Analysis of Torso Movement for Signing Avatar using Deep Learning

Shatabdi Choudhury
June 24, 2022
The American Sign Language Avatar Project, DePaul University

Importance of Torso Movement in signing

Analyzing and modeling the torso motion is crucial to accurately make the avatar mimic human movements
2014: Torso movements are critical for direct linguistic communication, McDonald et al.

The current quality of signing avatars is not satisfactory for producing human-like user experiences, making them less acceptable to the Deaf community.
2016: A web application for geolocalized signs in synthesized Swiss german sign language. In International Conference on Computers Helping People with Special Needs, pages 438–445. Springer, Ebling, S et al.

Capturing supporting motion is a complicated process

The linguistic annotation does not include supporting torso motions.

2016: Collecting and analysing a motion-capture corpus of french sign language. In 10th LREC Workshop on the Representation and Processing of Sign Languages: Corpus Mining, ELRA (Benchiheub, M., Berret, B., and Braffort, A)

2015: Avatar-independent scripting for real-time gesture animation. Kennaway, R.

Goal

Coordinate torso movements with other body parts to produce lifelike natural postures

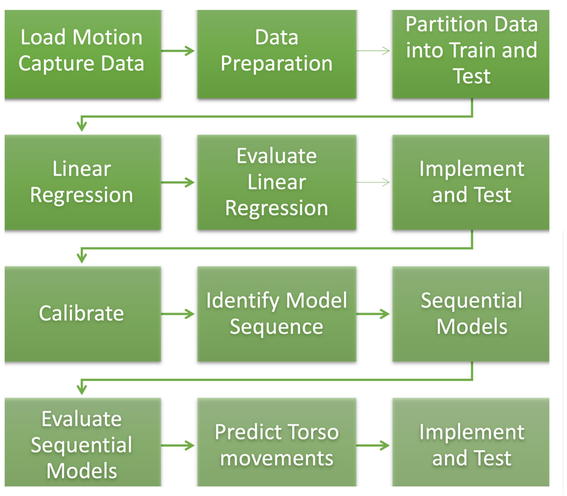
Experiment Details

The data were recorded with a motion capture system with sensors along the spine, neck, head, shoulders, elbow, and wrist orientations from four signers.
25 frames per second.

3DS Max software package was used to import the mocap data and convert it to match the avatar's coordinate system, resolve data issues, such as outliers, and derive new variables required for the ML model
the data to CSV files.

Tools used: Python, Scikit-learn, TensorFlow, Keras, 3DS Max

Machine Learning Models: Linear Regression, Chained Regression, Deep Learning. Linear Regression is used as the companion model.

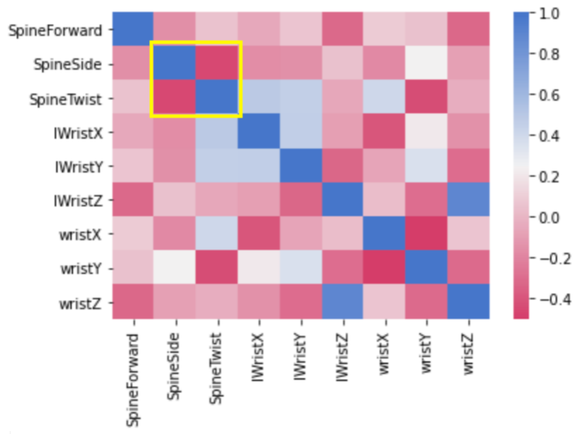


Work-Flow

Name	Action	Rotation Axis
Twist	Transverse twisting	Z-axis
Side	Lateral bending	X-axis
Forward	Sagittal bending	Y-axis

Targets

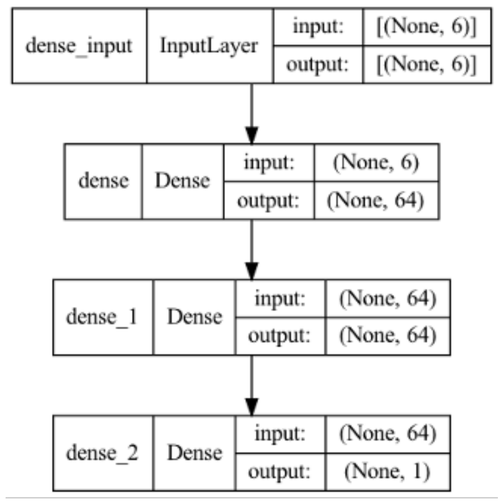
The predictor variables employed in the study are linear X, Y, and Z positions of the left and right wrists.



