
Robust matrix completion for rating-scale data

Aurore Archimbaud^{*1}, Andreas Alfons , and Ines Wilms

¹Toulouse Business School – Department of Information, Operations and Management Sciences – France

Résumé

Ces dernières années, la complétion de matrices de rang faible (“Low Rank Matrix Completion” - LRMC) a suscité un intérêt considérable. Cette approche vise à prédire les entrées inconnues d’une matrice partiellement observée en utilisant ses éléments connus ainsi qu’une contrainte de rang faible, tout en minimisant un critère spécifique tel que l’erreur quadratique moyenne. Bien que les applications courantes comportent des données d’échelle de notation discrète, comme le célèbre concours du Prix Netflix sur les systèmes de recommandation, les méthodes de type LRMC sont presque toujours conçues et étudiées dans le contexte de données continues. De ce fait, les propriétés statistiques de ces méthodes sur des données discrètes ont été très peu investiguées. De plus, bien qu’il existe de nombreux travaux sur la LRMC avec des données continues, seul un sous-ensemble relativement restreint de la littérature a envisagé la complétion matricielle en présence d’observations corrompues. Pourtant, il est très courant de se retrouver en présence de ce type d’observations lorsque l’on s’intéresse à l’évaluation de produits par des utilisateurs. Ce phénomène peut se produire dû à un manque d’attention de la personne mais aussi résulter d’une intention délibérée de celle-ci. Ces utilisateurs malveillants manipulent les évaluations afin d’influencer le système de recommandation à leur avantage dans le cadre de ce que l’on appelle des attaques. Pour combler ces lacunes, nous présentons un nouvel algorithme LRMC conçu spécifiquement pour des données d’échelle de notation discrète et robuste à la présence d’observations corrompues. De plus, nous évaluons les performances de notre méthode avec plusieurs approches concurrentes à travers une étude basée sur des simulations de données d’échelle de notation discrètes au lieu de données continues et en considérant divers scénarios d’attaques.

Mots-Clés: Rang faible, Robustesse, Données discrètes, Données creuses, Décomposition en valeurs singulières, Attaques, Systèmes de recommandation

^{*}Intervenant