Advanced data mining

This METEOR builds on the previous ***Data mining*** *METEOR and provide a more in-depth course on data mining with more advanced methods such as matrix factorization and an introduction to neural networks that is one of the approach of choice in modern applications. Data mining and machine learning have become central to numerous data processing studies in the recent years and this METEOR aim at providing advanced knowledge that can be used not only in astronomy but also in a more general professional environment.*

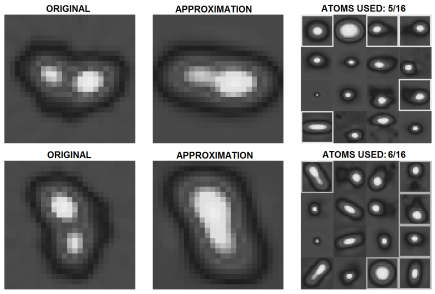
Theory

by C. RICHARD, R. FLAMARY

The theoretical part of the METEOR contains two main parts: data mining methods and data mining tools (optimization).

The methods that will be introduced are first dictionary learning and matrix factorization methods that are used in numerous applications among such as recommendation systems and clustering of observed objects. They have also been used recently to model the images of different families of astronomical objects. Another approach that will be introduced are neural networks and in particular their design for different learning tasks such as object recognition or image processing. Those methods have leads in recent year to tremendous results in artificial intelligence such as automatic image tagging and the Google AlphaGO whose level is now equivalent or better than humans.

Next, an introduction to large scale optimization techniques will be provided. All data mining method require the resolution of an optimization problem that can become very large on astronomical datasets. In the METEOR we will introduce recent approaches based on proximal methods that aim a promoting sparsity and interpretability in the solution. Those methods are commonly used nowadays for image reconstruction in astronomy. The use of distributed optimization algorithms that allow a global optimization problem to be solved on local machines/datasets will also be discussed.



Applications

by R. FLAMARY, C. RICHARD

Applications of the introduced method is central to the teaching of data mining. We will provided numerous practical sessions where the students will learn to use the Python language for data mining and its power-full data mining libraries such as scikit-learn in Python. Specific applications that will be implemented are dictionary learning on astronomical images for modeling and denoising, large scale spectral unmixing and identification and parallel code for running on high performance computers. At the end of the METEOR the student will have a practical experience with several machine learning methods and will be able to estimate efficiently data mining models on large datasets such as survey data. They will also have advanced machine learning and data mining skills for optimization, and data interpretation and denoising.

See also

- MAUCA courses FC1.6 (Signal/image processing) and FC2.2 (Statistical methods), Data Mining.
Introduction to data mining website.
- Data mining in astronomy
- Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data.
- AstroML Machine learning for Astronomy

Contact

☎ +33-4-92-07-63-94 (supervisor)
✉ cedric.richard@unice.fr
✉ remi.flamary@unice.fr