Modeling Involuntary Dynamic Behaviors to Support Intelligent Tutoring (Student Abstract)

Mononito Goswami,[1,2*†] **Lujie Chen,**[2 †] **Chufan Gao,**[2] **Artur Dubrawski**[2]
[1]Delhi Technological University, New Delhi, India
[2]Auton Lab, Carnegie Mellon University, Pittsburgh, USA
mononito_bt2k16@dtu.ac.in, {lujiec, chufang}@andrew.cmu.edu, awd@cs.cmu.edu

Abstract

Problem solving is one of the most important 21st century skills. However, effectively coaching young students in problem solving is challenging because teachers must continuously monitor their cognitive and affective states and make real-time pedagogical interventions to maximize students' learning outcomes. It is an even more challenging task in social environments with limited human coaching resources. To lessen the cognitive load on a teacher and enable affect-sensitive intelligent tutoring, many researchers have investigated automated cognitive and affective detection methods. However, most of the studies use culturally-sensitive indices of affect that are prone to social editing such as facial expressions, and only few studies have explored involuntary dynamic behavioral signals such as gross body movements. In addition, most current methods rely on expensive labelled data from trained annotators for supervised learning. In this paper, we explore a semi-supervised learning framework that can learn low-dimensional representations of involuntary dynamic behavioral signals (mainly gross-body movements) from a modest number of short time series segments. Experiments on a real-world dataset reveal a significant utility of these representations in discriminating cognitive disequilibrium and flow and demonstrate their potential in transferring learned models to previously unseen subjects.

1 Introduction

For young children, solving challenging non-routine math problems emulates real life challenges they will encounter later in their lives, and may often invite them to ride an "emotional roller-coaster" as the child advances through various stages of problem solving (Chen et al. 2016). Thus, effectively coaching young students for problem solving requires teachers to continuously monitor their cognitive and affective states and make real time pedagogical decisions such as when and how to best intervene. Moreover, teachers have to effectively handle the high cognitive load of monitoring diverse cohorts of students. Intelligent Tutoring Systems that attempt to teach problem-solving also face similar challenges. To lessen the cognitive load of teachers and improve the effectiveness of intelligent tutoring, we envision a decision support system which can monitor the cognitive and affective states of multiple students simultaneously in real time. The focus of this paper is on the state detection capability of such a system, specifically needed to discriminate between *cognitive disequilibrium* (CD) and *flow* states, which are the critical inputs to inform appropriate subsequent interventions. In this work, we investigate a method designed to discriminate between CD and flow using involuntary behavioral signals that are less prone to social editing, including head and eye movement, which can be non-invasively collected using inexpensive sensors such as cameras. To overcome limited supply of labeled data, while taking advantage of the large supply of unlabeled data, we explore a semi-supervised approach where deep embedding features are derived from unlabeled time series segments, which are then fed into a supervised learning algorithm.

2 Methodology

Our study is based on a dataset collected in one-to-one coaching scenarios for math problem solving. We extracted a number of features from the dataset along the visual (*Facial Action Units (FAUs), head & eye gaze orientations*) and writing channels (*writing speed*). We computed the first and second order derivatives of all visual features with the exception of FAUs.

The field of affective and cognitive computing relies on supervised learning algorithms, and is therefore heavily dependent on training data from expert annotators or self-reports by participants of a study. Since most advanced and powerful supervised learning algorithms require substantial amounts of training data to learn reliable decision functions, application of affective computing is severely limited by short supply of trained expert annotators or potentially biased self-reports. To this end, we investigated the utility of an unsupervised representation learning model proposed by (Franceschi, Dieuleveut, and Jaggi 2019), which can be trained on a large amount of unlabeled data to learn potentially useful feature representations. By automatically learning useful features for classifying raw data, representation learning algorithms replace manual feature engineering and

[*]Mobile Number: +91 8800592994

[†]The authors wish that it be known that, in their opinion, the first two authors should be regarded as joint First Authors.

allow systems to identify potential discriminators and use them to support a specific predictive task. Their model comprises of a deep neural network with dilated causal convolutions to handle time series and minimizes an unsupervised triplet loss function which assigns similar time series proximate embeddings based on the assumption that they occur in temporal proximity while a randomly chosen subseries is likely to be dissimilar.

The unsupervised representation learning model was trained on 248 time series segments each having 27 features over 300 time steps and returned 64-dimensional embeddings. Using these output embeddings as feature vectors and manually annotated labels, we trained a random forest classifier to predict the cognitive state (Flow or Cognitive Disequilibrium) of a time series segment. We chose random forests because they are able to learn non-linear and complex decision boundaries, work well with high-dimensional data and can be robust to outliers.

