The Kendall and Mallows Kernels for Permutations

Yunlong Jiao & Jean-Philippe Vert

MINES ParisTech – CBIO, PSL Research University, Institut Curie, INSERM U900, Paris, France

Overview

This study revolves around computationally attractive alternatives to more complex kernels over the symmetric group $\mathbb{S}_n$ to learn from rankings, or to learn to rank.

Kendall and Mallows Kernels for Total Rankings

A *total ranking* is a strict ordering of n items $\{x_1, x_2, \dots, x_n\}$,

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_n} \iff \sigma \in \mathbb{S}_n : \sigma(i_j) = n - j + 1.$$

rank σ	rank σ'
2	3
4	5
1	2
6	6
3	4
5	1

- The number of discordant pairs is

$$n_d(\sigma, \sigma') = \sum_{i<j} \mathbb{1}_{\sigma(i)<\sigma(j)} \mathbb{1}_{\sigma'(i)>\sigma'(j)} + \mathbb{1}_{\sigma(i)>\sigma(j)} \mathbb{1}_{\sigma'(i)<\sigma'(j)} .$$

- The number of concordant pairs is

$$n_c(\sigma, \sigma') = \binom{n}{2} - n_d(\sigma, \sigma') .$$

n_d is called *Kendall tau distance* [3] and can be computed efficiently in $O(n \log n)$ time [4], underlying two popular similarity measures between permutations

- The *Mallows kernel* defined for any $\lambda \geq 0$ by

$$K_M^\lambda(\sigma, \sigma') = e^{-\lambda n_d(\sigma, \sigma')} ,$$

- The *Kendall kernel* defined as

$$K_\tau(\sigma, \sigma') = \frac{n_c(\sigma, \sigma') - n_d(\sigma, \sigma')}{\binom{n}{2}} .$$

Theorem 1 (Main Theorem)**.** *The Mallows kernel K_M^λ, for any $\lambda \geq 0$, and the Kendall kernel K_τ are positive definite.*

Proof. Consider the explicit kernel mapping

$$\Phi : \mathbb{S}_n \to \mathbb{R}^{\binom{n}{2}}, \sigma \mapsto \Big(sgn(\sigma(i) - \sigma(j)) \Big)_{1 \leq i < j \leq n} .$$

Then one immediately sees that the Kendall and Mallows kernel correspond respectively to a linear and Gaussian kernel on a $\binom{n}{2}$-dimensional embedding of $\mathbb{S}_n$. □

Kendall Kernel for Partial Rankings

A *partial ranking* is in general of the form

$$X_1 \succ X_2 \succ \cdots \succ X_k,$$

where $X_i \cap X_j = \varnothing, X_i \subset \{x_1, \dots, x_n\}$.

- Interleaving partial ranking

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_k}, \quad k \leq n.$$

- Top-k partial ranking

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_k} \succ X_{\text{rest}}, \quad k \leq n.$$

Any kernel over $\mathbb{S}_n$ can be extended to partial rankings with *convolution kernel* [2]

$$K_\tau^\star(R, R') = \frac{1}{|R||R'|} \sum_{\sigma \in R} \sum_{\sigma' \in R'} K_\tau(\sigma, \sigma'),$$

where $R \subset \mathbb{S}_n$ is the set of full permutations that are compatible with all the partial orders of a partial ranking.

Theorem 2. *For interleaving partial rankings of size k or top-k partial rankings, the convolution kernel induced by Kendall kernel can be evaluated efficiently in $O(k \log k)$ time.*

Kendall Kernel for Rankings with Uncertainty

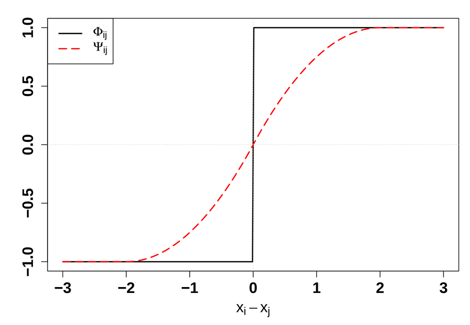

Figure 1: The (i, j)-th entry of smoothed feature mapping jittered with uniform noise ($\mathcal{U}[-1, 1])^n$.

Kendall mapping for quantitative vectors is discrete-valued and very sensitive to "almost ties", i.e.,

$$\Phi : \mathbb{R}^n \to \mathbb{R}^{\binom{n}{2}}, \mathbf{x} \mapsto \Big(\mathbb{1}_{x_i > x_j} - \mathbb{1}_{x_i < x_j} \Big)_{1 \leq i < j \leq n} .$$

We propose a noise-corrupted kernel mapping instead (similarly to [6])

$$\Psi(\mathbf{x}) = \mathbb{E}\Phi(\underbrace{\mathbf{x} + \epsilon}_{\tilde{\mathbf{x}}}) = \Big(\mathbb{P}\left(\tilde{x}_i > \tilde{x}_j\right) - \mathbb{P}\left(\tilde{x}_i < \tilde{x}_j\right) \Big)_{1 \leq i < j \leq n} .$$

