

Adaptive divergence for rapid adversarial optimization & $(1 + \epsilon)$ -class Classification: an Anomaly Detection Method for Highly Imbalanced or Incomplete Data Sets

Wednesday 21 October 2020 12:20 (20 minutes)

This talk contains 2 contributions:

1) Adaptive divergence for rapid adversarial optimization.

Adversarial Optimization provides a reliable, practical way to match two implicitly defined distributions, one of which is typically represented by a sample of real data, and the other is represented by a parameterized generator. Matching of the distributions is achieved by minimizing a divergence between these distributions, and estimation of the divergence involves a secondary optimization task, which, typically, requires training a model to discriminate between these distributions. The choice of the model has its trade-off: high-capacity models provide good estimations of the divergence, but, generally, require large sample sizes to be properly trained. In contrast, low-capacity models tend to require fewer samples for training; however, they might provide biased estimations. Computational costs of Adversarial Optimization becomes significant when sampling from the generator is expensive. One of the practical examples of such settings is fine-tuning parameters of complex computer simulations. In this work, we introduce a novel family of divergences that enables faster optimization convergence measured by the number of samples drawn from the generator. The variation of the underlying discriminator model capacity during optimization leads to a significant speed-up. The proposed divergence family suggests using low-capacity models to compare distant distributions (typically, at early optimization steps), and the capacity gradually grows as the distributions become closer to each other. Thus, it allows for a significant acceleration of the initial stages of optimization. This acceleration was demonstrated on two fine-tuning problems involving Pythia event generator and two of the most popular black-box optimization algorithms: Bayesian Optimization and Variational Optimization. Experiments show that, given the same budget, adaptive divergences yield results up to an order of magnitude closer to the optimum than Jensen-Shannon divergence. While we consider physics-related simulations, adaptive divergences can be applied to any stochastic simulation.

2) $(1 + \epsilon)$ -class Classification: an Anomaly Detection Method for Highly Imbalanced or Incomplete Data Sets

Anomaly detection is not an easy problem since distribution of anomalous samples is unknown a priori. We explore a novel method that gives a trade-off possibility between one-class and two-class approaches, and leads to a better performance on anomaly detection problems with small or non-representative anomalous samples. The method is evaluated using several data sets and compared to a set of conventional one-class and two-class approaches.

Primary authors: BORISYAK, Maxim (Yandex School of Data Analysis (RU)); GAINTEVA, Tatiana (Yandex School of Data Analysis (RU)); USTYUZHANIN, Andrey (Yandex School of Data Analysis (RU))

Presenter: BORISYAK, Maxim (Yandex School of Data Analysis (RU))

Session Classification: Workshop

Track Classification: 5 ML algorithms : Machine Learning development across applications