

Representing Time Series Data in Intelligent Training Systems*

Shengnan Hu¹, Zerong Xi¹, Greg McGowin¹, Gita Sukthankar¹, Stephen M. Fiore¹, and Kevin Oden²

University of Central Florida¹ Lockheed Martin²
Orlando, FL US

Abstract

Many of the most popular intelligent training systems, including driving and flight simulators, generate user time series data. This paper presents a comparison of representation options for two different student modeling problems: 1) early failure prediction and 2) classifying student activities. Data for this analysis was gathered from pilots executing simple tasks in a virtual reality flight simulator. We demonstrate that our proposed embedding which uses a combination of dynamic time warping (DTW) and multidimensional scaling (MDS) is valuable for both student modeling tasks. However, Euclidean distance + MDS was found to be a superior embedding for predicting student failure, since DTW can obscure important agility differences between successful and unsuccessful pilots.

Introduction

The advent of cheap commercial virtual reality headsets has increased the feasibility of training motion tasks such as vehicle control in simulation. Although it is possible to utilize many data sources including gaze tracking, physiological response, and knowledge assessment questions, user control data remains an important predictor of task performance. When applying machine learning to student modeling problems, the question arises of how best to represent the data. In raw form, the data usually comprises a multi-dimensional time series of either position or control inputs. However, it can be preprocessed and vectorized in a variety of ways.

This paper presents a new technique for representing time series data that combines dynamic time scaling (DTW) and multidimensional scaling (MDS) to create a distance embedding for modeling the student performance of flight tasks in simulation. Our research questions are:

- **RQ1:** is it valuable to preserve the temporal ordering of the data in a time series or is it better to vectorize the data using the distribution features?
- **RQ2:** does the same representation perform well across multiple user modeling tasks?

We hypothesize that disparate student modeling problems are not likely to benefit from the same distance embedding;

hence the data preprocessing pipeline must be configured for the research problem.

Data was gathered from 23 subjects in the Prepar3D flight simulator, which was designed to deliver immersive, experiential learning for both professional and academic pilot skills training. During a two hour period, novice subjects were trained to perform simple flight tasks in Prepar3D. Each flight task tested their ability to achieve a target direction, airspeed, and altitude while monitoring the correct instruments. Students were graded as successful if they were able to achieve and maintain the target direction, airspeed, and/or altitude. A time series was created from the user flight control data including the aileron (percentage), elevator (percentage) and throttle (percentage).

Figure 1 shows our proposed embedding method. First dynamic time warping (DTW) is applied to the time series to create a distance matrix between all the flight controller time series. Multidimensional scaling is then applied to embed the data as points in an abstract Cartesian space while preserving abstract distance relations. This output is then used for our student modeling problems: early failure prediction and classifying student activity.

Related Work

There is a rich literature on domain specific processing methods for time series data that includes financial analysis techniques such as autoregressive modeling, sequence learning for natural language processing, and sequence pattern mining for bioinformatics (Esling and Agon 2012). Our flight training problem is somewhat akin to sketch recognition in that we are aligning spatial time series data. However unlike sketch recognition, there is more variance in our time series, and we lack a large master dataset of well executed flight examples. Esling and Agon divide time series methods according to representation techniques, distance measures, and indexing methods. Warping to improve temporal alignment has been shown to be valuable across many domains. There are many variants of it including non-uniform warping to handle local variations, fastDTW (Salvador and Chan 2007), and softDTW. However, rather than using DTW directly, we create an embedding from the DTW output using multidimensional scaling.

*This research was supported by Lockheed Martin.
Copyright © 2021 by the authors. All rights reserved.

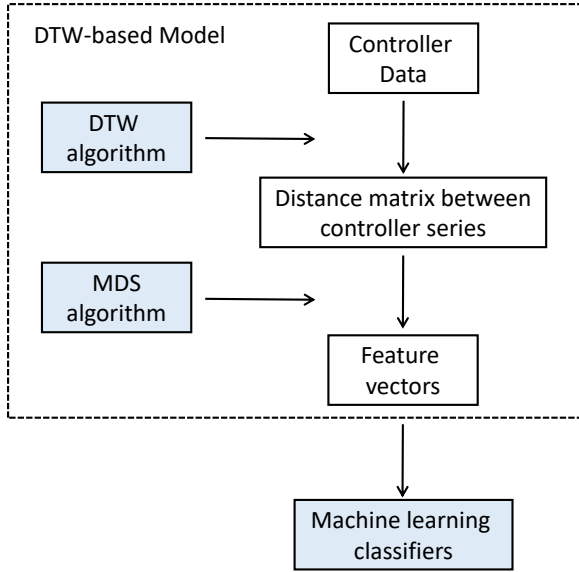


Figure 1: DTW-MDS embedding technique

Method

Our aim is to perform two student modeling tasks: 1) predict whether students will succeed or fail at a flight task and 2) identify the flight maneuver that the pilot is executing. Both of these are useful in the context of an adaptive training simulator.

