Improving Motion Classifier Robustness by Estimating Output Confidence

Durell Bouchard

Roanoke College, Salem VA 24153, USA

1 Introduction

Embodied conversational agents that can sense and respond to multiple modalities of user communication, like speech, gesture, and facial expressions, create a better impression and facilitate communication [1,2]. Responding to a user's gestures entails classifying the content and quality of each gesture, but classification performance is dependent on the selection of input sequence boundaries. Small changes in the boundaries of an input sequence can have a large effect on classifier output. Failing to correctly classify a user's gestures may cause an agent to respond incorrectly, which can negatively impact the agent's ability to communicate.

Motion classifiers must be robust to changes in input boundaries to create effective conversational agents. This poster outlines a method of modifying any learning based motion classifier to estimate the confidence of the classifier's output. The method calculates confidence by using multiple classifiers that are sensitive to different input sequence boundaries. Preliminary results show that the classification rate of a motion classifier improves by selecting input sequences with highest confidence estimation.

2 Model

The error rate of a motion classifier is dependent on the location of the input motion sequence's boundaries. A shift of a single frame in either the start or end boundary of an input motion sequence can have a negative impact on the classifier's error rate.

The effect of input sequence boundary shift on a motion classifier's error rate is a result of the selection of the boundaries of the training motion sequences. For example, if two identical classifiers are trained with the same motion sequences, except the start boundary of every training sequence is shifted by one frame forward or backward, then one classifier will have a lower classification error rate. If, however, the boundaries of input motion sequences are shifted the same way as the boundaries of the training motion sequences, then the output of the two classifiers is similar.

The similarity between two classifiers trained with the same motion sequences, but with different boundaries, can be used to estimate the confidence of a motion classifier. The less similar the outputs of the two classifiers are, the less likely it

T. Bickmore et al. (Eds.): IVA 2014, LNAI 8637, pp. 64–66, 2014.

Table 1. Classifier Error Rates

(a) Static Boundary

Training	Testing
0.09	0.35

(b) Variable Boundary

Training	Testing
0.08	0.12

is that the outputs are correct. The similarity of the two outputs is a measure of the confidence of their combined output. Furthermore, if there are more than two classifiers, trained with different permutations of boundary shift, the probability that all of the classifiers will produce the same incorrect output is low.

A compound classifier, multiple motion classifiers trained with different permutations of boundary shift, is used to search for input sequence boundaries with the highest output confidence. The confidence of the compound classifier's output, c, is the reciprocal of the average deviation of all of the sub-classifiers and is calculated as

$$c = 121/(\sum_{i=1}^{11}\sum_{j=1}^{11}|a - o_{ij}|) \tag{1}$$

where o is the output of a sub-classifier and a is the average output of all of the sub-classifiers. The compound classifier uses 121 sub-classifiers because the effect of boundary shift plateaus at 5 frames of total boundary shift and there are 121 permutations of shifting an input sequence's start and end boundaries by 5 frames or less. The average output of the classifiers with the highest confidence is the final output of the variable boundary compound classifier.

3 Preliminary Results

The variable boundary compound classifier is tested with a feed-forward neural network that uses simple kinematic features, such as average velocity and initial acceleration, to classify Laban Movement Analysis (LMA) Effort factors of motion capture data. Zacharatos et al.[3] demonstrated the ability and utility of classifying the LMA Effort factors of movements.

Table 1 summarizes the 24-fold cross validation error rates of the variable-boundary compound classifier and the static-boundary simple classifier. The 288 non-emblematic training motions represent a diversity of movements as defined by LMA. Note that the error rates of the static-boundary classifier and variable-boundary classifier are similar for the training set but are different for the testing set.

The training set error rates of the two classifiers are similar because the classifiers are sensitive to the boundaries of the training motion sequences. The testing set error rates are different because the compound variable-boundary classifier is more robust to the segment boundary shifts in sequences on which it was not trained. Note that the difference between training and testing error rates is 0.26 for the static-boundary simple classifier, but is 0.04 for the variable-boundary compound classifier.

The variable-boundary compound classifier is more robust to variability in input sequence boundary selection for an LMA neural network classifier. Future work will evaluate the impact of interacting with a conversational agent that uses a variable-boundary compound classifier. The compound classifier's confidence may also be useful in creating more sophisticated conversation agents. For example, a conversational agent should respond differently to a user who is definitely nervous than to a user who is potentially nervous. Future work should also evaluate using classifier confidence to create more nuanced conversational agents.

References

1. Cafaro, A., Vilhjálmsson, H.H., Bickmore, T., Heylen, D., Jóhannsdóttir, K.R., Valgarðsson, G.S.: First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 67–80. Springer, Heidelberg (2012)
2. Huang, L., Morency, L.-P., Gratch, J.: Virtual rapport 2.0. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS, vol. 6895, pp. 68–79. Springer, Heidelberg (2011)
3. Zacharatos, H., Gatzoulis, C., Chrysanthou, Y., Aristidou, A.: Emotion recognition for exergames using laban movement analysis. In: Proceedings of Motion on Games, pp. 39–44. ACM (2013)