

ClasSel : revue de mi parcours

15 septembre 2010

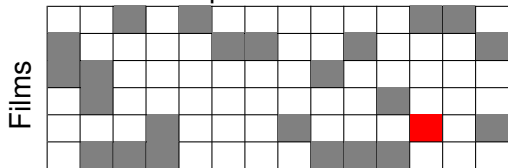


Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes

D. Fourdrinier
Université de Rouen, LITIS EA 4108

ClasSel en une diapositive

spectateurs



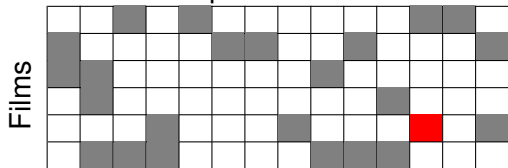
- ▶ 480 000 spectateurs
- ▶ 18 000 films
- ▶ 100 millions de notes
- ▶ matrice creuse (1,16 % $\neq 0$)

- ▶ Formation de groupes spectateurs-films
→ Classification croisée (modèles de mélanges)
- ▶ Nombre de groupes pertinents
→ sélection de modèle (estimation de coût)

- ▶ Classification croisée → HEUDIASYC
- ▶ Sélection de modèle → LITIS
- ▶ Plateforme de validation → CRIP5 → LIPADE
- ▶ Etudes de cas - démonstrateurs → NumSight

ClasSel en une diapositive

spectateurs



- ▶ 480 000 spectateurs
- ▶ 18 000 films
- ▶ 100 millions de notes
- ▶ matrice creuse (1,16 % $\neq 0$)

- ▶ Formation de groupes spectateurs-films
→ Classification croisée (modèles de mélanges)
- ▶ Nombre de groupes pertinents
→ sélection de modèle (estimation de coût)

- ▶ Classification croisée → HEUDIASYC
- ▶ Sélection de modèle → LITIS
- ▶ Plateforme de validation → CRIP5 → LIPADE
- ▶ Etudes de cas - démonstrateurs → Numsight

Organisation de la présentation

1. L'état du projet D. Fourdrinier (coordinateur)
2. tâche 1 : Classification croisée M. Nadif
3. tâche 2 : Sélection de modèle A Boisbunon
4. tâche 3 : Plateforme de validation S. Canu
5. tâche 4 : Les problèmes et les données T. Couronne
6. Le futur du projet D. Fourdrinier (coordinateur)

B Livrables et jalons

État	N°	Titre	Nature	Partenaire	Comment.
Livr. sept. 09	1	Web	Logiciel	HEUDIASYC	
Au fil du temps	2	Gestion	Rapport	LITIS	T0+6, T0+18
Livr. sept. 09	3	1.1	Rapport	HEUDIASYC, CRIP5	
Livr. juil. 09	4	1.2	Rapport	HEUDIASYC, CRIP5	
Rept. déc. 10	5	1.3	Rapport	LITIS	en retard
Prev. mars 11	6	1.4	Rapport	HEUDIASYC, CRIP5	en cours
Rept. déc. 10	7	1.5	Rapport	HEUDIASYC, CRIP5	en cours
Livr. juin 09	8	2.1	Rapport	LITIS	
Livr. sept. 10	9	2.2	Rapport	HEUDIASYC, CRIP5	
Livr. juin. 11	10	2.3	Rapport	LITIS	v. préliminaire
Livr. sept. 10	11	3.1	Rapport	LITIS	
Livr. mars 11	12	3.2	Rapport	HEUDIASYC, CRIP5	v. préliminaire

 livrables terminés

Perspectives

1. Poursuite du plan prévu et traitement des points reportés.
2. Organisation d'un Workshop à Paris le 4 mars 2011 avec conférencier invité (Iven Van Michelen, Université de Leuven).
3. Organisation d'une session dans une conférence internationale.

C.2 Résultats marquants

Arrivée de Numsight Consulting

C.3 Réunions et rendez-vous importants

- Réunions trimestrielles
- Organisation d'un Workshop le 04/03/2011

C.4 Commentaires et questions

Comment intégrer administrativement Numsight dans notre projet ?

D Impact du projet

D.1 Indicateurs d'impact :

8 publications : 3 multipartenaires et 5 monopartenaires

D.2 Listes des publications et communications

- ▶ 1 revue (Journal of Multivariate Analysis)
- ▶ 2 chapitres de livre
- ▶ 3 conférences internationales (ECML, ERCIM et ICMLA)
- ▶ 2 conférences nationales (Cap et les journées de stat.)

1. Blouza, A., Fourdrinier, D. and Lepelletier, P. A Green's formula in a weak framework: an application in statistics, Volume en l'honneur du professeur Strawderman W. E. édité par Springer, 2010 (accepté)
2. Fourdrinier, D. and Marchand, É. On Bayes estimators with uniform priors on spheres and their comparative performance with maximum likelihood estimators for estimating bounded multivariate normal means. *Journal of Multivariate Analysis*, 101, 1390-1399, 2010.
3. Fourdrinier, D. and Strawderman, W. E. Robust generalized Bayes minimax estimators of location vectors for spherically symmetric distribution with unknown scale, Volume en l'honneur du professeur Brown L., 2010 (accepté).
4. G. Govaert, M. Nadif. Un modèle de mélange pour la classification croisée d'un tableau de données continues. CAP'09, pp 287-302, Hammamet, Tunisie, 2009.
5. B. Hanczar, M. Nadif. Bagging for Biclustering: Application to Microarray Data. ECML/PKDD, pp 490-505, 2010.
6. M. Nadif, G. Govaert. Model-based Co-clustering for continuous data. ICMLA, Washington D.C, 12-14 december, 2010.
7. Ch. Keribin, G. Govaert, Gilles Celeux. Estimation d'un modèle à blocs latents par l'algorithme SEM. 42èmes journées de statistique, Marseille, 2010.
8. M. Nadif, G. Govaert. Different variants of model-based co-clustering of continuous data. 2nd Workshop of the ERCIM group on Computing and Statistics, Limassol, October 2009.

D4 Personnels recrutés en CDD (hors stagiaires)

- ▶ Aurélie Boisbunon, Doctorante, LITIS (30 mois).
- ▶ Grégoire Mesnil, Doctorant, LITIS (2 mois).
- ▶ Rudolphe Priam, Ingénieur Recherche, CRIP5 (3 mois).
- ▶ Lazhar Labiod, Post-doc, CRIP5 (9 mois).
- ▶ Fida El Baf, Post-doc, Heudyasic (9 mois).

D.5 État financier

Nom du partenaire	Crédits consommés	Commentaires
Université de Rouen, LITIS	95 %	1 salaire CDD d'ores et déjà provisionné
Université Paris 5, CRIP5	25 %	2 stagiaires, 1 post-doc, Workshop
Université de Technologie de Compiègne	52 %	1 stagiaire, 1 doctorante sur bourse ministère

La suite de ClasSel : travaux futur

- ▶ Poursuite du plan prévu et traitement des points reportés.
- ▶ Organisation d'un Workshop à Paris le 4 mars 2011 avec conférencier invité (Iven Van Michelen, Université de Leuven).
- ▶ Boite à outil matlab
- ▶ traitement des études de cas : les problèmes et les données de Numsight