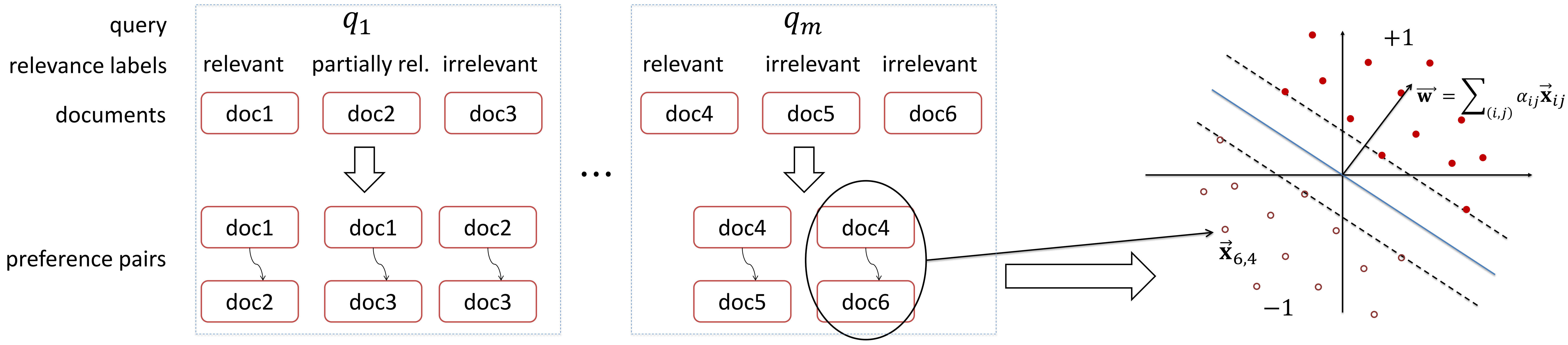
Modeling Parameter Interactions in Ranking SVM

Yaogong Zhang[1], Jun Xu[2], Yanyan Lan[2], Jiafeng Guo[2], Maoqiang Xie[1], Yalou Huang[1], Xueqi Cheng[2]
[1]College of Computer and Control Engineering, Nankai University
[2]Institute of Computing Technology, Chinese Academy of Sciences

1. Pairwise learning to rank: ranking as binary classification over preference pairs



Motivation: There exist significant interactions among the training pairs, e.g., (doc1, doc2) and (doc1, doc3) share doc1. Whether there also exist interactions among model parameters? How to utilize the interactions if the answer is yes?

2. Parameter interactions in Ranking SVM

Ranking SVM

Prime

$$\min_{\mathbf{w}\in R^n} \frac{1}{2}||\mathbf{w}||^2 + C \sum_{(i,j)\in P} \left[1 - \langle \mathbf{w}, \mathbf{x}_i - \mathbf{x}_j \rangle\right]_+$$

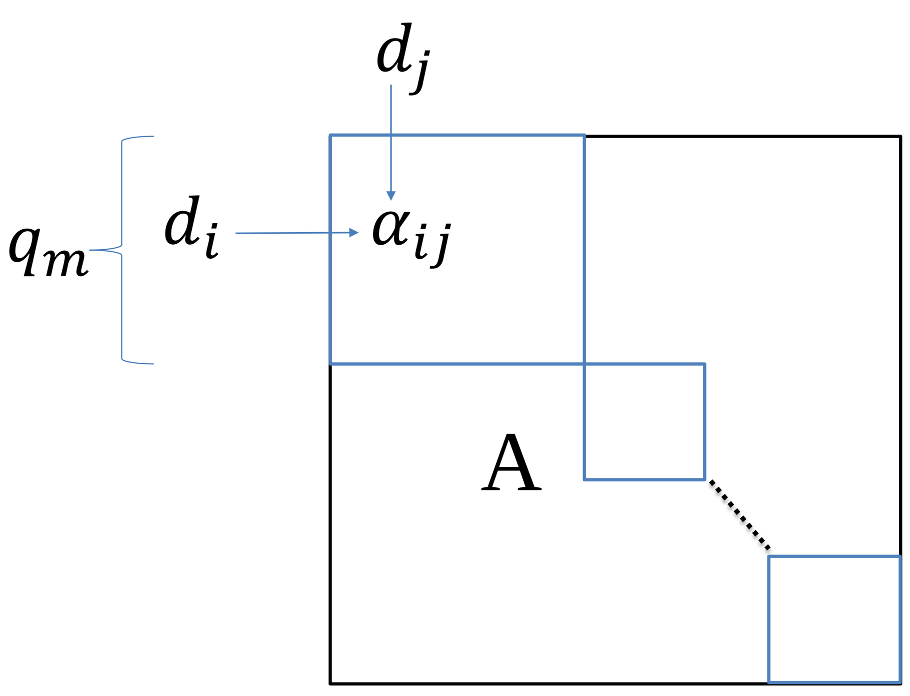
α_{ij} corresponds to preference pair (i,j)