First we create a distance matrix between the flight controller time series using Dynamic Time Warping (DTW). DTW is a time series analysis algorithm for measuring similarity between temporal sequences. It calculates an optimal match of minimal accumulated distance by warping the timeline while preserving the chronological order. In its simplest form, given two sequences s and t and a distance measurement $d(\cdot, \cdot)$, a table $\mathcal{D}[i, j]$, where i is the index from s sequence and j is that from t , is incrementally updated:

$$\mathcal{D}[i, j] = d(s[i], t[j]) + \min(\mathcal{D}[i, j-1], \mathcal{D}[i-1, j], \mathcal{D}[i-1, j-1]).$$

Given a specific flying task, generally the pilots will execute flight control commands in a similar order but not at an identical pace. Aligning the sequences with DTW removes these pace differences. For our experiments, we use fastDTW (Salvador and Chan 2007), an improved variant of DTW in terms of space and time.

Our final embedding is created using multidimensional scaling (MDS) (Kruskal 1964) which embeds data as points in an abstract Cartesian space while preserving pairwise distances. This has the additional benefit of dimensionality reduction. While DTW can be interpreted as a distance measurement on an infinite dimensional space, MDS is applied here to translate the obtained pairwise distance matrix into the continuous features that are preferred by machine learning algorithms.

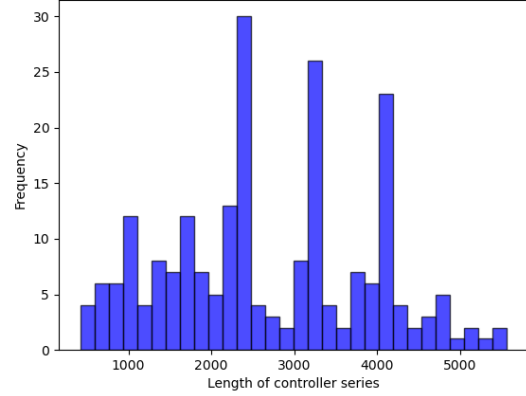


Figure 2: Distribution of controller series length

Results

This section presents the experimental results from applying our proposed DTW-MDS embedding technique to two student modeling problems: 1) predicting task failure and 2) identifying the flight maneuver. We also present an ablation analysis where we compare the performance of different vectorization methods, distance functions, dimensionality reduction techniques, and machine learning classifiers.

Both user modeling tasks can be represented as classification problems. For our experiments, we tested all the standard classifiers that have been shown to work well with small datasets: 1) Support Vector Machines (SVM), 2) Random Forest Classifiers (RF), 3) KNN, 4) Logistic Regression, 5) Gaussian Naive Bayes, and 6) Linear Discriminant Analysis. Stratified five-fold cross-validation process was applied, and the average accuracy score over five folds is reported.

The dataset includes time series from students executing nine different flight tasks in which they were asked to bring the aircraft to a target direction, altitude, airspeed, or combination of two, within a limited time. The dataset is balanced across maneuvers and reasonably well balanced between successful and failed student performances. Figure 2 shows the distribution of time series lengths across the data. Note that it is not simply possible to predict either student failure or task type by the length of the time series alone.

Early Student Failure Prediction

To predict student failure, we analyzed data from the first half of the time period and attempted to predict whether they would succeed at maneuvering the aircraft to the desired position by the time limit.

First we experimented with omitting the temporal information and simply using the descriptive statistics to represent the time series. The data was vectorized using the mean, median, standard deviation, skew, and kurtosis from the aileron, elevator and throttle percentages (Xi et al. 2020).

