Enhancing Human Body Pose Classification through Intermediate Joint Inclusion with SVM

Eric McKevitt
Department of Computer Science
Colorado School of Mines
Golden, Colorado
emckevitt@mines.edu

I. Introduction

Developing skeleton-based representations of joint positions in 3D space is a prevalent approach to modeling human skeletons and body postures. In this project, I represent body poses using labeled point data for various joints, constructing a series of histograms that denote joint positions and relative angles over time.

Two representations are implemented: Relative Distances and Angles (RAD) and a custom representation designed to incorporate intermediate joints between extremities and a body center, improving body pose estimation. The histograms are concatenated into featurized vectors and used as inputs for a Support Vector Machine (SVM) to classify the observations into distinct body positions.

The SVM is trained with various hyperparameter configurations to identify the optimal classification model. Each model's performance is evaluated using accuracy, precision, and confusion matrices.

II. RAD Featurization

The Relative Distances and Angles (RAD) featurization technique relies on calculating the distances from each extremity to the center of the body and the angles between each extremity. In this experiment, the 6 highlighted joints in Figure 1 are used.

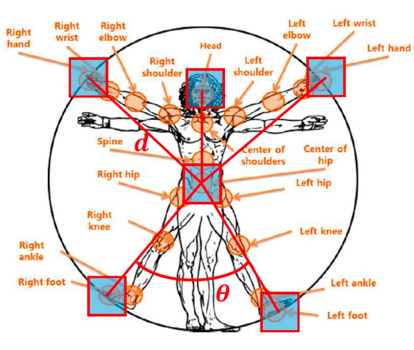

Fig. 1. RAD Representation for Body Pose

To find the distance $d_{i,j}$ between two joints, $i, j \in \mathbb{R}^3$, the following distance equation is used:

$$d_{i,j} = \sqrt{(i_x - j_x)^2 + (i_y - j_y)^2 + (i_z - j_z)^2} \tag{1}$$

To find the relative angle between two joints, we treat the joints as vectors drawn from the origin, and find the angle $\theta_{a,b}$ between the vectors $\vec{a}$ and $\vec{b}$ using the following equation [1].

$$\theta_{a,b} = cos^{-1}\left(\frac{a_x b_x + a_y b_y + a_z b_z}{\sqrt{(a_x^2 + a_y^2 + a_z^2) \cdot (b_x^2 + b_y^2 + b_z^2)}}\right) \tag{2}$$

Using a series of distance and angle values pertaining to the selected joints, histograms can be created, allowing for an effective representation of the pose data to be input into an SVM for classification.

III. Custom Featurization

In the custom featurization, I incorporate elbow and knee joints in addition to the five extremity joints. Including these intermediate joints aims to provide more granularity and capture nuances in arm and leg movements across different activities. While adding joints may increase the model's complexity, it potentially allows for a higher level of sophistication and performance.

IV. Histogram Creation

To create histograms for both RAD and custom featurization, the data was first divided into bins based on joint positions and angles. The number of occurrences in each bin was counted to form the histograms. These histograms were then normalized by dividing by the total number of frames in each data instance to account for varying frame counts. The histograms resulting from RAD featurization can be seen in Figure 2. The custom featurization method creates more histograms as there are more joints involved.

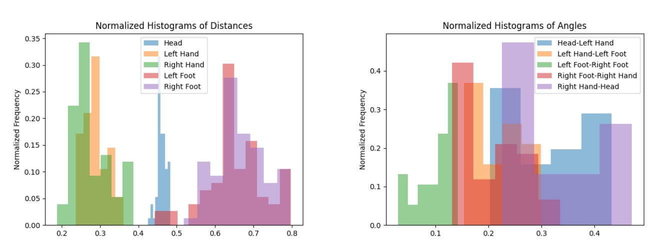

Fig. 2. RAD Distance and Angle Histograms

The histograms for each representation were concatenated into one-dimensional vectors, which will serve as the input features for the SVM classifier. By effectively representing body poses using histograms, the classifier should be able to learn and predict human activities from the dataset.

