❑ COLLABORATION

INSTITUT DE MATHEMATIQUES DE TOULOUSE

LABORATOIRE D'INFORMATIQUE, DE TRAITEMENT DE L'INFORMATION ET DES SYSTEMES, INSA ROUEN

LABORATOIRE LAGRANGE, UNIVERSITE DE NICE SOPHIA ANTIPOLIS

❑ RELATED DOCUMENTS

Léa Laporte, Rémi Flamary, Stéphane Canu, Sébastien Déjean et Josiane Mothe. Non-convex Regularizations for Feature Selection in Ranking with Sparse SVM. IEEE Transactions on Neural Networks and Learning Systems, Vol. 25 N. 6, p. 1118-1130, 2014. https://www.irit.fr/publis/SIG/2014_OTAO_LLFRCSDSMJ.pdf -

Léa Laporte, Sébastien Déjean et Josiane Mothe. *Séparateurs à Vaste Marge pondérés en norme L2 pour la sélection de variables en apprentissage d'ordonnancement*. Conférence en Recherche d'Information et Applications (CORIA), 19/03/2014 – 21/03/2014, Nancy.

Léa Laporte. *La sélection de variables en apprentissage d'ordonnancement pour la recherche d'information : vers une approche contextuelle*. Rapport de thèse, 2013.

❑ PARTICIPANTS

LEA LAPORTE
REMI FLAMARY
STEPHANE CANU
SEBASTIEN DEJEAN
JOSIANE MOTHE

❑ CONTACTS

J. MOTHE
Institut de Recherche en Informatique de Toulouse
Equipe Systèmes d'Informations Généralisées
Campus Université Paul Sabatier
118, Route de Narbonne – 31062 Toulouse Cedex 4
Email : josiane.mothe@irit.fr,
Tél. : (33/0) 561 55 64 44 Fax.: (33/0) 561 55 62 58

INSTITUT DE RECHERCHE EN INFORMATIQUE DE TOULOUSE

INSTITUT DE MATHEMATIQUES DE TOULOUSE

TEAM SYSTÈMES D'INFORMATIONS GÉNÉRALISÉES – J. MOTHE

SELECTION DE VARIABLES EN APPRENTISSAGE D'ORDONNANCEMENT VIA DES SVM PARCIMONIEUX ET DES REGULARISATIONS NON CONVEXES.

L'ALGORITHME RANKSVM-NC.

☐ Skill ☑ Prototype ☐ Product

UMR 5505 CNRS-INP-UPS 118 ROUTE DE NARBONNE 31062 TOULOUSE CEDEX 4

❑ FEATURE SELECTION FOR RANKING IN INFORMATION RETRIEVAL

In the field of Information Retrieval, learning to rank aims at automatically optimizing the ranking of the documents returned by a retrieval system. Learning to rank algorithms learn ranking functions that combine several IR models in order to rank documents according to their relevance to a query. The scores obtained by IR models are then features used by ranking functions. As the number of features can be very large, ranking algorithms have to deal with large scale data. This amount of data causes computational issues, such as an increase of computational costs. Feature selection is used to solve this issue.

❑ REGULARIZED SVM FOR FEATURE SELECTION IN RANKING

In the field of machine learning, when dealing with linear methods such as linear SVMs, a quite natural way to perform feature selection is to consider an L0-regularized problem, as the L0 norm is the number of nonzero coefficient in the linear model. Unfortunately, the L0 norm is nonconvex, not continuous and not differentiable : the optimization problem is NP-hard. One solution is to consider an L1-regularized problem, which is a convex relaxation of the L0-norm problem. Nevertheless, as L1-norm is not differentiable, quadratic optimization can not be used. Moreover, L1-norm can be biaised. Nonconvex sparse regularizations such as Log, Lp or MCP are then preferred. In our work, we solve the nonconvex regularizations problems by iteratively reweighting L1-regularized SVM for ranking.

❑ RANKSVM-NC : NONCONVEX REGULRIZATIONS FOR FEATURE SELECTION BY REWEIGHTIN L1-REGULARIZED SVM

RankSVM-NC performs feature selection for ranking with SVMs by using nonconvex regularization. A first algorithm, that we called RankSVM-L1, solve the L1-regularized ranking problem by using a Forward-Backward Splitting approach. The second algorithm, that we called RankSVM-NC, approximate the solutions of the optimization problem that use nonconvex regularizations by iteratively reweighting the L1-norm problem solved by RankSVM-L1.

MODULES

– *Feature selection and ranking on IR benchmarks* : this script allow the user to specify the regularization trem and the dataset from LETOR corpora he wants to use. Results files are generated, that contains the ranking model and the IR metrics (MAP, AP, NDCG) for train, validation and test datasets in each run. The script called the RankSVM-NC algorithm, that called also RankSVM-L1 (t*h*e FBS algorithm implemented by Remi Flamary).

DEVELOPPEMENTS

- The scrips are implemented in Matlab, but can by run either on Matlab or Octave. They incorporated an FBS algorithm implemented by Rémi Flamary. This algorithm is under Gnu Public Licence and all the other scripts heritated of this licence.

APPLICATIONS

- *Feature selection in ranking on documents corpora (LETOR benchmarks, etc)*

- *Feature selection and ranking on other kind of data*