

**Doctoral School of Information and Biomedical Technologies
Polish Academy of Sciences (TIB PAN)**

SUBJECT: Variational approach to learning from incomplete data

SUPERVISOR: Jan Mielniczuk, professor, Institute of Computer Science,
PAS

DESCRIPTION:

Inference for incomplete data is an important research area in Machine Learning and Statistics due to ubiquity of such data in practice. Important insight is provided by variational approach which considers a lower bound for interesting parameter, pertaining to a family of functions and then constructs the estimator as the maximiser or maximal value of the bound in this family. As such, the method is ideally suited for deep learning using as the family of functions specific family of neural networks. Such approach is successfully used e.g. for estimation of mutual information as well for posterior probability.

The research proposal focuses on the application of such method for posterior probability estimation for incomplete data such as positive unlabelled (PU) data and missing data and studying the properties of resulting classifiers. The starting point is application of Varadhan-Donsker representation and its task. The outcoming methods will be applied for text/image annotation and in recommender systems. The project builds upon former extensive research experience concerning inference for Positive Unlabelled data of the supervisor.

Candidate should have M.Sc. in Mathematics, Computer Science or Engineering, be knowledgeable in Machine Learning and Statistics, including both its mathematical and computational aspects, and possess sufficient computing skill to effectively implement and analyse proposed methods. Scientific curiosity and eagerness to learn are essential.

Candidate should contact the author of the proposal before formal submission of documents (miel@ipipan.waw.pl).

BIBLIOGRAPHY:

- [1] H. Chen, F. Liu, Y. Wang, L. Zhao (2020). , A variational approach for learning from positive and Unlabeled data, NIPS 2020
- [2] Belghazi, M. I., Baratin, A., Rajeswar, S., Ozair, S., Bengio, Y., Courville, A., and Hjelm, R. D. (2019). Mine: Mutual information neural estimation, ICML 2018
- [3] M.D. Donsker, S. R. Varadhan (1983). Asymptotic evaluation of certain Markov process expectations for large time. *Communications on Pure and Applied Mathematics*, 36(2):183–212.