Table 1 compares this method to our proposed DTW-MDS method. Our proposed method leverages the time series information effectively and is just as successful at pre-

dicting student failure with half the sequence as if it had the full sequence. Neither using the raw time series nor DTW alone performs well as a vectorization strategy (results not shown). Figure 3 shows the same trend across multiple classifiers. SVM is the best performer on the early failure prediction problem which is unsurprising since it often does very well on binary classification tasks.

Models	Controller length	
	Full sequence	Half sequence
SVM (Distribution Statistics)	0.77	0.62
RF (Distribution Statistics)	0.79	0.80
DTW+MDS SVM (Time Series)	0.926	0.926
DTW+MDS RF (Time Series)	0.878	0.890

Table 1: Comparison of vectorization strategies

Models	Training strategies	
	Training/Testing with all examples	Training/Testing with successful only
SVM	0.649	0.875
KNN	0.582	0.762
Logistic Reg	0.618	0.850
Gaussian NB	0.609	0.863
Random Forest	0.676	0.906
LDA	0.622	0.875

Table 2: Performance of DTW-MDS applied to flight maneuver identification

Identifying Flight Maneuver

Since we only need to identify the flight maneuver after the training simulation is complete, these experiments were conducted with the full control data sequence. Table 2 shows the performance of the proposed DTW-MDS embedding across different classification models. We conduct two separate evaluations. In the first we trained and tested models with all the controller data, both from students who succeeded and failed at the maneuver. However since including the poor examples weakens the quality of the flight maneuver dataset, we conduct another training/testing process with only the successful flight examples. We believe that this condition approximates the performance we would achieve with more expert pilots. The Random Forest model achieves the highest classification accuracy, since it is more resilient at handling data with varying scales.

Distance Metric Comparison

To further validate the effectiveness of our embedding strategy, we compare the usage of DTW vs. Euclidean distance as an input to MDS. Specifically, given two controller series S and T , we compute the Euclidean distance between them as:

$$d_e = \sqrt{(s_1 - t_1)^2 + (s_2 - t_2)^2 + \dots + (s_n - t_n)^2},$$

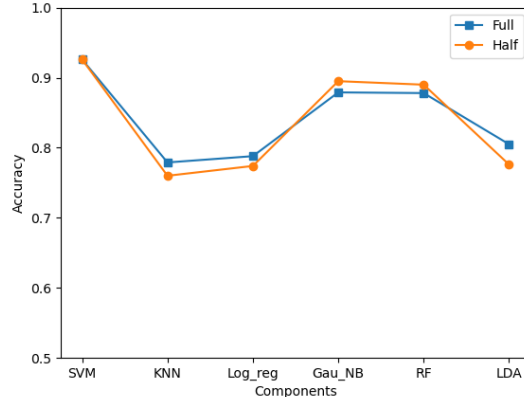


Figure 3: DTW-MDS embedding across different classifiers

where n is the length of series S and T . Note that the length of the two series may vary widely. To tackle this problem and align the sequences for comparison, we select the shorter sequence and pad it with zeros until the lengths of the two sequences are equal. After obtaining the distance matrix using this procedure, we pass it through MDS to create the embedding.

Figure 4 shows the comparison between the use of DTW vs. Euclidean distance for both user modeling tasks. For both training scenarios and all the machine learning models, DTW outperforms the Euclidean distance method at flight task prediction. We believe that DTW is very good at extracting the shape of the different flight trajectories for each of the maneuvers and is more robust to variations across pilots.

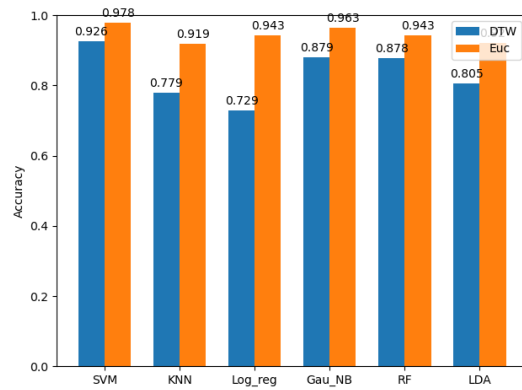
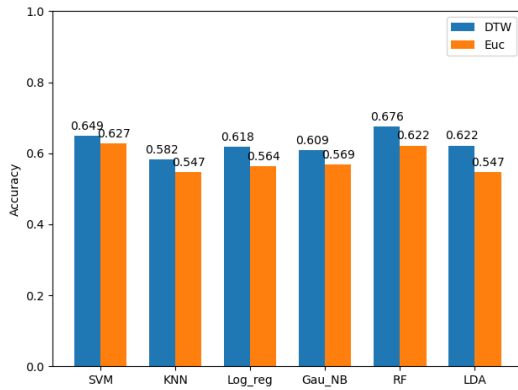
However for failure prediction we observe that the DTW embedding lags behind the use of Euclidean distance. This verifies our hypothesis that separate embeddings should be used for different modeling problems. DTW artificially obscures the pacing differences between the successful and unsuccessful pilots. Previous work on time series (Esling and Agon 2012) has shown that despite its shortcomings Euclidean distance is remarkably successful at many time series problems.

