Записки научных
семинаров ПОМИ
Том 525, 2023 г.

И. Ф. Азангулов, Д. А. Еремеев

СТЕПЕННЫЕ КОВАРИАЦИОННЫЕ ФУНКЦИИ В ЗАДАЧЕ О ВЫУЧИВАНИИ ПЕРЕСТАНОВОК

Введение

Рассмотрим задачу о выучивании перестановок [2]. А именно, рассмотрим некоторое множество $\mathcal{X}$ и предположим, что имеется неизвестная функция $f : \mathcal{X}^n \to \mathrm{S}_n$, которая упорядочивает элементы $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$. Примером может быть случай, когда $\boldsymbol{x}$ – набор документов поисковой выдачи, а $f(\boldsymbol{x}) = \sigma$ переставляет их в соответствии с их релевантностью. Задача выучивания перестановок состоит в том, чтобы научиться по набору примеров $\{(\boldsymbol{x}^{(j)}, f(\boldsymbol{x}^{(j)})\}_{j=1}^N$ предсказывать правильную перестановку $f(\boldsymbol{x})$ для произвольного $\boldsymbol{x} \in \mathcal{X}^n$.

Один из подходов к решению данной задачи был предложен в [2]. Первый шаг состоит в том, чтобы рассмотреть оценку функции $f : \mathcal{X}^n \to \mathrm{S}_n$ как функцию $\widehat{f} : \mathcal{X}^n \times \mathrm{S}_n \to \mathbb{R}$, при этом значение $\widehat{f}(\boldsymbol{x}, \sigma)$ соответствует уверенности в том, что σ – соответствующая $\boldsymbol{x}$ перестановка. Далее для нахождения $\widehat{f}$ применяются стандартные ядерные методы, описанные в [5, 7]. Напомним, что ядром на произвольном множестве $\mathcal{Y}$ называется симметричная неотрицательно определенная функция на $\mathcal{Y} \times \mathcal{Y}$. При этом качество построенной по ядру аппроксимации сильно зависит от того, насколько подходящее ядро было выбрано. В частности, хорошее ядро должно учитывать особенности задачи и структуру пространства, на котором оно определено. В нашем случае в качестве $\mathcal{Y}$ выступает $\mathcal{X}^n \times \mathrm{S}_n$, а $\widehat{f}$ по определению должна быть инвариантна относительно одновременной перестановки $\boldsymbol{x}$ и σ: $\widehat{f}(\boldsymbol{x}, \sigma) = \widehat{f}(\boldsymbol{x}_\tau, \sigma\tau^{-1})$, где $\boldsymbol{x}_\tau = (x_{\tau(1)}, \dots, x_{\tau(n)})$ – образ $\boldsymbol{x}$ под естественным действием τ.

Ключевые слова: симметрическая группа, положительно определенные функции, ковариации, ядерные методы, моделирование.

Исследование выполнено при поддержке гранта Российского научного фонда No. 21-11-00047.

В [2] было показано, что естественным выбором ядер в таком случае будут ядра вида

$$k((\boldsymbol{x},\sigma),(\boldsymbol{y},\pi)) = \frac{1}{n!}\sum_{\tau\in \mathrm{S}_n} k_{\mathrm{S}_n}(\tau,e)\prod_{j=1}^{n} k_{\mathcal{X}}(\boldsymbol{x}_{\sigma^{-1}(j)},\boldsymbol{y}_{\tau\pi^{-1}(j)}), \qquad (1)$$

получающиеся усреднением ядра $k_{\mathrm{S}_n}(\sigma,\pi)\prod_{j=1}^{n} k_{\mathcal{X}}(\boldsymbol{x}_j,\boldsymbol{y}_j)$ относительно S_n, где k_{S_n} и $k_{\mathcal{X}}$ – ядра на соответствующих пространствах, причем k_{S_n} должно быть инвариантно относительно действия S_n на себе левыми и правыми сдвигами.

