Time Series Extrinsic Regression for Physical Rehabilitation Assessment

Elvin Ismayilzada[1], Maxime Devanne[1], Jonathan Weber[1], and Germain Forestier[1]

IRIMAS, Université de Haute-Alsace
firstname.lastname@uha.fr

Abstract. Rehabilitation is the process of assisting people with disabilities in regaining their function and independence. As artificial neural networks are trained on large datasets using deep learning, rehabilitation can be improved by providing individualized and efficient treatment options. As human rehabilitation involves multivariate time series data, we review well-known algorithms for the classification of time series data. We also discuss the challenges and opportunities presented by the use of deep learning in rehabilitation, including the need for large and diverse datasets and the potential for bias in algorithms. Overall, our analysis indicates that deep learning has the potential to improve rehabilitation outcomes and the lives of disabled individuals. A comparison of many methodologies was conducted in order to establish a framework capable of supporting and reliably evaluating patients' workouts throughout recovery programs. In order to assess the algorithms, two datasets pertaining to human rehabilitation are used: KIMORE, and UI-PRMD for regression tasks.

1 Introduction

One of the most efficient methods to diagnose musculoskeletal issues and rehabilitate post-stroke participants is through physical therapy treatment through workouts on specific tasks. However, it is neither practical nor cost-effective for a physician to attend every rehabilitative exercise program. Patients typically execute these exercises at home, without the involvement of specialists or therapists. Despite the fact that patients are supposed to document and record their progress as well as contact the doctors on a regular basis, various medical organizations have stated that patients are not able to complete the exercises correctly, causing the recovery process to be extended. As a result, patients are unable to receive proper supervision and feedback for the required activity. These challenges make human rehabilitation a hot topic in a research environment [1]. Thanks to recent Computer Vision algorithms, it is possible to capture human motion by estimating from an image joints' 3D coordinates, forming a humanoid skeleton. In this paper, we use Deep Learning to analyze human rehabilitation programs, represented by 3D skeleton sequences, leveraging precise and outstanding performance. We propose to recast the problem of human rehabilitation movement assessment as a multivariate time series analysis. Time series analysis has been investigated for various tasks such as classification [2], clustering [3], averaging [4] and adversarial attacks [5]. In this work we particularly study a well-known architecture in the field of time series classification [6] for the task of rehabilitation assessment. However, there are two main issues that must be addressed prior to deployment:

1. First, we must switch the implementation domain from classification to regression in order forecast a single numerical score corresponding to patient's performance.

2. Second, because deep learning models assume fixed-size inputs, pre-processing procedures are mandatory before any deep learning model can be implemented on the human rehabilitation dataset.

We established our experiments on the KIMORE [7] and UI-PRMD [8] datasets. In particular, Kimore dataset is more practical for real-world situations since it incorporates both healthy and unhealthy subjects. We believe that our proposed method could be very helpful for physical rehabilitation assessment and could for instance be embedded into a autonomous system (like a robot coach [9]) for monitoring rehabilitation sessions in rehabilitation center or at home.

2 Background and Related Work

2.1 Multivariate Time Series

A multivariate time series [6] contains multiple time-dependent features. Multivariate time series $X = [X^1, X^2, ..., X^M]$ contains M individual univariate time-series where $X^i \in \mathbb{R}^T$.

2.2 Convolution

Convolution can be used to [6] perform and sliding filter through time series. It just has one-dimensional filters (time) rather than two dimensions like images. The filter can alternatively be viewed as a non-linear change of time series.

$$Ct = f(\omega * Xt - l/2 : t + l/2 + b) \mid \forall t \in [1, T], \tag{1}$$

2.3 Classification

Various works have been investigated in order to replace the expensive and arbitrary judgment of human experts with an automated process. Classification evaluations are used to estimate categorical values that indicate the ability level of the activities that are performed. Such evaluations categorize executed motions into distinct groups that belong to a rank but can be challenging to precisely define. Assessments may fall into one of two groups in a straightforward classification system (correct or incorrect). On the other hand, adding more classes allows us to make more exact distinctions between executed motions. More examples from each class are required for better performance. As a result, if there are numerous classes or features that enhance complexity, they may create a scaling difficulty.

