Matrix completion problem

Project Proposal

Team name: ROY
SkolTech, November 2016

Team

- **[Team Leader]** Artem Bochkarev
- Ilya Zharikov
- Roman Isachenko
- Anne-Laure Ducrocq

Background

The problem of matrix completion is newly emerged and rapidly growing field of numerical linear algebra. This field addresses a broad range of problems of significant practical interest, namely, the recovery of a data matrix from what appears to be incomplete, and perhaps even corrupted, information. In its simplest form, the problem is to recover a matrix from a small sample of its entries. This problem occurs in many applications, when it's very inefficient to store huge matrices or we want to recover some signal. For example, matrix completion is very useful in collaborative filtering, recommendation systems, dimensionality reduction and computer vision.

We can solve such problem if the rank of the matrix is low, compared to its dimension. There are several known approaches in this field, which exploit the that fact, for example singular value thresholding and nuclear norm minimization.

Problem formulation

Let denote $\mathbf{M} \in \mathbb{R}^{n \times m}$ is a matrix we would like to know as precisely as possible. However, the only information available about $\mathbf{M}$ is a sampled set of entries M_{ij}, $(i, j) \in \Omega$ where Ω is a subset of the complete set of entries $\{1, \dots, n\} \times \{1, \dots, m\}$. It will be convenient to summarize the information available via $P_\Omega(\mathbf{M})$, where the sampling operator $P_\Omega : \mathbb{R}^{n \times m} \to \mathbb{R}^{n \times m}$ is defined by

$$(P_\Omega(\mathbf{X}))_{ij} = \begin{cases} X_{ij}, & (i, j) \in \Omega; \\ 0, & \text{otherwise.} \end{cases}$$

Our goal is to find the matrix X by solving the following task. If the number of measurements is sufficiently large, and if the entries are sufficiently uniformly distributed, it is possible that there is only one low-rank matrix with these entries. And in this case we came to the next problem:

$$\begin{aligned} \text{minimize} \quad & \operatorname{rank}(\mathbf{X}) \\ \text{subject to} \quad & P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}). \end{aligned}$$

There are many approaches to solve this problem and relaxations of the problem that we also want to consider.

Data

We are considering several options here, as the problem of matrix completion can be used in many applications. First, we can evaluate our algorithm on purely synthetic low-rank matrices. After that we plan to test the approach on some known dataset from public repository. This data might vary from preferences matrix to euclidean distances between cities.

Related work

Usually we want to minimize nuclear norm in order to solve this problem. Several approaches for doing that were developed, such as singular value thresholding algorithm [1], nuclear norm penalization [2] and convex relaxation [3]. Also methods for completion of matrices with noise [4] and with use of alternating minimization [5] were developed. One of the main applications of listed approaches is collaborative filtering [6].

Scope

First, we will get familiar with a problem itself. We will study fundamental and recently published works in this field. After that, we are going to implement some of the existing algorithms as a baseline and then play with its parameters. We will also try to implement more efficient optimization methods in our algorithms.

Evaluation

How do you measure the performance of your solution? It is also a good place for including baselines.

Our main quality metric is root mean squared error

$$\mathrm{RMSE}(R, \hat{R}) = \sqrt{\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(r_{i,j} - r_{i,j})^2},$$

where $\hat{R}$ is an estimator of R.

References

[1] *Jian-Feng Cai, Emmanuel J. Candes, Zuowei Shen.* A Singular Value Thresholding Algorithm for Matrix Completion. // *SIAM Journal on Optimization*, 20(4):1956–1982. 2010.

[2] *Vladimir Koltchinskii, Karim Lounici, Alexandre B. Tsybakov.* Nuclear norm penalization and optimal rates for noisy low rank matrix completion. // *The Annals of Statistics*, 2302–2329. 2011.

[3] *Emmanuel J. Candes and Terence Tao.* The Power of Convex Relaxation: Near-Optimal Matrix Completion. // *IEEE Transactions on Information Theory*, 56(5):2053–2080. 2010.

[4] *Emmanuel J. Candes, Yaniv Plan.* Matrix Completion with Noise. // *Proceedings of the IEEE*, 98(6):925–936. 2010.

[5] *Prateek Jain, Praneeth Netrapalli, Sujay Sanghavi.* Low-rank Matrix Completion using Alternating Minimization. // *Proceedings of the 45-th annual ACM symposium on Theory of computing*, 665–674. 2013.

[6] *Gabor Takacs, Istvan Pilaszy, Bottyan Nemeth, Domonkos Tikk.* Scalable Collaborative Filtering Approaches for Large Recommender Systems to clustering. // *Journal of machine learning research*, 10(3):623–656. 2009.