When creating histograms, the number of bins is an important parameter that can affect the future SVM's performance. I experimentally tested combinations of bin widths for both the RAD representation and the custom representation, and found the optimal number of bins for each representation.

V. Support Vector Machines

Support Vector Machines (SVMs) are a powerful and widely used class of supervised machine learning algorithms for classification and regression tasks. SVMs are particularly well-suited for high-dimensional datasets, which makes this model a good fit for this problem as the dimensionality of the featurized data will be relatively high.

The SVM model aims to find an optimal linear division between classes such that the margin between the classes is maximized. For higher-dimensional data, the SVM will find a hyperplane to separate the data. In the case that the data is not linearly separable, the SVM can project the data into a higher dimension using a kernel function, and then find a hyperplane or decision boundary in the higher dimensional space to separate the data.

The softness of the margin (C), is one hyperparameter. The kernel function is another. Finally, there is a hyperparameter for gamma (γ), the strictness of the decision boundary. In this experiment, the `Scikit SVC` implementation of a SVM classifier was used [2].

VI. Hyperparameter Selection

In order to find the optimal set of hyperparameters, I defined a reasonable range of values for C and *gamma*, and then tested every combination with both sigmoid and rbf kernels using `ParameterGrid` from `sklearn`. The models were evaluated to find the configuration that maximized accuracy. This experiment was conducted separately for both the RAD featurized data as well as the custom featurized data. The findings are discussed in the following sections.

A. Hyperparameter Tuning for SVM with RAD Featurization

Values of C between 0.001 and 1000 with a roughly logarithmic scale were chosen for the parameter grid. Gamma values between 0.0001 and 5 were chosen, also following a roughly logarithmic scale. As previously stated, the selected kernel functions were sigmoid and rbf. Each combination of hyperparameters was used to construct an `SVC` model, which was fitted to the RAD featurized data and then evaluated for accuracy. The hyperparameter tuning results were visualized using a heatmap.

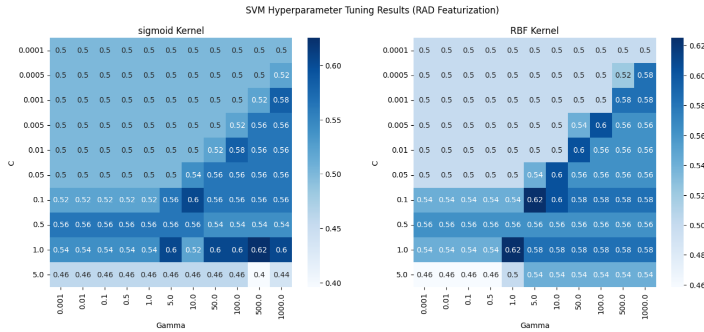

Fig. 3. RAD Hyperparameter Tuning Results

As seen in Figure 3, the configuration with the highest accuracy was an rbf kernel with $C = 1$ and $gamma = 1$. The accuracy for this model was 0.625, and the precision was 0.630. To better analyze the performance of the model, refer to the confusion matrix in Figure 4.

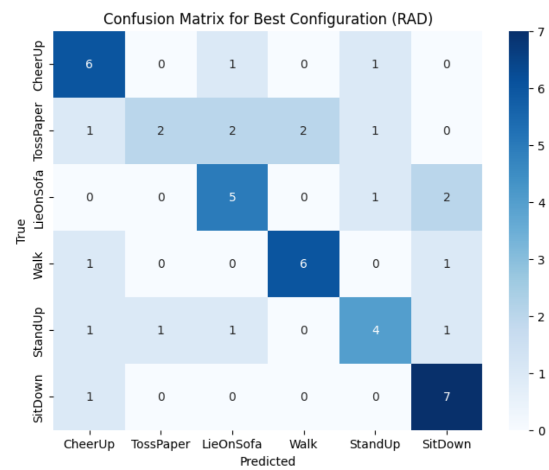

Fig. 4. RAD Optimal Configuration Confusion Matrix

The main diagonal is well-populated, indicating the model correctly predicted a good proportion of test samples. The most troubling action was `TossPaper`, which was predicted correctly only 25% of the time. In contrast, the model did quite well in predicting `SitDown`, which was correct 88% of the time.

