A New Approach to Collaborative Filtering: Operator Estimation with Spectral Regularization

Jacob Abernethy (UC Berkeley, USA)
Francis Bach (Ecole normale supérieure, France)
Theodoros Evgeniou (INSEAD, France)
Jean-Philippe Vert (Mines ParisTech, France)*

Jean-Philippe.Vert@ensmp.fr
Centre for Computational Biology
35 rue Saint-Honoré
77300 Fontainebleau, France

Collaborative filtering (CF) refers to the task of predicting preferences of a given "user" for some "objects" (e.g., books, music, products, people, etc...) based on his/her previously revealed preferences – typically in the form of ratings – as well as the revealed preferences of other users. In a book recommender system, for example, one would like to suggest new books to someone based on what he and other users have recently purchased or rated. The ultimate goal of CF is to infer the preferences of users for new to them objects.

Recently there has been interest in CF using regularization based methods [2]. These methods assume that the only data available are the revealed preferences, and no other information such as background information on the objects or users is given. In this case, one may formulate the problem as that of filling a matrix with users as rows, objects (e.g., books) as columns, and missing entries as currently unknown preferences. To make useful predictions within this setting, regularization based CF methods make certain assumptions about the *relatedness* of the objects and users. The most common assumption is that preferences can be decomposed into a small number of factors, both for users and objects, resulting in the search for a low-rank matrix which approximates the partially observed matrix of preferences [2]. The rank constraint can be interpreted as a regularization on the hypothesis space. Since the rank constraint gives rise to a non-convex set of matrices, the associated optimization problem will be a difficult non-convex problem for which only heuristic algorithms exist [2]. An alternative formulation, proposed by [3], suggests penalizing the predicted matrix by its *trace norm*, i.e., the sum of its singular values. An added benefit of the trace norm regularization is that, with a sufficiently large regularization parameter, the final solution will be low-rank [1].

However, a key limitation of current regularization based CF methods is that they do not take advantage of information, such as attributes of users (e.g., gender, age) or objects (e.g., book's author, genre), which is often available. Intuitively, such information might be useful to guide the inference of preferences, in particular for users and objects with very few known ratings. For example, at the extreme, users and objects with no prior ratings can not be considered in the standard CF formulation, while their attributes alone could provide some basic preference inference.

*JPV will present the paper at the conference.

Here we develop a general framework and specific algorithms for the more general CF setting where other information, such as attributes for users and/or objects, may be available. We consider that the spaces of users and objects are Hilbert spaces, either defined trivially when no attribute is considered, or defined by a positive definite kernel when attributes are available. We then show that CF, typically seen as a problem of matrix completion, can be thought of more generally as estimating a *linear operator* from the space of users to the space of objects. Equivalently, this can be viewed as learning a bilinear form between users and objects. We then develop *spectral regularization* based methods to learn such linear operators. Spectral regularization is obtained by constraining a penalty function for compact linear operators that only depends on the singular values of the operator. In the classical matrix completion case, the singulat values of the operator are simply the singular values of the matrix and spectral penalty include the Frobenius or trace norm, as well as low rank constraints. When dealing with general operators, rather than matrices, one may also work with infinite dimensions, allowing to consider arbitrary feature spaces induced by some kernels.

A key theoretical contribution is that we prove a new representer theorem which show that an operator which minimizes a empirical risk function subject to a spectral penalty constraint can be found in a finite-dimensional space spanned by the training points. This theorem generalizes to any spectral penalty the classical representer theorem for tensor product kernels, which corresponds to a particular spectral penalty: the Hilbert-Schmidt norm. It allows us to develop new methods that learn finitely many parameters also in the general case of infinite dimesional feature spaces that describe users or objects.

We also show that, with the appropriate choice of kernels for both users and objects, we may consider a number of existing machine learning methods as special cases of our general framework. In particular, we show that several CF methods, such as rank constraint, trace-norm regularization, and frobenius norm regularization based ones, can all be cast as special cases of spectral regularization on operator spaces. Moreover, particular choices of kernels lead to specific subcases such as regular matrix completion and *multitask learning*. In the specific application of collaborative filtering with the presence of atttributes, we show that our generalization of these subcases leads to better predictive performance, both on toy examples and on a database of movies.

References

[1] M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proc. American Control Conference*, volume 6, 2001.

[2] N. Srebro and T. Jaakkola. Weighted low-rank approximations. In T. Fawcett and N. Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning*, pages 720–727. AAAI Press, 2003.

[3] N. Srebro, J. D. M. Rennie, and T. S. Jaakkola. Maximum-margin matrix factorization. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Adv. Neural. Inform. Process Syst. 17*, pages 1329–1336, Cambridge, MA, 2005. MIT Press.

Topic: learning algorithm.
Preference: oral.