

Peer Review

Review of: "Synthetic Data Generation of Body Motion Data by Neural Gas Network for Emotion Recognition"

Julian Tejada¹

1. Department of Psychology, Universidade Federal de Sergipe, Brazil

This research introduces a novel application of the Neural Gas Network (NGN) algorithm for synthesizing body motion data and optimizing diversity and generation speed. The NGN learns skeletal structure topology by fitting neurons or gas particles on body joints, which are then used to synthesize new body postures and motions. The authors used a subset of motion capture data from the 100Style dataset to evaluate their approach. These subsets, which represent human movement expressing specific emotions, were enhanced with body motion data generated by NGN. The enhanced dataset is compared against others generated by GANs, VAEs, and another benchmark algorithm, using benchmark metrics such as Fréchet Inception Distance (FID), Diversity, and classification metrics. The findings demonstrate that the NGN algorithm produces more realistic and emotionally distinct body motion data with greater synthesizing speed than existing methods.

The authors' approach is clearly described in the manuscript; however, there are two things that I would like to ask them. First, the manuscript lacks a clear explanation of how their proposal is superior to a simple kinetic parameter analysis, in which the velocity, acceleration, or jerk can be used for emotion classification, without needing to enhance the dataset with NGN techniques. To further clarify the relationship between kinetic parameters and emotional states, please expand the matrix in Figure 6 to include emotions. This will allow for confirmation of the extent to which kinetic parameters, such as velocity, acceleration, or jerk, can be used to characterize emotional states as represented by human movement, if these parameters showed a high correlation with a specific emotion, as suggests Figure 5.

Second, it is also crucial to emphasize some of the drawbacks of the 100Style dataset, particularly in terms of expressing emotion. This is due to the dataset being constructed from the motion capture of only one individual, who acted out movements resembling emotions, moving in circles around the

camera frame. Since the dataset was created using only one actor and simulated emotions, the results may not generalize well to a wider population or real-world scenarios. The discussion should more thoroughly address these limitations because the model's performance may be inflated due to overfitting to the specific actor and the limited emotional range present in the dataset. Additionally, the discussion should also address the condition that the simulated emotions may not accurately capture the subtle nuances and complexities of real human emotions, which could impact the model's ability to accurately classify or generate emotionally expressive body motions.

Declarations

Potential competing interests: No potential competing interests to declare.