it allows for exploration of the data in an unsupervised fashion. Unsupervised learning has the advantage of not requiring hand-coding of the student performances. To this end, a neural network is used to perform feature extraction, and the resulting features are used in a clustering method. This method automatically

	Individual			Recurrent		
	Precision	Recall	F Score	Precision	Recall	F Score
Seconds						
1	0.085	0.113	0.097	0.294	0.600	0.395
2	0.173	0.130	0.149	0.418	0.509	0.459
3	0.323	0.382	0.350	0.303	0.388	0.340
4	0.355	0.408	0.380	0.368	0.548	0.441
5	0.337	0.361	0.349	0.352	0.434	0.389
6	0.323	0.361	0.341	0.346	0.417	0.378
7	0.346	0.355	0.350	0.373	0.444	0.405

Table 2: Results after relaxing the criteria for struggling students

decomposes the data into clusters, some of which indicate that the student will struggle. The broad impact of these "struggle" clusters, not only allows for prediction, but also for individual personalized feedback.

To examine this effect, the data is partitioned into clusters. Next, clusters which only contain non-struggling data are discarded, and the contents of the remaining clusters are examined. Of these clusters, after just 1 second, 59.9% of remaining datapoints are associated with a student who will eventually struggle. This means that given a point which falls into one of these clusters, there is a 59.9% likelihood that that student will struggle.

Clustering is of particular utility in analyzing students as it gracefully handles many of the problems associated with educational data mining. For one, students are a diverse group, with many different tendencies and behaviors, and clustering techniques allow for modeling of a population as a mixture of sub-populations. The other benefit is their lack of reliance on target data. As discussed, the evaluation of student performance is variable and prone to subjectivity. This can prevent a coherent model from developing in a supervised learning context. Since clustering establishes relationships within the data, this can have utility in other context as well, such as finding students that are similar to one another. These unsupervised methods may even be useful in finding similarities between students who appeared not to struggle and students who did, to create customized feedback.

6 Conclusion

This work has shown the ability of neural networks to process complex data-types such as human poses. Two network designs were proposed, one viewing samples as independent points in time and the other considering short sequences of pose information. This found that sequential modeling was more accurate than the modeling using an independence assumption. By relaxing strict assumptions about the definition of struggling, the results showed a marked improvement, yielding an F Score of 0.459. An unsupervised method yielded a model which could predict if a student would struggle with nearly a 60% accuracy after just 1 second.

There are many avenues for future work extending this study. Decomposing the human data into small windows proved successful in predicting future failure; however, this made a strong assumption which may hinder the overall performance of the model. Rather than assuming that the user was "doomed" to struggle from the start, it would be very helpful to examine the data to determine if there was a point at which the student began to struggle. Making such a determination would enhance the results by reducing the amount of noise, since the data currently suffers from noise due to encoding that the subject would eventually struggle even if the subject was not struggling yet. While the recurrent network can be somewhat more robust to these issues, further work may investigate if an automatic approach to finding such a "tipping point" were possible. Finally, a larger sampling combined with an unsupervised method may be able to deduce larger trends not just in body language but in determining new methods of student assessment.

References

[1] D. S. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, and M. B. Moussa. *Evaluating a dancer's performance using kinect-based skeleton tracking*. ACM, New York, New York, USA, Nov. 2011.

[2] R. L. Allington. Intervention all day long: New hope for struggling readers. *Voices from the Middle*, 2007.

[3] R. Baker, S. Gowda, and M. Wixon. Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. ... *Data Mining 2012*, 2012.

[4] R. S. Baker and P. S. Inventado. Educational Data Mining and Learning Analytics. In *Learning Analytics*, pages 61–75. Springer New York, New York, NY, 2014.

[5] I. Behoora and C. S. Tucker. Quantifying Emotional States Based on Body Language Data Using Non Invasive Sensors. *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages V01AT02A079–V01AT02A079, Aug. 2014.