2.4 Parametric Assessment

Rather than a classification task, a parametric assessment provides continuous value in order to evaluate the performance of rehabilitation exercises. These methods apply strategies similar to other approaches but they emphasize domain-specific factors. As a result, such methods offer helpful detail in their evaluations of executed human rehabilitation exercises. Some research for exercise assessment tasks often concentrates on learning distance measures [10]. Those techniques can find similarities between two random exercises, however, they can't represent task-specific exercises. To solve this issue, another line of research relies on probabilistic techniques for evaluating workouts, such as Hidden Markov models [11], [12] and Gaussian mixtures [13]. These techniques impede end-to-end processing as they need many pre-processing phases and the knowledge of experts in the particular field. We intend to analyze exercises using deep learning techniques since deep learning algorithms are better suitable for end-to-end processing.

2.5 Evaluation of Rehabilitation Exercise

There hasn't been enough research on this subject. [14] Lee et al. classified a variety of motions into true and false categories with the help of hand-crafted features. In the work of [15] for the evaluation of human rehabilitation exercises spatio-temporal architecture is suggested. In order to boost performance, multi-branch convolution, recurrent networks, and temporal pyramid. The downside of these approaches they employ a variety of feature engineering and pre-processing steps. In light of the recent achievement of deep learning approaches in various fields, we conduct human rehabilitation exercise assessments with deep learning frameworks. We use common time series classification algorithms [6] for this study since human rehabilitation activities involve coordinates of joint positions throughout time.

3 Proposed Approach

In this paper, we adapted the inception network to human rehabilitation exercises. The inception network [16] is convolutional neural network initially created for a more deep representation of time series classification problems. The inception network is made up of two main concepts: bottleneck layers and sliding several filters. By utilizing a bottleneck layer, time series data dimensions can be reduced while capturing complex features and overfitting issues can be minimized. Moreoever, this architecture allow to slide several filters of varying lengths over the given input time series at the same time in order to capture meaningful patterns at different scales. Our inception network is adapted to human rehabilitation and extrinsic regression to generate a numerical value that represents the score associated with the input motion sequence. Figure 1 illustrates the proposed architecture.

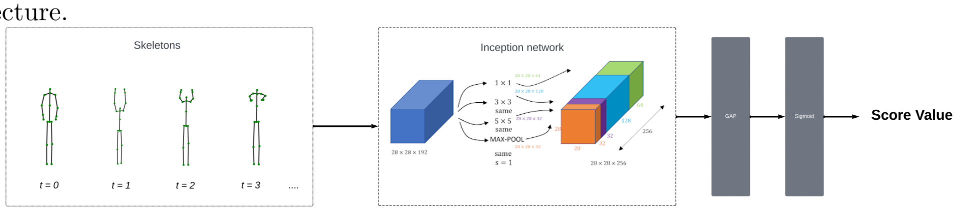

Fig. 1: The overall framework of our proposed rehabilitation exercise assessment

4 Experiments and Results

4.1 Datasets

For evaluating the proposed approach, we use two separate datasets: KIMORRE and UI-PRMD. KIMORE dataset [7] contains RGB-D videos collected by Kinect sensor and clinical scores evaluation of human motions. There are 5 distinct exercises done by control groups and an unhealthy group. The control group included 44 healthy people, 12 of whom were physiotherapists with competence in the treatment of postural and back pain, while the remaining 32 were non-expert healthy people. The unhealthy group consists of 34 patients who suffer from pain and postural issues and have chronic motor impairments. (2) UI-PRMD dataset [8] is made up of ten rehabilitation activities that were gathered from 10 healthy individuals utilizing Kinect and Vicon sensing devices. There were ten repetitions of the same activity performed by each individual. A comprehensive collection of data is provided, which includes the positions and angles of the joints throughout the body.

4.2 Normalization

Before feeding data through deep learning algorithms, normalization processes must be performed on the data. Because variability in input feature scales can increase the complexity of the model performance. In general, models that are constructed with heavy weights are unstable, meaning that they perform poorly during learning and exhibit high sensitivity to input values. The learning process can become unstable when a target variable has a wide spread of values, and the error gradients may be large, causing weight values to change drastically. As long as your output activation function has a scale of [0, 1], so the target values also fall inside that range. As a result, we normalize all human motion data into the range [-1, 1], and clinical values between 0 and 1. When we apply three different methods to normalize clinical scores:

- Initially, we divide the truth values by 100. The results are negatively affected by dividing by 100 since there is a large difference between the truth values of healthy and unhealthy subjects. Due to this, a model is not capable of capturing these differences between healthy and unhealthy clinical scores.
- Second, we apply minmaxscaler to clinical score labels to obtain values between 0 and 1. This strategy is effective when using the second and third splits, but for the first split, as train and test data is separated it treats the different distribution of data as the same values, therefore deep learning algorithms treat different clinical scores in the same manner. This leads models to mislead to find the desired output values correctly.

