Fiche de présentation

Maximum Margin Matrix factorization http://people.csail.mit.edu/jrennie/papers/icml05-mmmf.pdf

Nathan Srebro - Jason D. M. Rennie - Tommi S. Jaakkola

Résumé

Cet article présente une nouvelle approche pour le filtrage collaboratif. L'approche est très inspirée par la technique de discrimination linéaire à large marge. Les auteurs s'intéressent en particulier à l'adaptation de méthodes de réduction dimensionnelle linéaire des matrices *sparse*.

Introduction

Les auteurs rappelent l'état de l'art dans le cas de mimisation de fonction de coût, en particulier :

- Modèle probabilistique spécifique comme pLSA (T.Hofmann 2001, M.Collins) ou les paramètres naturels (S.Dasgupta et R.Schapire 2002) : problèmes avec multiples minimum locaux.
- pour le filtrage collaboratif, les techniques d'optimisation par SVD (Nathan Srebro and Tommi Jaakkola 2003) ne sont pas applicables
- Données : matrice Y contenant des appréciations (notations) d'utilisateurs sur des films.
- cas de matrices avec des données manquantes
- Objectif : prédire des entrées non-observées de la matrice Y basées sur un ensemble d'entrées observées Ys

Section 2

L'auteur introduit la notion de réduction dimensionnelle linéaire des matrices creuses (*sparse*) pour le filtrage collaboratif.

- On cherche à factoriser la matrice Y sous la forme $X = UV^{'}, U \in R^{n*k}, V \in R^{m*k}$
- Dans le filtrage collaboratif, U (vecteurs de caractéristiques) et V sont inconnus et doivent être estimés.
- On cherche la prédiction collaborative par *low-rank*, ce qui correspond à limiter la dimension de l'espace : chaque colonne est une prédiction linéaire dans un espace de dimension réduite
- avec une fonction coût appropriée (ou contraintes sur les entrées observées), cela correspond à une discrimnination linéaire à large marge (comme SVM).

Section 3

L'auteur introduit la Factorisation de Matrices à Marge Maximun ou MMMF et les notions de la factorisation de matrices à marge dure et floue. ou *MMMF*

- Simplification du problème : on utilise des labels binaires, $Y \in \pm 1^{n*m}$.
- Marge dure : recherche d'un minimum de la norme de la trace de la matrice, qui correspondent aux labels observés avec $X_{ia}Y_{ia} \geq 1$ pour $ia \in S$
- Marge molle : minimisation de la norme de la trace de X et l'erreur relative à Ys : minimise $\parallel X \parallel_{\sum} + c \sum max(0.1 - Y_{ia}X_{ia})$
- Il y a dépendance inverse entre la norme et la marge : minimiser la trace équivaut à maximiser la marge
- Prédire la matrice cible avec les signes du rang k de la matrice revient à projeter les "items" et les "utilisateurs" vers des hyperplans (*hard-margin low-max-norm prediction*). Chaque hyperplan sépare les items positifs des négatifs (*large marge*). Les notes des utilisateurs sont des combinaisons dun petit ensemble de facteurs communs.

Section 4

Dans cette section, les auteurs introduisent l'optimisation des problèmes d'apprentissage par MMMF

- Hypothèse : matrice non-négative
- Hypothèse : on peut reformuler la norme de la trace comme un problème semi-définit (SDP).
- Dans un problème typique de prédiction collaboratif, on observe une petite fraction des entrées de la matrice cible.

Prédiction pour des nouveaux utilisateurs :

- On considère que l'apprentissage est fait sur toutes les entrées de chaque ligne
- Il est confortable de pouvoir prédire des entrées dans une ligne non observée sur l'apprentissge initial (nouvel utilisateur)
- L'essentiel étant de résoudre un problème où V est connu et un nouveau U est calculé (prédiction pour le nouvel utilisateur), basé sur nouvelle observation de X
- Utiliser MMMF est un problème standard SVM

Section 5
A quel degrès peut-on être confiant pour prédir les entrées Y par rapport au Ys observés en terme d'erreurs?

- Les auteurs montrent que leur méthode est généralisable

Section 6 Implémentations et expérimentations
Les auteurs présentent des tests et résultats. Données tests : 100 utilisateurs sur un ensemble de 100K notes utilisteurs de la base de données MovieLens

- Comparaison entre 2 algorithmes comme référence/base : WLRA (weighted low rank approximation) et K-Medians
- 4 ensembles de données aléatoirement disposées
- Pour chacun des 4 sous-ensembles possibles, calcul de *3-fold cross-validation* sur le reste des données.
- Pour chacun des 4 sous-ensembles, on teste 2 MMMF et algoritmes références avec (*ZOEzero/one level-agreement error*) et MAE (erreur absolue moyenne) puis on mesure l'erreur avec les données tests

Discussion

- L'apprentissage avec MMMF est résolu avec une optimisation SDP.
- L'apprentissage est possible sur des dizaines de miliers de labels.
- Limitation : les entrées observées doivent être uniformement échantillonées. Hypothèse peu réaliste.
- Méthode généralisable

Commentaires
Les auteurs proposent dans cet article une nouvelle méthode pour le filtrage collaboratif (FC) intéressante. Elle permet de répondre au challenge de cas de matrices avec des données manquantes. De plus, d'après les auteurs, leur méthode est généraisable. De mon point de vue, cette méthode est très intéressante car elle apporte une réelle avancée dans un domaine récent (FC). Par contre, cette méthode demande une phase de calcul off-line qui peut-être importante. Enfin, d'une manière générale, la recherche dans ce domaine s'oriente vers l'utilisation de méthodes basées sur la réduction dimensionnelle de matrices creuses (parse). Cet article est intéressant et soigneusement détaillé.