Correlation plot

Based on correlation analysis and experiments with linear regression, it was determined that chained regression between the target variables was appropriate.

Applying Neural Network



Model Parameters:
- Keras Sequential Model
- Simple multi-layer perceptron with three layers with the shape of the predictors
- First and second layers contain 64 units
- output layer contains one unit
- Rectified linear activation function (ReLU),
- 'Adam' optimizer, a stochastic gradient descent method for the training model.
- 100 epochs.
- 'Mean Squared Error' as the regression loss function

Results

The model was trained from the descriptions of the first 80% of scenes and held out the rest as a test set.

The metrics reported are loss, root mean squared error (RMSE), and RMSE of the test set.

The proposed model improves accuracy over linear regression, the companion model, and also over the the heuristic model which is currently used on the avatar.

	Twist	Side	Forward
MSE	1.68	4.82	6.89
RMSE	1.30	2.20	2.63
R^2	0.95	0.70	0.48

Model Performance

Model Comparison

Model	R-Squared	RMSE
Neural Network	0.95	1.30
Linear Regression	0.86	3.62
Heuristic Model	-	4.60

Twist

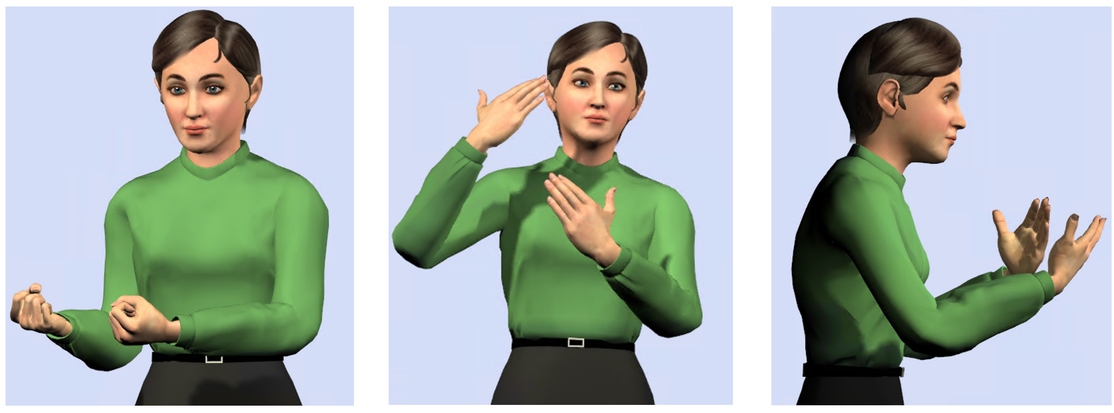
Model	R-Squared	RMSE
Neural Network	0.70	2.20
Linear Regression	0.24	11.29
Heuristic Model	-	3.5

Side

Model	R-Squared	RMSE
Neural Network	0.48	2.63
Linear Regression	0.17	21.73
Heuristic Model	-	3.2

Forward

Implementation



Proposed Model implemented on avatar

The proposed model is implemented on an avatar. The initial testing and validation produce satisfactory results. The model is implemented in the avatar using Python and tested against the original mocap positions. A user survey is planned to measure the improvement of naturalness.

Future Work

- Create a multi-target neural network model to combine the three models
- Handle overfit issue
- Incorporate personal signing style
- Include additional predictor variables and data.
- Leverage the framework to other sign languages,
- such as German or Mexican

Paper with references

Please see the published paper with references. Thank you!

Contact me

DePaul University