4.3 Evaluation Process

For evaluating regression tasks for human rehabilitation exercises we use two metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), as defined below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

The Root Mean Square Error indicates how condensed the values are close to the best-fit line. The Mean Absolute Error computes an absolute average gap between the actual and predicted data.

4.4 Experimental Results on Kimore

We train our model on the splitting technique using the leave-one-only cross-validation (LOOCV) on unhealthy samples. As part of this technique, one subject is used for testing, and the remaining data serve as a training set, which is combined with healthy and unhealthy samples. As Deep Learning models are dependent on their initial random initialization, we run our algorithm 5 times and report in Table 1 average values with standard deviations for both metrics on each exercise separately.

Table 1: Results of Inception algorithm 5 exercises (Ex) conducted on the KIMORE dataset by MAD and RMSE metrics

Metric	**Ex1**	**Ex2**	**Ex3**	**Ex4**	**Ex5**
RMSE	0.33 ± 0.04	0.31 ± 0.05	0.42 ± 0.04	0.34 ± 0.03	0.28 ± 0.02
MAE	0.19 ± 0.03	0.15 ± 0.03	0.27 ± 0.04	0.17 ± 0.03	0.16 ± 0.02

Considering that we used to leave one out cross-validation which improved model performance. Moreover, Figure 2 shows the comparison of real (green) and predicted (red) scores obtained by our approach on test sequences of exercise 2 (lateral tilt of the trunk with the arms in extension). We can see that our proposed method allow to correctly predict corresponding scores with low errors.

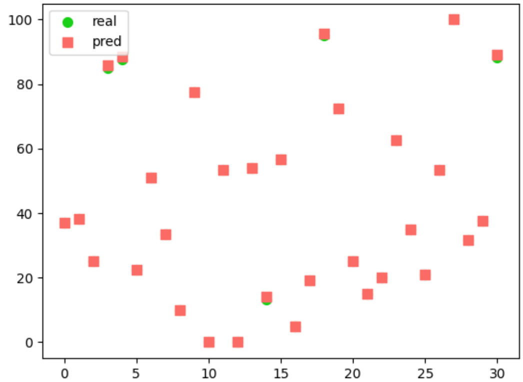

Fig. 2: The comparison plot between predicted and actual values for the Inception network on KIMORE dataset

The UI-PRMD dataset was subjected to the same experiments as the KIMORE dataset using the Inception model. In addition to other experiments related to human rehabilitation exercises, we applied subject-based leave-one-out cross-validation to prevent the use of the same subject information in both the train and test set at the same time.

Based on the Figure 3, we can conclude that our proposed method in the UI-PRMD dataset also shows promising results

Table 2: Results of 10 exercises conducted on the UI-PRMD dataset by MAD and RMSE metrics for the Inception model

Ex	**RMSE**	**MAE**
Ex1	0.0227 ± 0.0047	0.018 ± 0.0037
Ex2	0.0128 ± 0.0038	0.01 ± 0.0032
Ex3	0.0245 ± 0.0010	0.0178 ± 0.0011
Ex4	0.0282 ± 0.0026	0.0233 ± 0.0027
Ex5	0.242 ± 0.454	0.0841 ± 0.1442
Ex6	0.0186 ± 0.0017	0.0138 ± 0.0009
Ex7	0.0219 ± 0.0045	0.0172 ± 0.0029
Ex8	0.0316 ± 0.0050	0.0223 ± 0.0029
Ex9	0.02 ± 0.0022	0.017 ± 0.0025
Ex10	0.7436 ± 1.3082	0.2866 ± 0.4182

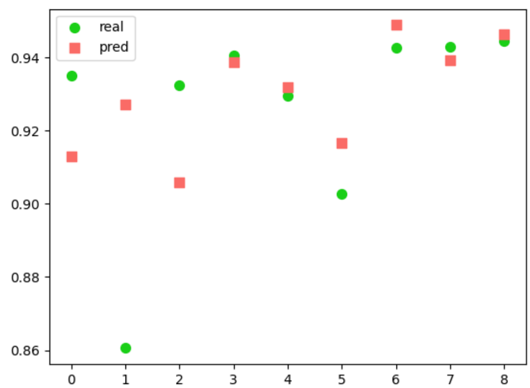

Fig. 3: The comparison plot between predicted and actual values for the Inception network on UI-PRMD dataset

5 Conclusion

In this paper, we proposed a human rehabilitation assessment approach using the Inception network. We modified the original architecture to provide numerical values by considering our analysis as a multivariate time series extrinsic regression problem. Our approach is evaluated on two datasets that represent rehabilitation exercises, the KIMORE dataset and the UI-PRMD dataset. Our results indicate that the Inception network algorithm can be used for human rehabilitation exercises with promising results. As future work, we aim at investigating explainability methods allowing to understand which part of the rehabilitation motion is more responsible of a given score.

