Protocol 15: Result Analysis

19.04.2019

Spatial Temporal Graph Convolutional Networks (ST-GCN)

The initial purpose of spatial temporal graph convolutional networks (st-gcn) [1] is to classify actions from videos. Due to some common factors such as spatio-temporal information for behavior understanding, the networks are applied to the task of lameness detection. A pre-trained model was used, and only the last fully-connected layer was trained.

1. Data

- **Classification:** The initial step treats the task as a classification problem. Even though the locomotion scores have five levels from 1 to 5, four classes are considered here:
  - Class1: LS 1-2
  - Class2: LS 2-3
  - Class3: LS 3-4
  - Class4: LS 4-5 (containing 5)

- **Dataset:** Each data sample contains the coordinates of 18 skeleton joints from 300 video frames and a locomotion score as the label. The dataset has 501 samples, divided into training and test sets with a 70/30 ratio.

- **Class distribution:** Figure 1 shows the class distribution of both the training and test sets.
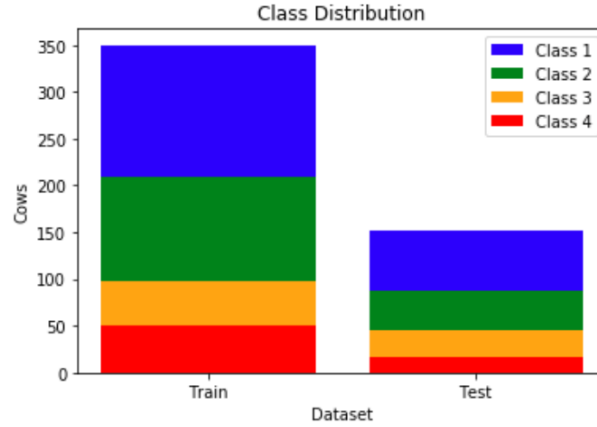

Figure 1: Class distribution of training and test sets.

2. Result and Discussion

The model was trained with a batch size of 16 and an initial learning rate of 0.1 for 80 epochs. As the initial step of result analysis, only the test result of one model is analyzed below. Table 1 is the confusion matrix, and the overall accuracy is 0.497. Even though the model did not perform well, the prediction seems to has a normal distribution around the true label for class 1 and 2. Since there are not many samples for class 3 and 4, the prediction looks somehow random. From a closer analysis of the worse prediction (e.g. class 1 predicted as class 4), one cause of poor prediction results from poor pose estimation (the coordinates of skeleton joints).

- **Confusion matrix:**

Table 1: Confusion matrix

	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 4
Actual Class 1	33	24	6	1
Actual Class 2	8	24	9	0
Actual Class 3	8	6	13	2
Actual Class 4	3	1	8	5

Table 2 lists the precision and recall of the classification result. While the recall is similar for all the classes except fro class 4, the precision value is higher for class 1 and 4. A low recall value for class 4 means the model has low sensitivity of detecting lameness, which may be caused by relatively less data of the severely lame cows.

- **Precision and recall:**

Table 2: Precision and recall

Class	Precision	Recall
1	0.635	0.516
2	0.436	0.585
3	0.361	0.448
4	0.625	0.294

3. Next Step

To improve the model for lameness detection, the first step is to improve the data quality, so the pose estimation will be carried out to extract more accurate skeleton joints. Another possibility is to use different skeleton joints that may extract more important features for lameness detection. After getting cleaner data, the model will be trained and tested again.

Bibliography

[1] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.