Maximum Margin Matrix Factorization

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

Fabien Lepercque - Fouille de données

March 7, 2007

Plan

(1) Introduction

(2) Contexte

(3) Factorisations de Matrices à Marge Maximun (MMMF)

(3) Résultats

(3) Conclusion

Objectifs

- Nouvelle approche pour le filtrage collaboratif
- Inspiré par la technique de discrimination linéaire à large marge (SVM)
- Cas de matrices avec des données manquantes
- Pour le filtrage collaboratif, les techniques d'optimisation par SVD (Nathan Srebro and Tommi Jaakkola 2003) ne sont pas applicables
- Objectif : prédire des entrées non-observées de la matrice Y basé sur un ensemble d'entrées observées Ys

Données d'entrées

movies

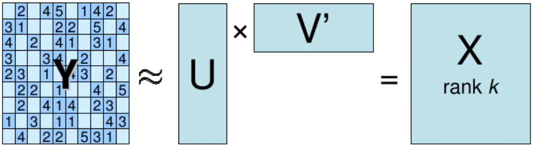
users

	2		1			4				5	
	5		4				?		1		3
		3		5			2				
4			?			5		3		?	
		4		1	3				5		
			2				1	?			4
	1					5		5		4	
		2		?	5		?		4		
	3		3		1		5		2		1
	3				1			2		3	
	4			5	1			3			
		3				3	?			5	
2	?		1		1						
		5			2	?		4		4	
	1		3		1	5		4		5	
1		2			4				5	?	

Données d'entrées

- On cherche a 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.



- Chaque colonne est une prédiction linéaire dans un espace de dimension réduite

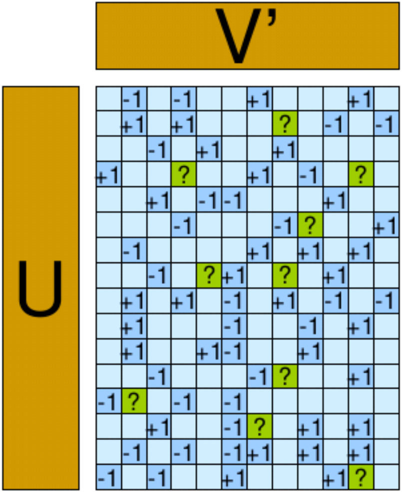
MMMF

- 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 = 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 la fonction coût relative à Ys : minimise $\parallel X \parallel_{\sum} + c \sum max(0.1 - Y_{ia} X_{ia})$

Optimisation

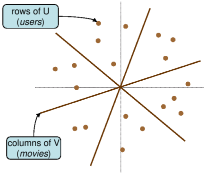
- Hypothèse: Matrices non négative
- Hypothèse: on peut reformuler la norme de la trace comme un problème semi-définit (SDP).
- Dans un problème typique de prédiction collaboratif, on observe une petite fraction des entrées de la matrice cible.

Prédiction pour des nouveaux utilisateurs :



Prédiction pour des nouveaux utilisateurs :

- Objectif : prédire des entrées dans une ligne non observée sur l'apprentissage initial (nouvel utilisateur)
- Prédire la matrice cible avec les signes du rang k de la matrice revient à projeter les “items” et les “utilisateurs” vers des hyperplans
- Chaque hyperplan sépare les items positifs des négatifs (*large marge*).



Expérimentations

- Données tests : 100 utilisateurs sur un ensemble de 100K (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 (*zero/one level-agreement error*) et MAE (erreur absolue moyenne) puis on mesure l'erreur avec les données tests

Résultats

Test Set	Method	ZOE CV	ZOE Test	Method	MAE CV	MAE Test
1	WLRA rank 2	0.547	0.575	K-Medians K=2	0.678	0.691
2	WLRA rank 2	0.550	0.562	K-Medians K=2	0.686	**0.681**
3	WLRA rank 1	0.562	0.543	K-Medians K=2	0.700	0.681
4	WLRA rank 2	0.557	0.553	K-Medians K=2	0.685	0.696
Avg.			0.558			0.687
1	max-norm C=0.0012	0.543	**0.562**	max-norm C=0.0012	0.669	**0.677**
2	trace norm C=0.24	0.550	**0.552**	max-norm C=0.0011	0.675	0.683
3	max-norm C=0.0012	0.551	**0.527**	max-norm C=0.0012	0.668	**0.646**
4	max-norm C=0.0012	0.544	**0.550**	max-norm C=0.0012	0.667	**0.686**
Avg.			**0.548**			**0.673**

Conclusion

- Le calcul de 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 peut réaliste
- Méthode généralisable