Dual

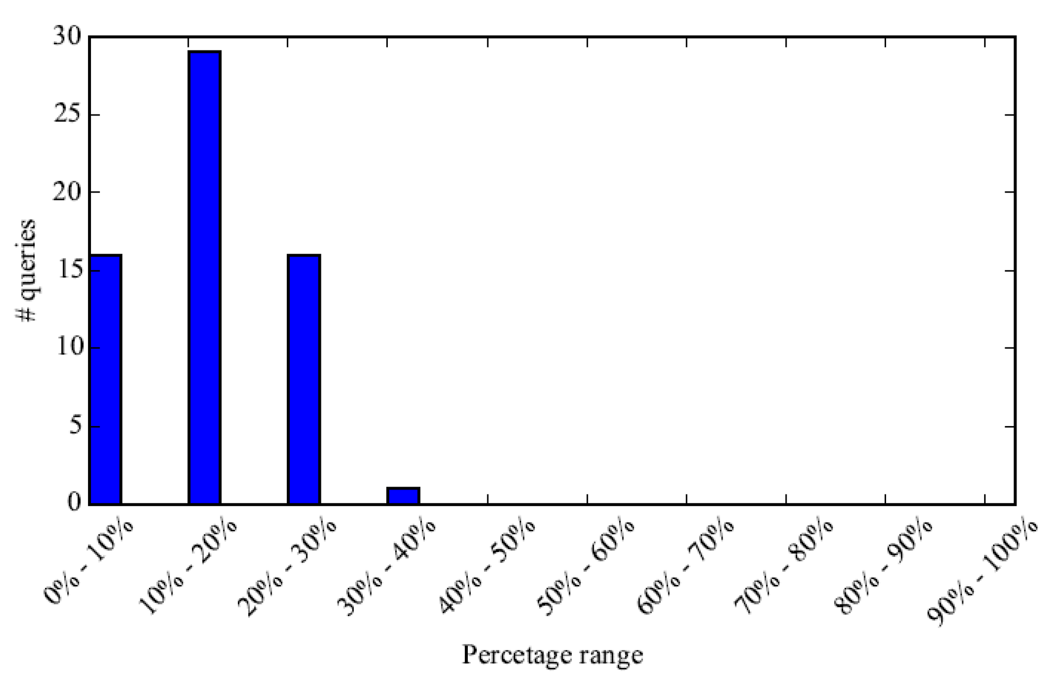
$$\min_{\alpha} \frac{1}{2}\boldsymbol{\alpha}^T \mathrm{M} \boldsymbol{\alpha} - \boldsymbol{e}^T \boldsymbol{\alpha}$$

