*#Title*: CIMAP: Clustering for Identification of Muscle Activation Patterns

*#Tags*: Python, surface electromyography, pattern clustering, machine learning

*#Authors*: Gregorio Dotti, Marco Ghislieri, Samanta Rosati, and Gabriella Balestra

*#Affiliations*: Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy; PoliToBIOMed Lab, Politecnico di Torino, Turin, Italy

*#Summary*

The main part of the movements of everyday life embodies a cyclical nature. Some examples are walking, running, cycling, stair climbing, and swimming. Gait analysis is broadly used for the clinical assessment of patients. To perform a quantitative assessment of muscle functionality during gait analysis, it is possible to investigate non-invasively muscular activity by acquiring sEMG (surface electromyography). From the sEMG is possible, using several methods [@bonato\_knaflitz; @ghislieri\_lstm], to identify the intervals when the muscle can be considered active or not. But, muscular activity, even during cyclical movements, such as walking, shows high intra-subject variability [@agostini\_1] both in the number of activation intervals within the same cycle (also called modality) and in their duration. The high variability in the activation patterns of the same muscle while performing a cyclical task reduce the reliability and complicate the interpretability of the obtained results. To address such issue, Statistical Gait Analysis (SGA) [@SGA] has been recently proposed. The SGA consists of a "statistical" description of gait Spatio-temporal parameters and parameters derived from EMG signals. The purpose of this approach is to describe gait functionality in a condition like everyday walking. From the SGA methods, the CIMAP (Clustering for Identification of Muscle Activation Pattern) [@cimap1; @cimap2] algorithm was developed to perform pattern analysis on the activation profiles extracted from the sEMG recording of walking. The aim CIMAP algorithm is to overcome the limitation in interpretability introduced by the residual variability that is still present after the application of the SGA methods. The clustering performed on the activation intervals allows the identification of few common muscle activation patterns among all the cycles to give the user easier to interpret results.

CIMAP is a Python algorithm based on agglomerative hierarchical clustering that aims at characterizing muscle activation patterns during cyclical movements by grouping strides with similar muscle activity. From muscle activation intervals to the graphical representation of the clustering results, the proposed algorithm offers a complete analysis framework for assessing muscle activation patterns that can be flexibly modified at need and applied to cyclical movements different from walking. CIMAP is addressed to scientists of any programming skill level working in different research areas such as biomedical engineering, robotics, sports, clinics, biomechanics, and neuroscience.

*# Statement of Need*

Traditional gait analysis most frequently analyses only a few cycles of a subject. However, this procedure does not allow for capturing the natural behavior of a subject and its cycle-to-cycle variability, especially in gait analysis. To address the issue, the practice is leaning towards the acquisition of high numbers of gait cycles when performing gait analysis, and new methods, like SGA [@SGA], have been developed. CIMAP algorithm [@cimap1; @cimap2] has been developed to help clinicians with the interpretation of EMG data in gait analysis highlighting with clustering the relevant activation patterns that characterize a subject while walking. Even though the algorithm was developed for gait analysis, it can be applied to all sorts of cyclical tasks.

The CIMAP package also allows the customization of results with the possibility of coloring the graphical results for a more in-depth analysis of the results.

*# Analysis workflow*

The typical workflow when using the CIMAP algorithm consists of the following 6 steps:

1. Data preparation (i.e., to read input data and convert them into the format needed for the following steps);
2. Data pre-processing (i.e., to split input data based on the number of muscle activations within each cycle);
3. Agglomerative Hierarchical Clustering [@clustering]:
   1. Creation of a hierarchical tree using two distance metrics (Manhattan and Chebyshev);
4. Selection of the optimal number of clusters [@cimap2];
   1. Identification of the optimal cutting point through the analysis of both the intra-cluster variability;
   2. Identification of the distance metric that has the best inter-cluster variability;
5. Cluster representation (available also at points 3 and 4, see Figure 1 for a representative example);
6. Data saving (clustering results are saved in an easy-to-read and open-source format).

Figure 1 | Example of sEMG activation intervals clustering for the XX muscle acquired from a representative healthy subject during a 5-minute overground walking at a self-selected speed. Add a detailed description of the figure.

A typical analysis workflow can be synthetically written as follows:

Default setting parameters are specifically optimized for the analysis of human locomotion. However, they can be easily overridden by modifying the arguments of the relevant functions. Further details are available on GitHub.

*# Availability*

The latest stable release of CIMAP is freely available on GitHub. Documentation and representative examples are freely available in each version’s readme file. The latest stable release of CIMAP can be easily installed through the bash shell or command prompt with the following command:

pip install CIMAP

Further details on the setup process and Python minimum requirements are available on GitHub.

*# Acknoledgments*

The authors are grateful to (in alphabetical order): Prof. Valentina Agostini, Dr. Cristina Castagneri, and Prof. Marco Knaflitz for their contributions to the development and validation of this approach. An up-to-date list of contributors is available on the GitHub page.

*# References*

@article{ghislieri\_lstm,

title={Long short-term memory (LSTM) recurrent neural network for muscle activity detection},

author={Ghislieri, Marco and Cerone, Giacinto Luigi and Knaflitz, Marco and Agostini, Valentina},

journal={Journal of NeuroEngineering and Rehabilitation},

volume={18},

number={1},

pages={1--15},

year={2021},

publisher={BioMed Central}

}

@article{bonato\_knaflitz,

title={A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait},

author={Bonato, Paolo and D'Alessio, Tommaso and Knaflitz, Marco},

journal={IEEE Transactions on biomedical engineering},

volume={45},

number={3},

pages={287--299},

year={1998},

publisher={IEEE}

}

@article{agostini\_1,

title={Normative EMG activation patterns of school-age children during gait},

author={Agostini, Valentina and Nascimbeni, A and Gaffuri, A and Imazio, P and Benedetti, MG and Knaflitz, M},

journal={Gait \& posture},

volume={32},

number={3},

pages={285--289},

year={2010},

publisher={Elsevier}

}

@article{SGA,

title={Statistical gait analysis},

author={Agostini, V and Knaflitz, M and Acharya, RU and Molinari, F and Tamura, T and Naidu, DS and Suri, JS},

journal={Distributed diagnosis and home healthcare (D2H2)},

volume={2},

pages={99--121},

year={2012},

publisher={American Scientific Publishers Stevendon Ranch, CA, USA}

}

@article{cimap1,

title={Muscle activation patterns during gait: A hierarchical clustering analysis},

author={Rosati, Samanta and Agostini, Valentina and Knaflitz, Marco and Balestra, Gabriella},

journal={Biomedical Signal Processing and Control},

volume={31},

pages={463--469},

year={2017},

publisher={Elsevier}

}

@inproceedings{cimap2,

title={Muscle contractions in cyclic movements: Optimization of CIMAP algorithm},

author={Rosati, Samanta and Castagneri, Cristina and Agostini, Valentina and Knaflitz, Marco and Balestra, Gabriella},

booktitle={2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)},

pages={58--61},

year={2017},

organization={IEEE}

}

@book{clustering,

title={Pattern classification and scene analysis},

author={Duda, Richard O and Hart, Peter E and Stork, David G},

volume={3},

year={1973},

publisher={Wiley New York}

}