Как видно, ядро является суммой $n!$ слагаемых и, следовательно, вычисление его даже для одной пары точек является вычислительно затратным. В связи с этим актуальна задача построения аппроксимаций. В [2] были предложены подходы решения этой задачи для низкоранговых ядер на S_n. Однако данный подход не применим для степенных ядер, введённых в [8]. В этой статье мы восполняем этот пробел и предлагаем специальный метод аппроксимаций для этого класса.

В §1 мы напомним определение степенных ядер. В §2 дана аппроксимация ядра (1) в случае, когда k_{S_n} принадлежит этому классу, и получена оценка скорости сходимости. Наконец, в §3 мы рассмотрим задачу о частичных порядках [3], тесно связанную с исходной, и продемонстрируем некоторый прогресс для этого случая.

§1. Степенные ковариационные функции

В данной работе рассматривается случай, когда k_{S_n} взято из семейства степенных ковариационных функций, предложенного в [8]. Одним из основных преимуществ ядер из данного семейства является возможность их эффективного вычисления. Кратко напомним соответствующее определение.

Пусть $d \in \mathbb{Z}_+, m \in \mathbb{N}$, обозначим $p_d : \mathbb{C}^m \to \mathbb{C}$ степенную сумму Ньютона:

$$p_d(z_1,\ldots,z_m) = z_1^d + \cdots + z_m^d.$$

Для разбиения $\lambda = (\lambda_1,\ldots,\lambda_l)$, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_l$ определим $p_\lambda = \prod_{j=1}^{l} p_{\lambda_j}$. Также сопоставим каждой перестановке $\tau \in \mathrm{S}_n$ разбиение $\mu(\tau) = (\mu_1,\ldots,\mu_l)$, где l – общее количество циклов в τ, а μ_j есть длина j-го по величине цикла.

Определение. *Для вектора* $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{R}^m_{\geqslant 0}$, *являющегося параметром ядра, положим*

$$k^{\boldsymbol{z}}_{\mathrm{S}_n}(\sigma, \pi) = p_{\mu(\sigma\pi^{-1})}(z_1, \ldots, z_n).$$

Из того, что $p_{\mu(\sigma\pi^{-1})}$ – однородный многочлен степени n, следует, что

$$k^{\boldsymbol{z}}_{\mathrm{S}_n}(\sigma, \pi) = |\boldsymbol{z}|_1^n p_{\mu(\sigma\pi^{-1})}(\tilde{\boldsymbol{z}}) = |\boldsymbol{z}|_1^n \prod_{j=1}^{n} (\tilde{z}_1^j + \ldots + \tilde{z}_m^j)^{l_j},$$

где $\tilde{\boldsymbol{z}} = \boldsymbol{z}/|\boldsymbol{z}|_1$, а l_j – количество циклов длины j. Иными словами, ядро по определению мультипликативно "штрафует" на $\tilde{z}_1^j + \ldots + \tilde{z}_m^j$ за каждый цикл длины j в частном $\sigma\pi^{-1}$.

§2. Вычислительно эффективное представление

Как было упомянуто ранее, наивное вычисление ядра (1) по определению представляется вычислительно трудной задачей, и возникает необходимость в поиске более вычислительно эффективных представлений. Одно из них даёт следующая теорема.

Обозначим через $\mathbf{R}^{m,n}$ случайную $m \times n$ матрицу, элементы которой – независимые одинаково распределенные случайные величины с нулевым математическим ожиданием и единичной дисперсией. В дальнейшем индексы m, n будут опускаться, если их значения ясны из контекста.

Пусть $\mathbf{D}(\boldsymbol{z}) = \operatorname{diag}(z_1, \ldots, z_m)$ – диагональная матрица, соответствующая вектору $\boldsymbol{z} \in \mathbb{R}^m$, а $\mathbf{P}_\sigma$ обозначает матрицу перестановки, соответствующую σ.

Для ядра $k_{\mathcal{X}}(\cdot, \cdot)$ на пространстве $\mathcal{X}$ и $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}^n$ обозначим через $\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}} = \big(k_{\mathcal{X}}(x_i, y_j)\big)_{i,j=1}^n$ матрицу попарных ковариаций между $\boldsymbol{x}$ и $\boldsymbol{y}$. Наконец, будем обозначать $\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}} = \mathbf{P}_\sigma \mathbf{K}_{\boldsymbol{x},\boldsymbol{y}} \mathbf{P}_\pi^T$.

