Ranking SVM on LETOR

Introduction to Ranking SVM

The basic idea of Ranking SVM is to formalize learning to rank as a problem of binary classification on instance pairs, and then to solve the problem using Support Vector Machines.

The details of Ranking SVM can be found from http://www.research.microsoft.com/~rherb/papers/herobergrae99.ps.gz and http://svmlight.joachims.org/.

Learning Parameters

We use the package of SVMlight in our experiments. Here we used linear model. We tuned -c parameter, and the best value of the parameter selected by validation set is shown as follows. To reduce training time, we set -# 5000 for SVMlight.

Know issues

- We used version 6.0.1 of SVMlight. Different versions of SVMlight may output slightly different models with same parameters and datasets. Thank Ming-Feng Tsai for pointing out this.
- We used 64 bit version of SVMlight for TD2004 datasets. Thank Keping Bi for pointing out this.
- The selected parameter for the 5th fold of NP2004 is 0.0009 instead of 0.0005. Thank Keping Bi for pointing out this.

Dataset	-c (from Fold1 to Fold5)
OHSUMED	0.01, 0.001, 0.0001, 0.6, 0.1
TD2003	0.009, 0.003, 0.0005, 0.0007, 0.07
TD2004	0.05, 0.01, 4, 0.03, 0.9
HP2003	0.1, 0.01, 0.00003, 0.0002, 0.001
HP2004	3, 0.0009, 0.5, 30, 0.00009
NP2003	0.002, 1, 0.00005, 0.01, 0.01
NP2004	0.0003, 70, 0.09, 0.0007, 0.0005 0.0009

Papers & Docs

R. Herbrich, T. Graepel, and K. Obermayer. Large Margin Rank Boundaries for Ordinal Regression. Advances in Large Margin Classifiers, 115-132, Liu Press, 2000.

T. Joachims. Optimizing Search Engines Using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2002.

BibTex
@inbook{citeulike:477598,
   author = {Herbrich, Ralf   and Graepel, Thore   and Obermayer, Klaus },
   booktitle = {Advances in Large Margin Classifiers},
   citeulike-article-id = {477598},
   editor = {Smola and Bartlett and Schoelkopf and Schuurmans},
   keywords = {ml_for_ir, reranking, svm, um},
   priority = {2},
   publisher = {MIT Press, Cambridge, MA},
   title = {Large margin rank boundaries for ordinal regression},
   url = {http://citeseer.ist.psu.edu/contextsummary/1891774/0},
   year = {2000}
}

@inproceedings{775067,
   author = {Thorsten Joachims},
   title = {Optimizing search engines using clickthrough data}

```
title = {Optimizing search engines using clickthrough data},
  booktitle = {KDD '02: Proceedings of the eighth ACM SIGKDD international conference on
Knowledge discovery and data mining},
  year = {2002},
  isbn = {1-58113-567-X},
  pages = {133--142},
  location = {Edmonton, Alberta, Canada},
  doi = {http://doi.acm.org/10.1145/775047.775067},
  publisher = {ACM},
  address = {New York, NY, USA},
}
```

Notes

This document was written by Jun Xu, and the experiments were conducted by Chaoliang Zhong. If any problem, please contact letor@microsoft.com