One-Bit Matrix Completion for Pairwise Comparison Matrices

Andrew K. Massimino and Mark A. Davenport
Georgia Institute of Technology
Email: {massimino,mdav}@gatech.edu

In this paper we consider the related problems of ranking and of recovering a matrix of pairwise comparisons from binary observations. In the context of ranking, our objective is to create an ordered list of n items based solely on comparisons between two items at a time. The items could be consumer products, sports teams, tweets or online forum comments, candidate drug therapies, job applications, academic papers or proposals, or any other collection of objects or courses of action. We will assume that each item can be assigned a numerical ranking (or score) $r_i \in \mathbb{R}$. Item i is preferred to item j if $r_i > r_j$. These rankings taken together form the rating vector $\boldsymbol{r} \in \mathbb{R}^n$. Next consider the $n \times n$ matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ in which each element M_{ij} represents the difference between the ratings for items i and j, i.e., $M_{ij} = r_i - r_j$. Equivalently, we can write $\boldsymbol{M} = \boldsymbol{r}\boldsymbol{e}^T - \boldsymbol{e}\boldsymbol{r}^T$, where $\boldsymbol{e} = [1, 1, \dots, 1]^T$.

Clearly, given the matrix $\boldsymbol{M}$ we could directly recover the vector $\boldsymbol{r}$ (up to a constant shift), giving us a rank-ordered list of all items. However, in practice we rarely have access to the matrix $\boldsymbol{M}$. Obtaining this matrix would require comparisons between every possible pair of items, which can be prohibitive in practice for even moderate sized n. For example, in constructing a ranking of sports teams, we can rarely expect every team to play every other possible team. Moreover, in most settings the outcomes of these comparisons will be extremely noisy and are often highly "quantized"—we may only observe which team wins, and have no strict assurance that the winning team was in fact the superior one. Similar problems arise whenever the comparisons are the result of human judgements, which can be notoriously unreliable and may only tell us which of two items was preferred. In this paper we consider the question of whether the underlying matrix $\boldsymbol{M}$ can be recovered from such noisy one-bit observations when the observations are potentially highly incomplete. We approach the problem from the perspective of *one-bit matrix completion*, which asserts that since the matrix $\boldsymbol{M}$ is low-rank, we should be able to recover $\boldsymbol{M}$ accurately via a simple convex program [1].

Specifically, suppose that $\boldsymbol{Y}$ denotes a complete set of noisy, binary comparisons given by

$$Y_{ij} = \{ +1 \text{ w.p. } f(M_{ij});\ -1 \text{ w.p. } 1 - f(M_{ij}) \},$$

where $f(M_{ij}) = \mathbb{P}\{\text{item } i \text{ is preferred to item } j\}$. The choice of f determines the chance that the outcomes are "upsets" with respect to the underlying ratings. In this paper we will consider the case where f is given by the standard logistic function $f(x) = (1 + e^{-x})^{-1}$, but other natural choices are also possible (see [1]). In the one-bit matrix completion setting, we assume that we are able to observe $\boldsymbol{Y}$ on a subset of indices indexed by Ω. In this case, if we let Ω^+ denote the subset of Ω where $Y_{ij} = +1$ and similarly for Ω^-, the log-likelihood function is given by

$$F_\Omega(\boldsymbol{X}|\boldsymbol{Y}) = \sum_{(i,j)\in\Omega^+} \log f(X_{ij}) + \sum_{(i,j)\in\Omega^-} \log f(1 - X_{ij}).$$

If the underlying matrix $\boldsymbol{M}$ has rank s and $\|\boldsymbol{M}\|_\infty \le \alpha$, then the approach advocated in [1] is to set

$$\widehat{\boldsymbol{M}} = \arg\max_{\boldsymbol{X}} F_\Omega(\boldsymbol{X}|\boldsymbol{Y}) \text{ s.t. } \|\boldsymbol{X}\|_\star \le \alpha n \sqrt{s} \text{ and } \|\boldsymbol{X}\|_\infty \le \alpha.$$

In the case where $\boldsymbol{M} = \boldsymbol{r}\boldsymbol{e}^T - \boldsymbol{e}\boldsymbol{r}^T$, we have that $s = 2$, and so the upper bounds on $\|\boldsymbol{M} - \widehat{\boldsymbol{M}}\|_F$ in [1] apply and we can accurately recover $\boldsymbol{M}$.

However, this naïve application of generic matrix completion techniques ignores the fact that in this specific context, the matrix $\boldsymbol{M}$ has considerable additional structure beyond having low rank, and it may be possible to leverage this structure to obtain improved results. For example, if $\boldsymbol{M} = \boldsymbol{r}\boldsymbol{e}^T - \boldsymbol{e}\boldsymbol{r}^T$ then $\boldsymbol{M}$ is skew-symmetric. In [2] this skew-symmetry is added as an additional constraint to a nuclear norm minimization program to obtain a more accurate recovery from an incomplete set of (unquantized) aggregated ratings. In this paper we consider an even more direct path to leveraging this structure. Specifically, we can replace the likelihood function with

$$F_\Omega(\boldsymbol{x}|\boldsymbol{Y}) = \sum_{(i,j)\in\Omega^+} \log f(x_i - x_j) + \sum_{(i,j)\in\Omega^-} \log f(1 - (x_i - x_j)),$$

and consider an optimization problem of the form

$$\widehat{\boldsymbol{r}} = \arg\max_{\boldsymbol{x}} F_\Omega(\boldsymbol{x}|\boldsymbol{Y}) \text{ s.t. } x_i \ge 0 \text{ for all } i, \text{ and } \|\boldsymbol{x}\|_1 \le \rho.$$

If desired, we can then obtain $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{r}}\boldsymbol{e}^T - \boldsymbol{e}\widehat{\boldsymbol{r}}^T$. In this paper we show how to adapt the techniques in [1] to obtain upper bounds on $\|\boldsymbol{M} - \widehat{\boldsymbol{M}}\|_F$ that significantly improve upon the bounds available for the naïve adaptation of one-bit matrix completion.

One limitation of the approach described above is that it is only possible to record one observation per item pair. However, it is easy to modify the objective function to allow for multiple comparisons between the same pair of items. Given this modification, we show that up to a set of monotonic transformations in $\boldsymbol{r}$, our model is actually equivalent to the classic *Bradley-Terry-Luce* model, which has been studied extensively in the context of ranking. From this perspective, one can view our proposed algorithm as only a slight variant of a classic approach described in [3], with the difference being in the particular form of our constraints on $\boldsymbol{r}$. We thus provide new insight into this traditional approach to ranking.

To evaluate our approach, we include a suite of numerical simulations that compare the performance of our proposed approach with the naïve adaptation of one-bit matrix completion on a synthetic example where the true underlying ratings are known. We find that our proposed approach leads to significant performance gains.

References

[1] M. Davenport, Y. Plan, E. van den Berg, and M. Wootters, "1-bit matrix completion," *Arxiv preprint arxiv:1209.3672*, 2012.

[2] D. Gleich and L.-H. Lim, "Rank aggregation via nuclear norm minimization," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Diego, CA, Aug. 2011.

[3] L. Ford, "Solution of a ranking problem from binary comparisons," *Amer. Math. Monthly*, vol. 64, no. 8, pp. 28–33, 1957.