# Move to Health! Temporal information aware semi-supervised learning to memorize proprioceptive effects in latent space

**Keywords:** Time Series, spatio-temporal varational autoencoder, short- and long-term analysis of temporal effects in kinematics, analysis of proprioceptive perception, effects on emotions.

**Abstract.** Industrialization, digitization and, most recently, AI brought many positive advances for mankind. However, these come at a high price: since 2010, daily screen time has increased dramatically while daily exercise has decreased proportionally. This leads to both short- and long-term psychological and physical degeneration and irreparable consequential damage, resulting in higher mortality in the long-term. To stop the degeneration, we suggest the following. First, the short- and long-term analysis of temporal effects in movement kinematics, the analysis of proprioceptive perception, and the effects on emotions. Second, deriving qualitative feedback in everyday life. The aim is to research a metric that automatically records and analyzes posture and movement as well as their effects on the psyche of people in everyday life in order to correctly identify possible complaints such as postural errors or chronic and mental illnesses in movement patterns. The metric enables an early warning system to inform users in good time and to provide suggestions for improvement via a software application for wearables. As the current data situation is poor, semi-supervised methods are used for data collection and analysis. Thus, they collect data from various information sources (phones and watches) and learn to map compressed representations in latent space.

**Problem.** Deep learning has become the dominant approach for supervised learning of labeled data [1-2]. However, in many applications the data labels are not available or not reliable. Here, unsupervised learning is used to learn more about the underlying structure of a data set without the need of ground truth labels. The process of learning anomalies in movement behavior and proprioception using inertial sensors is represented with time series data. Time series data are defined as any data that contain multiple measurements in sequence. Examples of time series data are the human gait cycle, facial expressions and gestures, which are recorded by, e.g., inertial sensors. So, a common learning task is to divide a number of time series, respectively movement and perception, into clusters. However, clustering time series data is a challenging problem, since time series data can be high-dimensional, noisy, and not segmented.

**Related work.** A variety of techniques have been developed for unsupervised learning in which the algorithms draw conclusions about unlabeled data. Known clustering algorithms such as hierarchical- or k-means clustering algorithms are susceptible to noise and time shifts. The improved metric Dynamic Time Warping [3] provides an invariance for time shifts, but is expensive to calculate. Unsupervised shapelets [4] also reduce problems with shifts and noise, but are limited to extracting a single feature from each time series. An autoencoder (AE) learns low-dimensional projections of high-dimensional data that are resolved by any clustering algorithm. In addition, AEs denoise the raw data and offer a degree of invariance against time drift. Semi-supervised learning increases the performance of the unsupervised models using small amounts of labeled data. However, latest approaches such as He et al. [5] show scalability problems with large amounts of data. Moreover, state-of-the-art AE methods suffer from significant model divergence when only few labels are available in training. As the quality of the latent space generated is not optimal, they cannot extend the clustering method to time series with missing values, and their generalization to spatio-temporal multi-channel inputs remains unclear.

**Idea.** Based on the ideas of Johnson et al. [5] and Griffiths et al. [6] we will investigate a semi-supervised deep learning model that will be based on a novel spatio-temporal varational autoencoder (STVAE) and that will be used to collect data from various information sources (mobile phones, smartwatches, headphones), learn to map compressed representations in latent space, and

cluster time series data sets. Upstream, we will map spatial features with convolutional networks and temporal features with recurrent neural networks or transformers to memorize features in the latent space of a varational autoencoder. In a quantitative and supervised review cycle, we combine these latent embeddings into final categories using known clustering processes. We argue that the clustering performance will significantly improve if we provide a small number of labeled time series in a supervised way while we gently optimize the VAE such that the reconstruction and the disentanglement (interpretability) are optimally balanced. Thus, our information metric represents learnable variants of temporal information of movements and its effects in both the input data and in latent space. Hence, with the help of the mutual information and information entropy metrics, the relationship between labeled inputs, the temporal features in latent space, and the final categories will be measured quantitatively. Since the input data contain multivariate data from various domains (accelerometer, gyroscope, WiFi, BLE, and GPS), we adapt the semi-supervised methods to multisensory data and we will also examine whether a cascaded architecture with several independent STVAEs and a global surrogate merging in a clustering process yields an optimal solution. In addition, the optimal preprocessing of the input data will be examined to embed abstract characteristics of the time series without loss of information and to enable a single embedding for multivariate data from various domains. The novel STVAE enables the semi-supervised learning of anomalous movement behavior and its effects of mankind. The STVAE clusters the temporal and causal information and checks the assignment with an external information pool, e.g., human supervisor, complementary features or loosely coupled external sensors. Finally, anomalies will be identified and a user will be notified and given suggestive feedback (e.g., posture and gait correction) for sustainable health.

## Overall goals.

- (1) Novel method for unsupervised motion analysis;
- (2) Evaluation and a posterior re/training (continual learning functionality);
- (3) Academic paper.

## Timetable (6 months, in person weeks [PW]).

- Data [1.5PM]: Literature review of publicly available data sets [0.25PM]; Data acquisition (optional) [1.25PM];
- Framework [2PM]: Investigation and implementation of a multi-hypothesis filter for multivariate data [1.5PM]; Toy example for a minimal set of sensors, e.g., accelerometer, gyroscope, and Wi-Fi/BLE/GPS [0.5PM];
- Semi-un/supervised methods [3.25PM]: Literature review of state-of-the-art semi-un/supervised methods [0.25PM]; Adaptation and improvement of the most modern methods [3PM];
- Anomaly Detection [4.25PM]; Literature review on information metrics for temporal and causal characteristics [0.25PM]; Investigation of information metrics and their effects on the data at hand [3PM]; Decorrelation and unbundling of temporal and spatial information in latent space with a substitute classifier (ground truth) [1PM];
- Evaluation [2PM]: Quantitative and experimental evaluation of the STVAE and framework [1PM]; Comparison with potential state-of-the-art of different fields of application [1PM];
- Publication [1PM]: At least one conference publication at ICLR, ICML, NeurIPS, AAAI, CHI, SIGCHI, PLoS ONE or CVPR.

| Student: | .. |
| Supervisor/s: | Tobias Feigl |
| Client: | .. |
| Thesis: | Bachelor or Masters Thesis |

## Expected results and scientific contributions.

- Unsupervised model and continual learning of novel and anomalous motion;
- Benchmark results of state-of-the-art and the novel models on prominent datasets;
- Scientific publication and GitHub publication.

# References

[1] LeCun Y. et al. (2015). Deep learning. Jo. Nature.

[2] Schmidhuber, Jürgen. (2015). Deep learning in neural networks: An overview. Jo. Neural networks.

[3] Berndt D. et al. (1994). Using dynamic time warping to find patterns in time series. KDD workshop.

[4] Zakaria J. et al. (2012). Clustering time series using unsupervised-shapelets. Intl. Conf. on Data Mining.

[5] He G. et al. (2019). A fast semi- supervised clustering framework for large-scale time series data. IEEE Transactions on Systems, Man, and Cybernetics.

[6] Johnson M. et al. (2017). Composing graphical models with neural networks for structured representations and fast inference. arXiv:1603.06277 [stat.ML].

[7] Griffiths T. et al. (2011). The Indian Buffet Process: An Introduction and Review. JMLR.