$$s.t.\quad 0 \le \alpha_{ij} \le C, \forall (i,j) \in P$$

Low rank structure among Lagrange multipliers α_{ij}



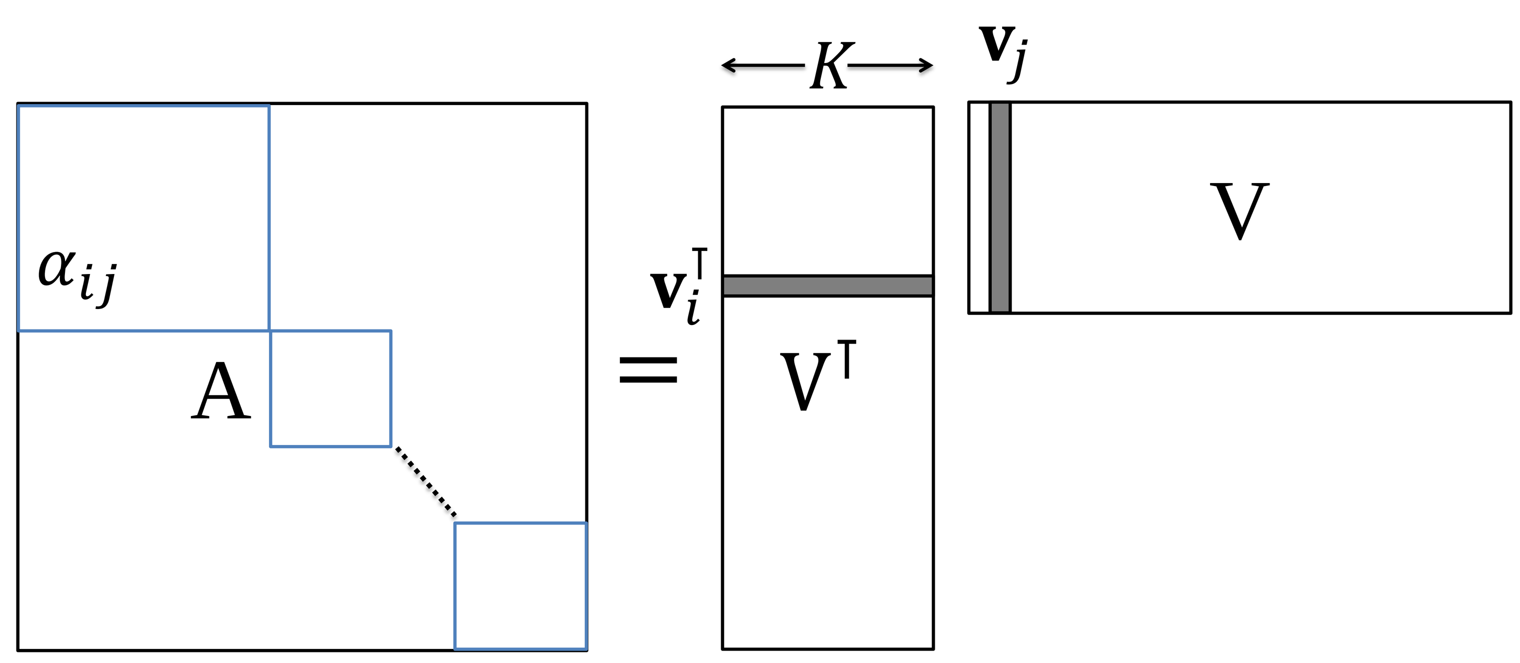
$$A(i,j) = \begin{cases} \alpha_{ij} & (i,j) \in P \\ 0 & \text{otherwise.} \end{cases}$$



(a) Capturing 90% energy

3. Factorized Ranking SVM

Directly modeling the low rank structure: $\alpha_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$



New loss function

new parameters

$$\min_{\mathbf{v}_1,\cdots,\mathbf{v}_N} \frac{1}{2}\left\|\sum_{(i,j)\in P} \langle \mathbf{v}_i, \mathbf{v}_j \rangle (\mathbf{x}_i - \mathbf{x}_j)\right\|^2 + C\sum_{(k,l)\in P}\left[1 - \left\langle \sum_{(i,j)\in P} \langle \mathbf{v}_i, \mathbf{v}_j \rangle (\mathbf{x}_i - \mathbf{x}_j), \mathbf{x}_k - \mathbf{x}_l \right\rangle\right]_+$$

Number of parameters: $O(N^2) \rightarrow O(KN)$

4. Experiments

Results on OHSUMED (dense preference pairs)

	MAP	NDCG@1	NDCG@3	NDCG@5
RSVM	0.4427	0.5289	0.4553	0.4392
RankNet	0.404	0.4007	0.3616	0.3388
ListNet	0.4443	0.5134	0.4664	0.4530
Fac-RSVM	**0.4463**	**0.5507**	**0.4798**	**0.4546**

Results on MQ2008 (sparse preference pairs)

	MAP	NDCG@1	NDCG@3	NDCG@5
RSVM	0.4713	**0.3686**	0.4277	0.4730
RankNet	0.4522	0.3414	0.3991	0.4500
ListNet	0.4415	0.3244	0.3916	0.4396
Fac-RSVM	**0.4714**	0.3660	**0.4289**	**0.4731**

1) Factorized Ranking SVM outperformed all baselines including Ranking SVM.
2) More improvements can be achieved on datasets with denser preference pairs.

5. Conclusion

1) There exists a low-rank structure among the Lagrange multipliers of Ranking SVM.
2) Factorized Ranking SVM decomposes each Lagrange multiplier as a dot product of two low-dimensional vectors.
3) Factorized Ranking SVM decreases space complexities from $O(N^2)$ to $O(KN)$.
4) Experimental results showed that Factorized Ranking SVM outperformed all baselines.