2021-22 NCTS Undergraduate Research Project

Title: Data Science with Application to Sleep Stage Automatic Scoring

Supervisor: Gi-Ren Liu (NCKU)

Students:

邱能泰Neng-Tai Qiu (NYCU, School of Medicine)

郝嘉誠 Jia-Cheng Hao (NYCU, Department of Electrical Engineering)

蘇彥維 Yan-Wei Su (NYCU, Department of Applied Mathematic)

曾以諾 Yi-Nuo Zeng (NCKU Department of Mathematics)

Background and Motivation:

Sleep is a critical factor contributing to physical and mental health, particularly in fatigue recovery and memory consolidation. Absence of sleep can cause different diseases such as obesity, heart related diseases, diabetes and as a result reduces life expectancy. Correct cognition of sleeping is helpful for diagnosis of varies disease. The purpose of this project is to design algorithms for sleep dynamics visualization and automatic annotation by applying time-frequency analysis and manifold learning to extract features from various physiological signals. We will introduce the short-time Fourier transform, wavelet transform, spectral clustering, and the diffusion maps. From this project, students can understand how to apply the Fourier analysis, linear algebra and Markov chain to analyze real data of interest.

Plan:

2021/10-2021/11, Stage 1: Processing data by Matlab

The students will study how to read medical data through the software Matlab and apply the scattering transform to extract features from them.

2021/11-2022/1 Stage 2: Dimension Reduction

We plan to study how to reduce the dimension of the features by the principle component analysis, canonical correlation analysis, spectral clustering and diffusion maps.

2022/03-2022/06, Stage 3: Project-oriented learning and research

We plan to use the scattering transform, Synchrosqueezing transform, dimension reduction methods, and the support vector machine to design an automatic sleep stage classification system. Apart from the electroencephalography (EEG) signals, we will also explore the possibility of using other signals, including the respiratory signal and the photoplethysmography (PPG) signals, to carry out accurate inference for sleep stages.

Reference:

[1] Bandeira, A. S., Singer, A. &, Strohmer, T. *Mathematics of Data Science* https://people.math.ethz.ch/~abandeira/BandeiraSingerStrohmer-MDS-draft.pdf

[2] Vershynin, R. (2018). *High-dimensional Probability: An Introduction with Applications in Data Science* (Vol. 47). Cambridge university press.

[3] Panchenko, D. (2019). *Introduction to Probability Theory*.

[4] Blum, A., Hopcroft, J., & Kannan, R. (2020). *Foundations of Data Science*. Cambridge University Press.

[5] Mallat, S. (1999). A Wavelet Tour of Signal Processing. Elsevier.

[6] Liu, G. R., Lo, Y. L., Malik, J., Sheu, Y. C., & Wu, H. T. (2020). Diffuse to fuse EEG spectra–intrinsic geometry of sleep dynamics for classification. *Biomedical Signal Processing and Control*, **55**, 101576.