Learning Theory for Uplift Modelling

Szymon Jaroszewicz
Institute of Computer Science PAS, Warsaw
sj@ipipan.waw.pl

1. Project Description

Uplift modelling is a branch of machine learning which aims to model the effect of some action (medical treatment, marketing campaign) on specific individuals. This is achieved by using two sets of data in the learning process: experimental, with cases subjected to the action and control with cases on which the action was not performed, serving as a background. The task is then to predict the difference in class probabilities between the two groups. Popularity of uplift modelling is growing in marketing and medicine but there are few theoretical results available.

The proposal is to develop rudiments of learning theory for the uplift modelling problem. The questions we would answer are: when is effective learning possible, what role does VC-dimension play in uplift modelling, how many training samples do we need to provide specific guarantees on model quality. The knowledge we gain in the process would also be applied to development of new uplift modelling algorithms.

2. Requirements

    a. M.Sc. of Computer Science, Mathematics or Physics

    b. Knowledge of probability and statistics

    c. Programming skills

    d. Good command of English

References

1. S. Jaroszewicz, Uplift Modeling. In: Encyclopedia of Machine Learning and Data Mining, Springer, 2016

2. Ł. Zaniewicz, S. Jaroszewicz, Lp-Support vector machines for uplift modeling, Knowledge and Information Systems, http://link.springer.com/content/pdf/10.1007/s10115-017-1040-6.pdf