Supplement Materials for "Large-scale Kernel RankSVM"

I Another Evaluation Criterion

We further consider mean average precision (MAP), which is the mean of average precision (AP) of each query.

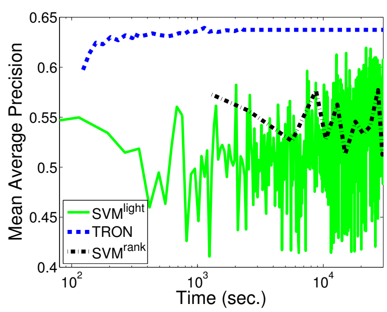
$$\text{Mean Average Precision} \equiv \frac{\sum_{q \in S} \text{AP}(q)}{|S|}$$

Average precision is the average of precisions at all possible recall levels.

$$\text{AP}(q) \equiv \sum\nolimits_{m=1}^{l_q} \text{Precision@}m \cdot \Delta r(m)$$

where $\Delta r(m)$ is the change in recall from position $m-1$ to m.

To calculate the precision, a relevance level that is less than the medium of all relevance levels is considered to be negative and others are considered to be positive. Figures (I) and (II) show the experimental results.



(a) Mean average precision

Figure (I): Experimental results on MQ2007

II Best Parameters for Each Method and Each Criterion

In Section 4, we mentioned that a grid search was conducted to decide the parameters to be used. The best parameters found for each method and each criterion are listed in Tables (III)-(V).

III Comparison of Performances Between Different Formulations

Because each method compared in our experiments considers different optimization problems, we conduct more examinations on their performances using other commonly considered evaluation criteria. Here we consider precision@5 and mean reciprocal rank. For a given query q, its mean reciprocal rank is defined as follows.

$$\text{Mean reciprocal rank} \equiv \frac{1}{l_q} \sum_{i=1}^{l_q} \frac{1}{|\{j \mid y_{\bar{\pi}(j)} > y_{\bar{\pi}(\pi^{-1}(i))}\}|}.$$

We report the average of this value among all $q \in S$.

Our purpose here is to compare the performance of different formulations in spite of the optimization methods. Thus, after selecting the best parameters for each criterion, we run each package for a long enough training time (min(24 hours, the training time to fulfill the default stopping condition)) to obtain stable results. The best parameters found for each method and each criterion are listed in Tables (VI)-(VII). The results are reported in Tables (VIII) and (IX). As mentioned earlier, the package RV-SVM failed to run on MQ2007

Data set	TRON	SVM^{light}	SVM^{rank}	RV-SVM
MQ2007	$(-4,-4)$	$(-2,-5)$	$(-1,-4)$	NA
MQ2008	$(1,-6)$	$(-1,-5)$	$(-4,3)$	$(4,-2)$
MQ2007-list	$(-4,-5)$	$(1,-5)$	$(-5,-6)$	NA

Table (III): Best parameter $(\log_2 C, \log_2 \gamma)$ for each method on pairwise accuracy.

Data set	TRON	SVM^{light}	SVM^{rank}	RV-SVM
MQ2007	$(-2,-5)$	$(-2,-4)$	$(-1,-4)$	NA
MQ2008	$(3,-5)$	$(3,-6)$	$(-4,-6)$	$(0,-2)$

Table (IV): Best parameter $(\log_2 C, \log_2 \gamma)$ for each method on mean NDCG. Note that mean NDCG is not available for MQ2007-list because the overflow of $2^{\bar{\pi}(i)}$ caused by the large k.

Data set	TRON	SVM^{light}	SVM^{rank}	RV-SVM
MQ2007	$(-4,-4)$	$(1,-6)$	$(3,-2)$	NA
MQ2008	$(1,-4)$	$(3,-6)$	$(-4,-6)$	$(-4,-2)$

Table (V): Best parameter $(\log_2 C, \log_2 \gamma)$ for each method on mean average precision. Note that we follow LETOR 4.0 not to consider MAP on MQ2007-list.

and MQ2007-list because the required memory is beyond our machine capacity.

From the results, we can see that the optimization problem considered in our approach is comparable to or better than state of the art. This indicates that our method does not sacrifice the final performance to obtain a faster training time.

Data set	TRON	SVM^{light}	SVM^{rank}	RV-SVM
MQ2007	$(-1,-4)$	$(0,-5)$	$(-3,-2)$	NA
MQ2008	$(0,-3)$	$(-3,-2)$	$(-5,-5)$	$(0,-1)$

Table (VI): Best parameter $(\log_2 C, \log_2 \gamma)$ for each method on precision@5. Note that we follow LETOR 4.0 not to consider precision@5 on MQ2007-list.

Data set	TRON	SVM^{light}	SVM^{rank}	RV-SVM
MQ2007	$(-2,-4)$	$(2,-5)$	$(3,-6)$	NA
MQ2008	$(1,-4)$	$(4,-6)$	$(-2,-5)$	$(-5,-2)$
MQ2007-list	$(5,-4)$	$(1,-5)$	$(-5,-3)$	NA

Table (VII): Best parameter $(\log_2 C, \log_2 \gamma)$ for each method on mean reciprocal rank.

Data set	TRON	SVM^{light}	SVM^{rank}	RV-SVM
MQ2007	0.3154	0.3125	0.2262	NA
MQ2008	0.3974	0.4103	0.4038	0.4038
MQ2007-list	0.8095	0.4613	0.7054	NA

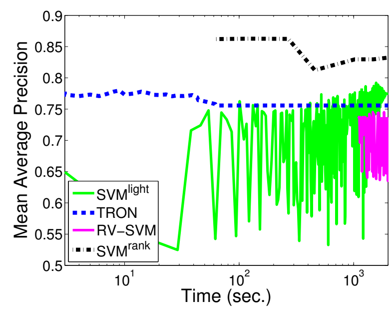
Table (VIII): Comparison of precision@5 between different formulations. We run each package long enough to exclude the difference resulted from optimization methods.

IV The Relation Between Data Size and Training Time

To check the performance of the proposed method when the number of instances grows, we sub-sampled different number of instances from MQ2007, and report the training time required. The experimental result is shown in Figure (III). The X-axis is the number of instances, and the Y-axis is the training time required. Note that both axes are log-scaled, and the results show that the proposed method remains efficient as the number of instances grows.

Data set	TRON	SVM^{light}	SVM^{rank}	RV-SVM
MQ2007	0.7503	0.7436	0.6629	NA
MQ2008	0.8127	0.8045	0.7966	0.8147
MQ2007-list	0.9579	0.8457	0.9300	NA

Table (IX): Comparison of mean reciprocal rank between different formulations. We run each package long enough to exclude the difference resulted from optimization methods.



(a) Mean average precision

Figure (II): Experimental results on MQ2008

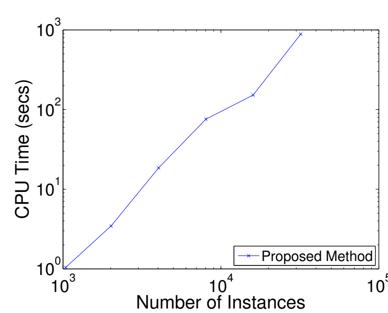


Figure (III): Experimental results on the relation between data size and training time.