B. Hyperparameter Tuning for SVM with Custom Featurization

The same selection of values for C, *gamma*, and kernel were chosen for testing with the custom featurized data. Using the same procedure as before, the following heatmap was obtained:

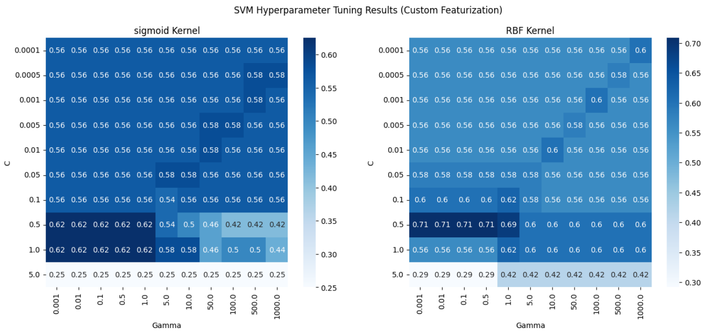

Fig. 5. Custom Featurization Hyperparameter Tuning Results

The optimal configuration, as seen in Figure 5, is given by $C = 0.001$, $gamma = 0.5$, and an rbf kernel. This configuration gives an accuracy of 0.708 and a precision of 0.716. Refer to Figure 6 for the confusion matrix generated by the optimal model for the custom featurization.

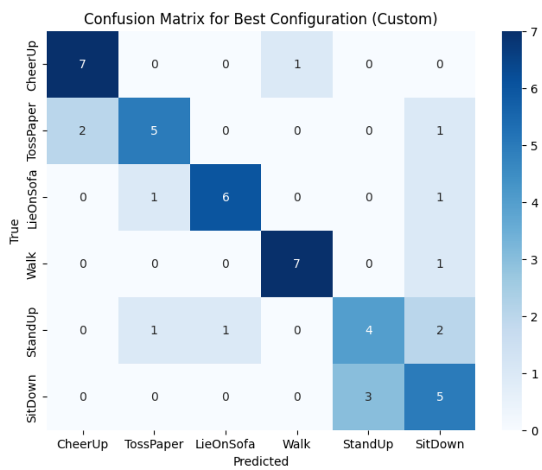

Fig. 6. Custom Featurization Optimal Configuration Confusion Matrix

An immediate observation is that the main diagonal for the custom featurization is much more defined than that of the RAD featurization, which further demonstrates the increase in accuracy. `TossPaper`, which was the most troublesome action for the RAD model to predict, saw an increase from 0.25% to 0.625% correctly labeled observations. 4 out of 6 actions saw increases in prediction accuracy, while `StandUp` saw no change and `SitDown` saw a decrease in performance by 25%.

Contrary to expectation, adding intermediate joints did not help distinguish between `StandUp` and `SitDown`. It did, however, increase the model's general performance, as suspected.

VII. Conclusion

In this study, we explored two featurization techniques, Relative Distances and Angles (RAD) and a custom featurization incorporating intermediate joints, to effectively represent human body poses for classification tasks. Support Vector Machines were employed to classify these poses based on the featurized data. The selection of appropriate hyperparameters played a crucial role in determining the performance of the SVM models.

The results indicate that the custom featurization technique, which included additional intermediate joints, demonstrated improved performance compared to the RAD featurization. Both featurization techniques produced models with relatively high accuracy and precision.

The optimal configuration of hyperparameters for the custom featurization resulted in an 8.3% increase in accuracy over the optimal RAD featurized model. Further research in this areas may focus on refining the featurization process and exploring other machine learning algorithms to enhance classification performance.

References

[1] "linear algebra - Calculating an Angle from 2 points in space," Mathematics Stack Exchange, Jul. 20, 2010. https://math.stackexchange.com/questions/59/calculating-an-angle-from-2-points-in-space (accessed Apr. 24, 2023).

[2] scikit-learn developers, "sklearn.svm.SVC — scikit-learn 0.22 documentation," Scikit-learn.org, 2019. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

[3] "sklearn.metrics.confusion_matrix — scikit-learn 0.21.3 documentation," Scikit-learn.org, 2019. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html