Acknowledgement

This work was supported by the Fondation Matmut Paul Bennetot, Paris, France (grant AP-FPB-21-001). The authors would also like to thank the creators and providers of the KIMORE and UI-PRMD dataset.

References

1. Blanchard, A., Nguyen, S.M., Devanne, M., Simonnet, M., Le Goff-Pronost, M., Rémy-Néris, O.: Technical feasibility of supervision of stretching exercises by a humanoid robot coach for chronic low back pain: The r-cool randomized trial. BioMed Research International **2022**(1) (2022) 1–10
2. Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G.: Deep learning for time series classification using new hand-crafted convolution filters. In: IEEE International Conference on Big Data (BigData), IEEE (2022)
3. Ismail-Fawaz, A., Ismail Fawaz, H., Petitjean, F., Devanne, M., Weber, J., Berretti, S., Webb, G., Forestier, G.: Shapedba: Generating effective time series prototypes using shapedtw barycenter averaging. In: ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data. (2023)
4. Terefe, T., Devanne, M., Weber, J., Hailemariam, D., Forestier, G.: Estimating time series averages from latent space of multi-tasking neural networks. Knowledge and Information Systems **65**(11) (2023) 4967–5004
5. Pialla, G., Ismail Fawaz, H., Devanne, M., Weber, J., Idoumghar, L., Muller, P., Bermgeir, C., Schmidt, D., Webb, G., Forestier, G.: Smooth perturbations for time series adversarial attacks. In: Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Springer (2022) 1–12
6. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Deep learning for time series classification: a review. Data mining and knowledge discovery **33**(4) (2019) 917–963
7. Capecci, M., Ceravolo, M.G., Ferracuti, F., Iarlori, S., Monteriu, A., Romeo, L., Verdini, F.: The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering **27**(7) (2019) 1436–1448
8. Vakanski, A., Jun, H.p., Paul, D., Baker, R.: A data set of human body movements for physical rehabilitation exercises. Data **3**(1) (2018) 2
9. Devanne, M., Nguyen, S.M., Rémy-Néris, O., Le Gals-Garnett, B., Kermarrec, G., Thépaut, A.: A co-design approach for a rehabilitation robot coach for physical rehabilitation based on the error classification of motion errors. In: Workshop on Collaboration of Humans, Agents, Robots, Machines and Sensors (CHARMS), in conjunction with IEEE IRC, IEEE (2018) 352–357
10. Su, C.J., Chiang, C.Y., Huang, J.Y.: Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic. Applied Soft Computing **22** (2014) 652–666
11. Lin, J.F.S., Karg, M., Kulić, D.: Movement primitive segmentation for human motion modeling: A framework for analysis. IEEE Transactions on Human-Machine Systems **46**(3) (2016) 325–339
12. Vakanski, A., Ferguson, J., Lee, S.: Mathematical modeling and evaluation of human motions in physical therapy using mixture density neural networks. Journal of physiotherapy & physical rehabilitation **1**(4) (2016)
13. Wang, Z., Yan, W., Oates, T.: Time series classification from scratch with deep neural networks: A strong baseline. In: 2017 International joint conference on neural networks (IJCNN), IEEE (2017) 1578–1585
14. Lee, M.H., Siewiorek, D.P., Smailagic, A., Bernardino, A., Badia, S.B.i.: Learning to assess the quality of stroke rehabilitation exercises. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. (2019) 218–228
15. Liao, Y., Vakanski, A., Xian, M.: A deep learning framework for assessing physical rehabilitation exercises. IEEE Transactions on Neural Systems and Rehabilitation Engineering **28**(2) (2020) 468–477
16. Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D.F., Weber, J., Webb, G.I., Idoumghar, L., Muller, P.A., Petitjean, F.: Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery **34**(6) (2020) 1936–1962