Теорема 1. *Ядро на* $\mathcal{X}^n \times S_n$, *заданное симметризацией степенного ядра* $k^{\boldsymbol{z}}_{\mathbb{S}_n}$ *и произвольного ядра* $k_{\mathcal{X}}$ *на* $\mathcal{X}$, *может быть представлено следующим образом:*

$$k((\boldsymbol{x}, \sigma), (\boldsymbol{y}, \pi)) = \frac{1}{n!} \mathbb{E} \prod_{j=1}^{n} \big(\mathbf{R}^* \mathbf{D}(\boldsymbol{z}) \mathbf{R} \mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\big)_{j,j}.$$

Доказательство теоремы будет основано на следующей лемме.

Лемма 1. *Пусть* $I : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. *Тогда*

$$\mathbb{E} \prod_j (\mathbf{R}^* \mathbf{D}(\boldsymbol{z}) \mathbf{R})_{j, I(j)} = \begin{cases} k^{\boldsymbol{z}}_{\mathrm{S}_n}(I, e), & \text{если } I \text{ – перестановка,} \\ 0, & \text{иначе.} \end{cases}$$

Доказательство. Через $\mathbf{U}$ обозначим матрицу $\sqrt{\mathbf{D}(\boldsymbol{z})}\mathbf{R}$ с вектор-столбцами $\mathbf{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$.

Пусть I не является перестановкой. Пусть $j_0 \notin \operatorname{Im}(I)$, тогда

$$\begin{aligned} \mathbb{E} \prod_j (\mathbf{R}^* \mathbf{D}(\boldsymbol{z}) \mathbf{R})_{j, I(j)} &= \mathbb{E} \prod_j (\mathbf{U}^* \mathbf{U})_{j, I(j)} \\ &= \mathbb{E} \prod_j \left(\boldsymbol{u}_j^* \boldsymbol{u}_{I(j)} \right) \\ &= \mathbb{E} \boldsymbol{u}_{j_0}^* \mathbb{E} \boldsymbol{u}_{I(j_0)} \prod_{j \neq j_0} \left(\boldsymbol{u}_j^* \boldsymbol{u}_{I(j)} \right) = 0. \end{aligned}$$

Пусть теперь τ – перестановка. Рассмотрим произвольный цикл $(k_1 \ldots k_d)$ в τ. Тогда

$$\begin{aligned} &\mathbb{E} \boldsymbol{u}_{k_1}^* \boldsymbol{u}_{k_2} \boldsymbol{u}_{k_2}^* \boldsymbol{u}_{k_3} \ldots \boldsymbol{u}_{k_d}^* \boldsymbol{u}_{k_1} \\ &\quad = \mathbb{E} \left(\left(\sum_{j_1 = 1}^m \overline{\mathbf{U}_{j_1, k_1}} \mathbf{U}_{j_1, k_2} \right) \ldots \left(\sum_{j_d = 1}^m \overline{\mathbf{U}_{j_d, k_d}} \mathbf{U}_{j_d, k_1} \right) \right) \\ &\quad = \sum_{j_1, \ldots, j_d} \mathbf{1}_{j_1 = \ldots = j_d} \mathbb{E} |\mathbf{U}_{j_1, k_1}|^2 \ldots |\mathbf{U}_{j_d, k_d}|^2 \\ &\quad = \sum_{j=1}^m z_j^d = p_d(z_1, \ldots, z_m). \end{aligned}$$

Пользуясь независимостью столбцов $\mathbf{R}$ и только что доказанным равенством, получим требуемое. □

Теперь мы готовы доказать саму теорему.