[6] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan. *Analysis of emotion recognition using facial expressions, speech and multimodal information*. ACM, New York, New York, USA, Oct. 2004.

[7] G. Dekker and M. Pechenizkiy. Predicting students drop out: A case study. ... *Data Mining 2009*, 2009.

[8] S. K. D'Mello, S. D. Craig, A. Witherspoon, B. McDaniel, and A. Graesser. Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1-2):45–80, 2008.

[9] S. K. D'Mello, A. M. Olney, N. Blanchard, B. Samei, X. Sun, B. Ward, and S. Kelly. *Multimodal Capture of Teacher-Student Interactions for Automated Dialogic Analysis in Live Classrooms*. ACM, New York, New York, USA, Nov. 2015.

[10] M. Feng, N. Heffernan, and K. Koedinger. Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, 19(3):243–266, 2009.

[11] J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced Computer Vision With Microsoft Kinect Sensor: A Review. *Cybernetics, IEEE Transactions on*, 43(5):1318–1334, 2013.

[12] Ł. Kidziński, M. Giannakos, D. G. Sampson, and P. Dillenbourg. A Tutorial on Machine Learning in Educational Science. In *State-of-the-Art and Future Directions of Smart Learning*, pages 453–459. Springer, Singapore, Singapore, 2016.

[13] H. S. Koppula and A. Saxena. Anticipating human activities for reactive robotic response. In *International Conference on Intelligent Robots and Systems*, pages 2071–2071. IEEE, 2013.

[14] M. Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, N. Dong, B. Elder, and L. Guan. An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(2):23–37, May 2015.

[15] R. V. Lindsey, M. Khajah, and M. C. Mozer. Automatic Discovery of Cognitive Skills to Improve the Prediction of Student Learning. *NIPS*, pages 1386–1394, 2014.

[16] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (CERT). In *Gesture Recognition (FG 2011)*, pages 298–305. IEEE, 2011.

[17] J. Mainprice and D. Berenson. Human-robot collaborative manipulation planning using early prediction of human motion. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2013)*, pages 299–306. IEEE, 2013.

[18] D. McColl and G. Nejat. Affect detection from body language during social HRI. In *2012 RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 1013–1018. IEEE, 2012.

[19] A.-T. Nguyen, W. Chen, and M. Rauterberg. Online feedback system for public speakers. In *2012 IEEE Symposium on E-Learning, E-Management and E-Services (IS3e)*, pages 1–5. IEEE, 2012.

[20] M. O. Z. S. Pedro, R. S. J. d. Baker, S. M. Gowda, and N. T. Heffernan. Towards an Understanding of Affect and Knowledge from Student Interaction with an Intelligent Tutoring System. In *Artificial Intelligence in Education*, pages 41–50. Springer, Berlin, Heidelberg, Berlin, Heidelberg, July 2013.

[21] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th International Conference on Computer Vision (ICCV)*, pages 261–268. IEEE, 2009.

[22] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.

[23] A. N. Rafferty, J. L. Davenport, and E. Brunskill. Estimating Student Knowledge from Paired Interaction Data. *Educational Data Mining*, 2013.

[24] C. Romero and S. Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135–146, July 2007.

[25] J. K. Torgesen and R. F. Hudson. Reading fluency: Critical issues for struggling readers. *What research has to say about fluency ...*, 2006.

[26] C. S. Tucker, B. Dickens, and A. Divinsky. Knowledge Discovery of Student Sentiments in MOOCs and Their Impact on Course Performance. *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages V003T04A028–V003T04A028, Aug. 2014.

[27] M. Wen, D. Yang, and C. Rose. Sentiment Analysis in MOOC Discussion Forums: What does it tell us? *Educational Data Mining*, 2014.

[28] Z. Zhang. Microsoft Kinect Sensor and Its Effect. *MultiMedia, IEEE*, 19(2):4–10, 2012.