1. Models

Models are the core of data in LSP. They can represent a single loop to one hundred loops or even more. A model is designed to best represent a loop or group of loops. It has 5 primary components:

1. A probabilistic sequence
2. , the angle between first SSE and the structure
3. , the angle between the second SSE and the structure
4. The distance between the atom touching the structure in both neighboring SSEs
5. The atom positions relative to the first SSE

In order to obtain atom positions relative to the first SSE, we transform the atom positions from global space to a frame within global space that is oriented with the protein. In order to establish a reference frame for the protein in 3D, we use 2 vectors – the vector along the first SSE and the vector from the first SSE to the last SSE.[[1]](#footnote-1) We normalize the vector from the first SSE to the last SSE and use that as as the x-axis vector. We cross the x-axis vector with the vector along the first SSE and normalize that to obtain our z-axis vector. Finally, because we don’t know if our x-axis is orthogonal to the first SSE, we cross the x-axis and z-axis to get a y-axis. The origin for the frame is the closest atom to the structure in the first SSE. We can then use this information to get a relative position for the atoms. For an atom and frame , .[[2]](#footnote-2)

1.1 Model comparison

Model comparison has two stages: deciding if two models are similar enough to be compared and giving a comparison value for those models. Deciding if two models are similar enough to be compared is, unfortunately, decided by some magic numbers. First, we compare the structures using an RMSD calculation for each atom. If the average of the distances is more than 2, we determine the two structures not similar enough.[[3]](#footnote-3) If they are similar enough, we then compare the values for , , and anchor distance. We keep track of the global minimum and maximum values for each, then give a score between 0 and 1 for the value. For example, ’s a score would be . If this value for any of , , or anchor distance is greater than , we consider the models too different to compare.

During training, if the models are too different to compare, we say they have a score of , so we know they are not at all similar. If they aren’t too different to compare, we give them a score that is the sum of their , , and anchor distance scores. This can range between 0-0.15 due to the maximum value of 0.05 for each component. During testing, we will always compare two models, using the same metric as we did with training. However, because we don’t use the maximum value of 0.05 for each component during testing, our comparison score can range between 0-3.

1.2 Model merging

Model merging is a fairly straightforward process that relies a lot on averaging. This is possibly a place to improve. In order to merge 2 models, we consider each of their 5 components. For the , , and anchor distance, we average the two values, weighting based on the number of loops composing each model. For example, if we make model from models and , where model was created from 10 loops and model was created from 5, we would have . For the atom positions, the position of each of the atoms in the new model is the weighted average of the x,y,z coordinates for each atom, again weighted on the number of loops in both of the merged models

1. Cite: <http://www.cs.dartmouth.edu/~cs77/slides/04_transforms.pdf> [↑](#footnote-ref-1)
2. Cite: Algorithm from <http://www.cs.dartmouth.edu/~cs77/assignments/assignment02.zip> [↑](#footnote-ref-2)
3. Cite Yoonjoo here [↑](#footnote-ref-3)