Kendall kernel stabilized alternative is given by

$$G\left(\mathbf{x}, \mathbf{x}'\right) = \Psi(\mathbf{x})^\top \Psi(\mathbf{x}') = \mathbb{E} K_\tau(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}') .$$

Mallows Kernel vs. Diffusion Kernel over $\mathbb{S}_n$

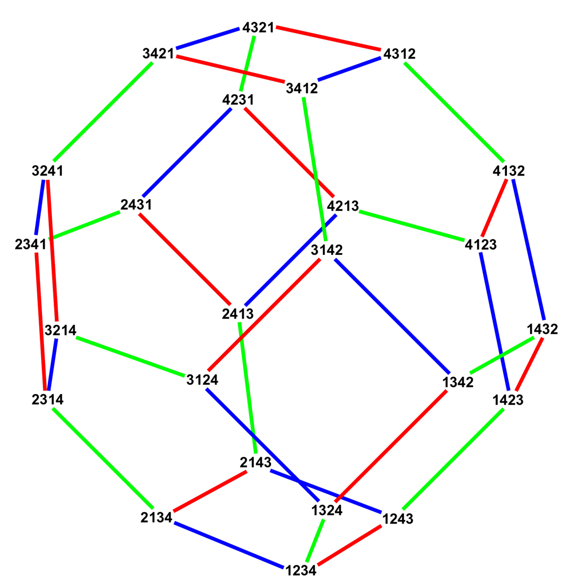

Figure 2: Cayley graph of $\mathbb{S}_4$, generated by the transpositions (1 2) in blue, (2 3) in green, and (3 4) in red.

Diffusion kernel [5] is defined by

$$K_{\text{dif}}^\beta(\sigma, \sigma') = [e^{\beta \Delta}]_{\sigma, \sigma'} ,$$

where Δ is the laplacian matrix of the Cayley graph of $\mathbb{S}_n$, i.e.,

$$\Delta_{\sigma, \sigma'} = \begin{cases} 1 & \text{if } \sigma \sim \sigma' \\ -(n-1) & \text{if } \sigma = \sigma' \\ 0 & \text{otherwise} \end{cases} ,$$

Mallows kernel is written

$$K_M^\lambda(\sigma, \sigma') = e^{-\lambda n_d(\sigma, \sigma')} ,$$

where $n_d(\sigma, \sigma') = d_{\mathcal{G}}(\sigma, \sigma')$ the shortest path distance on graph.

Experimental Results

Datasets

Dataset	No. of features	No. of samples (training/test)	
		C_1	C_2
Breast Cancer 1	23624	44/7 (Non-relapse)	32/12 (Relapse)
Breast Cancer 2	22283	142 (Non-relapse)	56 (Relapse)
Breast Cancer 3	22283	71 (Poor Prognosis)	138 (Good Prognosis)
Colon Tumor	2000	40 (Tumor)	22 (Normal)
Lung Cancer 1	7129	24 (Poor Prognosis)	62 (Good Prognosis)
Lung Cancer 2	12533	16/134 (ADCA)	16/15 (MPM)
Medulloblastoma	7129	39 (Failure)	21 (Survivor)
Ovarian Cancer	15154	162 (Cancer)	91 (Normal)
Prostate Cancer 1	12600	50/9 (Normal)	52/25 (Tumor)
Prostate Cancer 2	12600	13 (Non-relapse)	8 (Relapse)

Methods

- Kernel machines Support Vector Machines (SVM) and Kernel Fisher Discriminant (KFD) with Kendall kernel, linear kernel, Gaussian RBF kernel, polynomial kernel.
- Top Scoring Pairs (TSP) classifiers [1].
- Hybrid scheme of SVM + TSP feature selection algorithm.

