Interaction Information and its usage in Genome Wide Interaction Studies

Jan Mielniczuk

Institute of Computer Science, Polish Academy of Sciences

miel@ipipan.waw.pl, https://home.ipipan.waw.pl/j.mielniczuk/

A vital part of genetic research in humans is devoted to detection of epistasis or significant gene-gene interactions and quantification of their strength in occurrence of diseases (Genome Wide Interaction Studies, cf. e.g. Cordell (2002)). Interaction information is one of important non-parametric tools used to this aim and is widely applied in interaction detection systems (cf. e.g. Wan et al (2010)). Results obtained recently (Kubkowski i Mielniczuk (2020) can be applied to construct new more precise methods which would take into account dependence of genes. The aim of the project is to study properties of such methods, in particular their control of False Discovery Rate and to apply them in practical situations. Such analysis will involve methods of statistical inference and information theory as well as data mining methods for the applications' part. Moreover, usage of Interaction Information in selection of important genes based on Möbius approximation will be studied extending Mielniczuk and Teisseyre (2019). The aim here is to detect active genes and their active interactions in a high-dimensional framework.

References

[1] H. Cordell (2002), *Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans ,* Human Molecular Genetics. 111, 2463-2468

[2] X. Wan et al (2010), *BOOST: A fast approach to detecting Gene-Gene interactions in Genome-wide Case-Control Studies,* American Journal of Human Genetics. 87, 325-340

[3] M. Kubkowski, J. Mielniczuk (2020), *Asymptotic distributions of empirical Interaction Information,* Methodology and Computing in Applied Probability

[4] J. Mielniczuk, P. Teisseyre (2019), *Stopping rules for mutual information-based feature selection,* Neurocomputing, 358, 255-274