Motion Karaoke

Xiaochao Yang

Introduction:

Machine learning has been applied into digital entertainment in various ways. In this project, I will explore the possibility of a motion Karaoke system. Like traditional Karaoke system, an amateur will follow the instructions and perform the same action. Instead of vocal removed songs, here the instruction is in the form of prerecorded motions given by skilled performers. The system will evaluate the quality of a specific movement performed by an amateur and give a grade accordingly. This grading system has several potential applications: learning dancing or martial-arts at home without an expensive tutor, contesting with friends for fun and so on.

Problem reformulation:

The raw input of the motion is joint angle vectors data obtained from a motion capture system. The system should model the output grade and the similarity measure between the performance of an amateur (in the training set) and that of a professional, derive a function between the two in the offline learning procedure, and determine the grade of a new outlier (in the test set) performing the same motion task. This is a regression problem.

Let v_i ($i=1, 2, \ldots$)be a joint angle vector, the motion data of a professional with T_p frames and that of an amateur with length T_a frames can be expressed as:

$$\tilde{\mathrm{P}} = \left[\tilde{\mathrm{v}}_1, \tilde{\mathrm{v}}_2, \ldots, \tilde{\mathrm{v}}_{\mathrm{T_p}}\right] \text{ and } \widetilde{\mathrm{A}} = \left[\tilde{\mathrm{v}}_1', \tilde{\mathrm{v}}_2', \ldots, \tilde{\mathrm{v}}_{\mathrm{T_a}}'\right].$$

To make two movements comparable in time, we use dynamic time warping (DTW) to align $\tilde{\mathrm{P}}$ and $\widetilde{\mathrm{A}}$, then we have two matrices with the same size:

$$\mathrm{P} = [\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_{\mathrm{T}}] \text{ and } \mathrm{A} = [\mathrm{v}_1', \mathrm{v}_2', \ldots, \mathrm{v}_{\mathrm{T}}'].$$

When we look at the each angle, i.e. each element of vector in *v*, it is a time series with length *T*. So we can find the difference between two motion patterns from the time series of each angle by treating P and A row-wisely. P and A can be rewritten as:

$$\mathrm{P} = \left[\mathrm{p}_1^{\mathrm{T}}, \mathrm{p}_2^{\mathrm{T}}, \ldots, \mathrm{p}_{\mathrm{N}}^{\mathrm{T}}\right]^{\mathrm{T}} \text{ and } \mathrm{A} = \left[\mathrm{a}_1^{\mathrm{T}}, \mathrm{a}_2^{\mathrm{T}}, \ldots, \mathrm{a}_{\mathrm{N}}^{\mathrm{T}}\right]^{\mathrm{T}}.$$

where N is the number of joint angles. Now the similarity between the professional's motion and the amateur's can be evaluated on each angle:

$$\mathrm{x_i} = \mathrm{SIM}(\mathrm{p_i}, \mathrm{a_i}) \;\; \mathrm{i} = 1, 2, \ldots, \mathrm{N}$$

The similarity measure function SIM can be Euclidean distance, cosine distance etc. For Euclidean distance, the result should be normalized before further processing, since the magnitude of *x* depends on dynamic time warping result *N* which is instance dependent. Now we have a vector $\mathrm{S} = [\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_{\mathrm{N}}]$ representing the similarity. In training procedure, we assign m vectors with different grades according to the quality of the performance. Then we have a set of tuples:

$$D = \{(x^1, y^1), (x^2, y^2), \ldots, (x^m, y^m)\}$$

The goal is to learn $y = f(x)$. This is typically a regression problem. If N is too large, we can use only a portion of joint angles with drastic movements, or apply some dimensionality reduction method to reduce the length of the vector.

Schedule

I will ask two dancers to perform some specific movements (which need skill and training), and ask other volunteers to reproduce the movements in several ways: try his/her best to imitate, intentionally ruin his/her performance, and do an average job. For the movement, I think it should not contain too many rotations, so the amateur can stare at the screen and keep the movement sync with the professional's (That's what the Karaoke does).

Before I have the first hand data collected by milestone, I plan to implement trivial parts of the system: dynamic time warping, several regression algorithms, and get familiar with the motion data processing using public dataset.

Through this project, I can learn two things: getting familiar human movement analysis with motion capture, and gaining the experience in regression. I should learn how to solve the problem encountered in the project, and foster capability of analytical thinking.