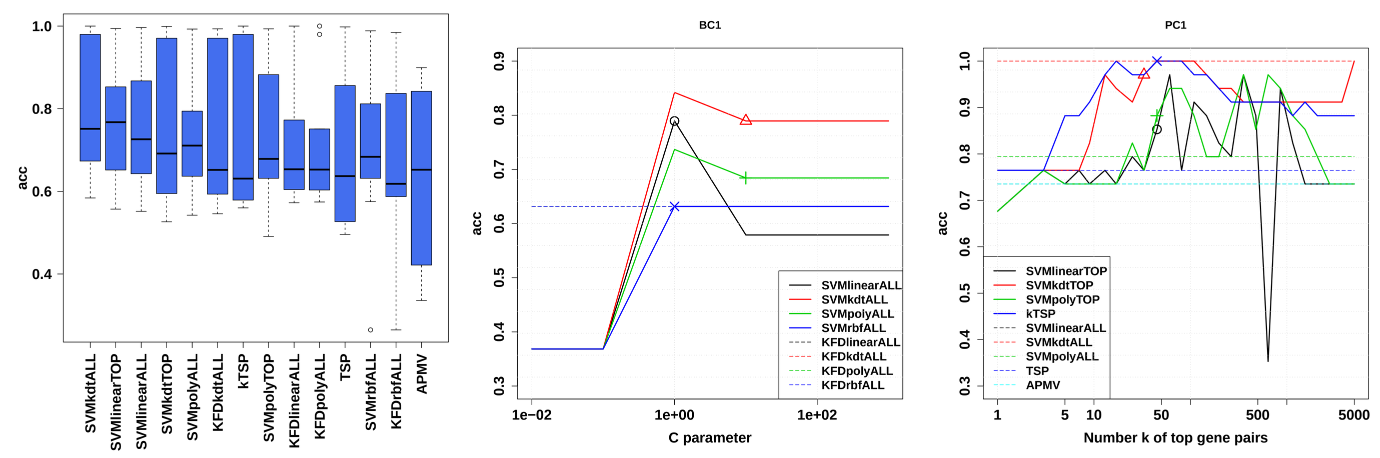

Figure 3: Left: Model performance comparison (ordered by decreasing average accuracy across datasets). **Middle**: Sensitivity of kernel SVMs to C parameter on the *Breast Cancer 1* dataset. **Right**: Impact of TSP feature selection on the *Prostate Cancer 1* dataset. (Special marks are returned by cross-validation.)

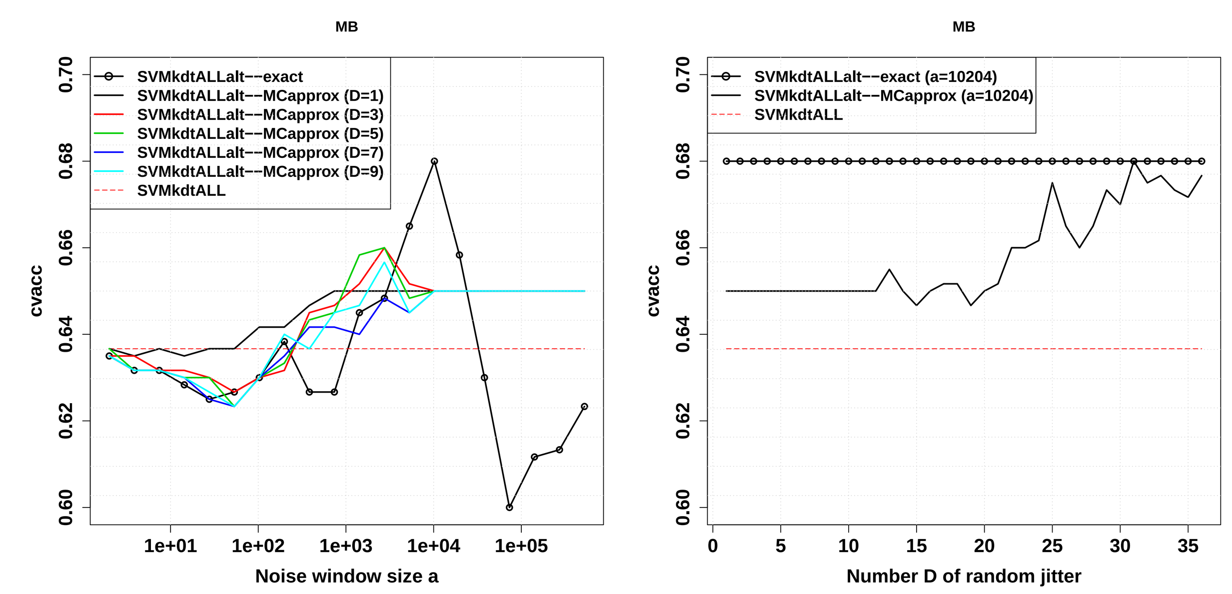

Figure 4: Left: Empirical performance of smoothed alternative to Kendall kernel on the *Medulloblastoma* dataset. **Right**: Empirical convergence of Monte Carlo approximate at the fixed window size attaining maximum underlying accuracy from the left plot.

Conclusion

- The Kendall and Mallows kernels are computationally efficient positive definite kernels over the symmetric group $\mathbb{S}_n$.
- Kendall kernel SVM shows very competitive performance for high-dimensional classification problems in biomedical applications.

Acknowledgements

This work was supported by the European Union 7th Framework Program through the Marie Curie ITN MLPM grant No 316861, and by the European Research Council grant ERC-SMAC-280032.

References

[1] D. Geman, C. d'Avignon, D. Q. Naiman, and R. L. Winslow. Classifying gene expression profiles from pairwise mRNA comparisons. *Stat. Appl. Genet. Mol. Biol.*, 3(1):Article19, 2004.

[2] D. Haussler. Convolution kernels on discrete structures. Technical Report UCSC-CRL-99-10, UC Santa Cruz, 1999.

[3] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[4] W. R. Knight. A computer method for calculating Kendall's tau with ungrouped data. *J. Am. Stat. Assoc.*, 61(314):436–439, 1966.

[5] I. R. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proceedings of the Nineteenth International Conference on Machine Learning*, volume 2, pages 315–322, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[6] K. Muandet, K. Fukumizu, F. Dinuzzo, and B. Schölkopf. Learning from distributions via support measure machines. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Adv. Neural. Inform. Process Syst.*, volume 25, pages 10–18. Curran Associates, Inc., 2012.