

Large scale online Collaborative Filtering

In collaborative filtering the data consists typically of a collection of ratings $(x_{ij})_{(i,j) \in D}$, where x_{ij} is the rating of item j by user i ; and D being a small subset of all user \times item pairs. In that formulation collaborative filtering is a matrix completion problem in which the goal is to reconstruct the remaining entries of $X = (x_{ij})_{ij}$.

A classical approach to collaborative filtering is to assume that X is low rank and to find the best low rank matrix that fits the data. This is formulated as a non-convex optimization problem that can be written as a matrix factorization problem. Specifically, X is then approximated by a product UV^T of matrices U and V with respectively few columns and few rows. A now classical convex relaxation of the best low rank approximation can be obtained using the so-called trace-norm [Abernethy et al., 2009].

Whether one considers the factorization formulation or its relaxation, these optimization problems are challenging to solve for very large matrices (millions of users, ten of thousands of items, and millions of known ratings). Several algorithms have been proposed recently to solve the corresponding optimization problems for such large matrices using *stochastic* and *distributed* optimization algorithms [Teflioudi et al., 2012, Recht and Ré, 2011].

The goal of this internship is to propose a new *stochastic* and possibly *distributed* algorithm to tackle the matrix factorization problem or other variants for matrix completion.

Student interested by the topic should contact **Guillaume Obozinski** to discuss further.

The project of the internship can potentially be extended into a PhD in the Imagine group, at Ecole des Ponts et Chaussées.

Pre-Requisites : Convex optimization, core machine learning classes.

Length : 4-6 months

Contact : Guillaume Obozinski (Guillaume.Obozinski@ens.fr)

<http://www.di.ens.fr/~obozinski/>

Laboratory : Equipe IMAGINE, École des Ponts ParisTech,

6, Avenue Blaise Pascal – Cité Descartes, Champs-sur-Marne, 77455 Marne-la-Vallée cedex 2 - France.

<http://imagine.enpc.fr/index.html>

Références

[Abernethy et al., 2009] Abernethy, J., Bach, F., Evgeniou, T., and Vert, J. (2009). A new approach to collaborative filtering : Operator estimation with spectral regularization. *The Journal of Machine Learning Research*, 10 :803–826.

[Recht and Ré, 2011] Recht, B. and Ré, C. (2011). Parallel stochastic gradient algorithms for large-scale matrix completion. *Optimization Online*.

[Teflioudi et al., 2012] Teflioudi, C., Makari, F., and Gemulla, R. (2012). Distributed matrix completion. *ICDM*.