The unsupervised representation learning model coupled with a random forest classifier can function as a semi-supervised model, where the former learns embeddings (features) from a large number of time series segments in a completely unsupervised fashion, and the latter uses these features and a limited number of annotations to learn a decision function. Such a semi-supervised paradigm can be extremely useful in practice of affective computing, where obtaining vast amounts of unlabeled data is extremely easy, but its annotation can be expensive.

3 Results and Discussion

Predictive Utility of Deep Features

We evaluated the predictive utility of deep embedding features by feeding them into a random forest classifier. We conducted two types of experiments: (1) "Random" experiments which make a *random split* between train and test sets. These experiments could yield inflated algorithm performance as the information from the same session and subject may appear in both the training and testing sets, allowing the model to succeed by hooking-onto personal characteristics of some distinct subjects. (2) Leave-one-person(subject)-out experiments which represent a "cold start" scenario where the model is trying to predict for a completely unseen subject. Due to varying degrees of information sharing between training and test set, we expect the performance will degrade from the upper bound case of random split to the conservative (but of high practical utility) LOPO experiments. We also compared the performance of our semi-supervised model with *ResNet*. (Fawaz et al. 2019) in a recent and comprehensive survey found that ResNet can significantly outperform other deep learning approaches in classifying time series on the UCR/UEA and MTS archives. In addition, they found encouraging results (comparable predictive performance and significantly less training & testing time) while comparing ResNet to other state-of-the-art time series classification algorithms such as HIVE-COTE.

	Semi-supervised		Supervised (ResNet)	
Experiments	Random	LOPO	Random	LOPO
Precision	0.83 (0.037)	0.81 (0.063)	0.81 (0.016)	0.77 (0.074)
Recall	0.82 (0.04)	0.7 (0.111)	0.81 (0.023)	0.78 (0.071)
F1	0.82 (0.04)	0.71 (0.092)	0.81 (0.024)	0.77 (0.069)
Accuracy	0.82 (0.04)	0.7 (0.111)	0.81 (0.022)	0.78 (0.072)
AUC	0.83 (0.052)	0.79 (0.058)	0.8 (0.016)	0.73 (0.081)

Table 1: Performance of semi-supervised model vs. ResNet.

Table 1 compares ResNet and the deep semi-supervised model introduced above on several performance metrics. The results reveal that while ResNet achieved higher accuracy (0.78%) in the LOPO experiments, there was no significant difference (within 95% confidence interval) between the models in terms of AUC in both evaluation scenarios.

How Much Supervision is Necessary?

We conducted sensitivity analysis to demonstrate the utility of unsupervised embeddings in the prediction task. In these set of experiments, we fixed the held-out test set, varied the size of the training set and reported the performance of our semi-supervised approach accordingly. Our experiments (refer Supplementary Material for details) revealed that with deep embeddings features, the model is able to learn effective discrimination with even a small number of labeled data points, and the resulting performance is comparable with a potent fully supervised deep learning alternative which often requires large extents of supervision.

Discussion

In this paper, we explored a semi-supervised framework to model the dynamics of involuntary behavioral signals in order to discriminate between cognitive disequilibrium and flow. Experimental results with a modestly sized multimodal multi-sensor dataset, collected from young children practicing problem solving in a naturalistic environment, reveal several insights. We found that our semi supervised approach was able to effectively generalize from training subjects to previously unseen subjects, as demonstrated by its robust performance with leave-one-person-out experiments. When further validated with a more diverse set of subjects, the proposed approach has the promise to scale up practicality of the task of cognitive and affective state detection that is often bottle-necked by high costs of label acquisition even with abundant unlabeled data. Practically relevant capability of generalization to unseen subjects is also encouraging as the proposed approach would often be expected to work well with out-of-sample subjects in the real world use cases.

References

Chen, L.; Li, X.; Xia, Z.; Song, Z.; Morency, L.-P.; and Dubrawski, A. 2016. Riding an emotional roller-coaster: A multimodal study of young child's math problem solving activities. *International Educational Data Mining Society*.

Fawaz, H. I.; Forestier, G.; Weber, J.; Idoumghar, L.; and Muller, P.-A. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33(4):917–963.

Franceschi, J.-Y.; Dieuleveut, A.; and Jaggi, M. 2019. Unsupervised scalable representation learning for multivariate time series. *arXiv preprint arXiv:1901.10738*.