Comparison of MDS with PCA

Here we compare the second part of our embedding technique (MDS) vs. another commonly used dimensionality reduction technique (PCA). Dimensionality reduction not only helps preserve discriminative information, but also filters out noisy features. For these experiments, we used the Random Forest Classifier since it was consistently a top performer.

The comparison of MDS vs. PCA is shown in Figure 5. The MDS algorithm slightly outperforms the PCA algorithm at maneuver identification (as shown in Figure 5 (a)), while the performance of both two algorithms on failure prediction is comparable (as shown in Figure 5 (b)). This validates our choice of MDS as an embedding technique.

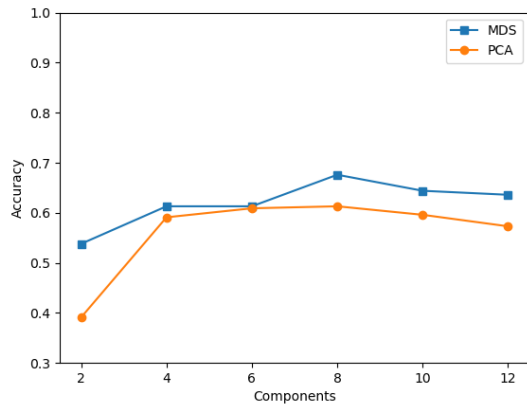
We also evaluated the performance of both algorithms



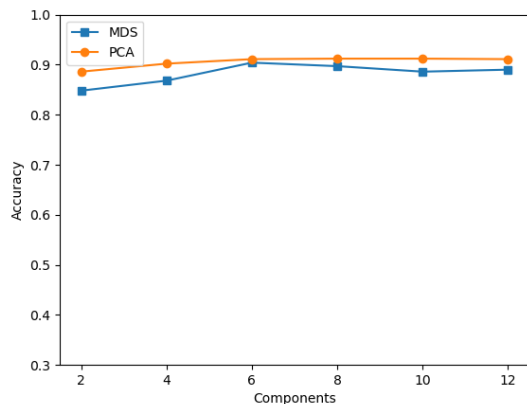
(a) Maneuver identification

(b) Failure prediction

Figure 4: DTW-MDS vs. Euclidean-MDS



(a) Maneuver identification



(b) Failure prediction

Figure 5: DTW-MDS vs. DTW-PCA

with different levels of dimensionality reduction. The best classification performance occurs with about eight components, but failure prediction appears to be less sensitive to this hyperparameter than maneuver identification.

Conclusion

This paper introduces an embedding technique, DTW-MDS, for representing time series data in intelligent training systems. We demonstrate its utility at two different pilot training problems in a virtual reality flight simulator: 1) early failure prediction and 2) maneuver identification. DTW-MDS enables us to leverage temporal information from the flight controller time series more effectively than vectorizing it using distribution statistics (**RQ1**). DTW-MDS performs fairly well across all conditions (**RQ2**) but it is outperformed by an alternate embedding (Euclidean-MDS) at failure prediction.

This verifies our belief that disparate student modeling problems are not likely to benefit from the same distance embedding; hence the data preprocessing pipeline must be configured for the specific research problem. In future work, we are continuing to expand our training system to include additional user modeling tasks such as problem diagnosis and new embedding techniques for knowledge assessment questions.

References

- Esling, P., and Agon, C. 2012. Time-series data mining. *ACM Comput. Surv.* 45(1).
- Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*.
- Salvador, S., and Chan, P. 2007. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* 11(5):561–580.
- Xi, Z.; Newton, O.; McGowin, G.; Sukthankar, G.; Fiore, S.; and Oden, K. 2020. Predicting student flight performance with multimodal features. In *SBP-BRIMS*, 277–287. Springer.