M.Sc. internship (stage) project proposal
2011-2012

Designing lean classifiers for detectors and triggers

Advisor	Balázs Kégl
Laboratory	Linear Accelerator Laboratory (LAL) and Computer Science Laboratory (LRI) University Paris-Sud/CNRS/INRIA
Research group	Machine Learning and Applied Statistics (AppStat/LAL) and Machine Learning and Optimization (TAO/LRI)
Address	University Paris-Sud, 91898 Orsay Cedex
Mail	balazs.kegl@gmail.com
Web	http://users.web.lal.in2p3.fr/kegl
Keywords	particle physics, machine learning
Required background	statistics or machine learning or physics

Summary

There are numerous applications where the **computational requirements of classifying a test instance** are as important as the performance of the classifier itself. **Object detection** in images [1] and **web page ranking** [2] are well-known examples. A more recent application domain with similar requirements is **trigger design in high energy physics** [3]. Most of these applications come with another common feature: the negative class (usually called noise or background) have sometimes orders of magnitudes higher probability than the positive class. Beside the testing time constraints, this also makes training difficult: traditional classification-error-based measures are not adequate, and using prior class probabilities in constructing training samples leads to either enormous data sizes or little representativity of the positive class.

A common solution to these problems is to design **cascade classifiers**. A cascade classifier consists of *stages*. In each stage a binary classifier attempts to eliminate background instances by classifying them negatively. Positive classification in inner stages sends the instance to the next stage, so detection can only be made in the last stage. By using simple and fast classifiers in the first stages, "easy" background instances can be rejected fast, shortening the expected testing time. Classical cascade classifiers (such as the seminal Viola-Jones (VJ) algorithm [1]) have many disadvantages, succinctly summarized in [4]. We have recently proposed to overcome these problems by **casting the cascade design problem in a Markov Decision Process** framework [5]. The goal of the project is to adapt this approach to trigger design is experimental physics. Our main testbed for the algorithm is the **trigger of the LHCb** experiment. LHCb is one of the experiments installed on the Large Hadron Collider (LHC), the particle accelerator ring of 27 km in circumference at the CERN Laboratory in Switzerland. Its objective is to help to explain the **matter-antimatter asymmetry** of the Universe. The experiment is a ideal testbed for our proposed automatic trigger design: on the one hand, classifiers trained by AdaBoost are parts of the existing trigger chain, and on the other hand, the six different measurements carried out by the detector, similarly to multi-view object detection in images, would require slightly different classifiers, making the classical cascade design suboptimal.

The ideal candidate has a strong background either in computer science, statistics, or physics, and an open mind to acquire the necessary knowledge in the other two disciplines. The project can be modulated between these three angles based on the candidate's profile. The project can be followed by a Ph.D. thesis.

References

[1] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.

[2] O. Chapelle and Y. Chang, "Yahoo! learning to rank challenge overview," in *Yahoo Learning to Rank Challenge (JMLR W&CP)*, vol. 14, (Haifa, Israel), pp. 1–24, 2010.

[3] V. Gligorov, "A single track HLT1 trigger," Tech. Rep. LHCb-PUB-2011-003, CERN, 2011.

[4] L. Bourdev and J. Brandt, "Robust object detection via soft cascade," in *Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 236–243, IEEE Computer Society, 2005.

[5] D. Benbouzid, R. Busa-Fekete, and B. Kégl, "MDDAG: designing sparse decision DAGs using Markov decision processes," in *submitted*, 2011.