Доказательство теоремы. Имеем

$$
\begin{aligned}
&\frac{1}{n!}\mathbb{E}\prod_j \left(\mathbf{R}^*\mathbf{D}(\boldsymbol{z})\mathbf{R}\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{j,j} \\
&\quad = \frac{1}{n!}\mathbb{E}\prod_j \left(\sum_k \left(\mathbf{R}^*\mathbf{D}(\boldsymbol{z})\mathbf{R}\right)_{j,k}\left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{k,j}\right) \\
&\quad = \frac{1}{n!}\sum_I \left(\mathbb{E}\prod_j \left(\mathbf{R}^*\mathbf{D}(\boldsymbol{z})\mathbf{R}\right)_{j,I(j)}\right)\left(\prod_j \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{I(j),j}\right),
\end{aligned}
$$

где суммирование ведется, вообще говоря, по всем отображениям

$$
I : \{1,2,3,\dots,n\} \to \{1,2,3,\dots,n\}.
$$

Пользуясь доказанной леммой, получаем

$$
\begin{aligned}
&\frac{1}{n!}\sum_{\tau\in \mathrm{S}_n} k^{\boldsymbol{z}}_{\mathrm{S}_n}(\tau^{-1},e)\left(\prod_j \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{\tau^{-1}(j),j}\right) \\
&\quad = \frac{1}{n!}\sum_{\tau\in \mathrm{S}_n} k^{\boldsymbol{z}}_{\mathrm{S}_n}(\tau^{-1},e)\left(\prod_j \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{j,\tau(j)}\right) \\
&\quad = \frac{1}{n!}\sum_{\tau\in \mathrm{S}_n} k^{\boldsymbol{z}}_{\mathrm{S}_n}(\tau,e)\prod_{j=1}^n k_{\mathcal{X}}\left(\boldsymbol{x}_{\sigma^{-1}(j)},\boldsymbol{y}_{\tau\pi^{-1}(j)}\right) \\
&\quad = k((\boldsymbol{x},\sigma),(\boldsymbol{y},\pi)). \qquad \square
\end{aligned}
$$

Теорема 1 позволяет приблизить ядро при помощи Монте-Карло аппроксимации:

$$
k((\boldsymbol{x},\sigma),(\boldsymbol{y},\pi)) \approx \frac{1}{L}\sum_{l=1}^{L}\frac{1}{n!}\prod_{j=1}^{n}\left(\mathbf{R}_l^*\mathbf{D}(\boldsymbol{z})\mathbf{R}_l\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{j,j},
$$

где $\mathbf{R}_l$ независимы и одинаково распределены.

Оценка скорости сходимости данного приближения выражается в терминах дисперсии случайной величины

$$
X := \frac{1}{n!}\prod_{j=1}^{n}\left(\mathbf{R}^*\mathbf{D}(\boldsymbol{z})\mathbf{R}\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{j,j}.
$$

Отметим, что случайная величина X зависит от совместного распределения элементов $\mathbf{R}$. В дальнейшем мы будем полагать, что $\mathbf{R}_{i,j} \sim \mathcal{N}(0,1)$ – независимые одинаково распределенные стандартные нормальные величины.

Теорема 2. *Справедливо неравенство*

$$\mathbb{E}|X|^2 \leqslant \left(\frac{1}{n!}\right)^2 \left(\prod_{j=1}^{n} \left(\sum_{k=1}^{n} \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)^2_{k,j}\right)\right) \mathbb{E}|\boldsymbol{u}_1|^{4n}.$$

Доказательство. Имеем

$$\mathbb{E}|X|^2 = \left(\frac{1}{n!}\right)^2 \mathbb{E} \prod_{j=1}^{n} \left|\mathbf{R}^*\mathbf{D}(\boldsymbol{z})\mathbf{R}\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right|^2_{j,j} = \left(\frac{1}{n!}\right)^2 \mathbb{E} \prod_{j=1}^{n} \left|\mathbf{U}^*\mathbf{U}\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right|^2_{j,j}.$$

Используя неравенство Гёльдера для нескольких множителей, получим

$$\mathbb{E}|X|^2 \leqslant \left(\frac{1}{n!}\right)^2 \prod_{j=1}^{n} \left(\mathbb{E}\left|\mathbf{U}^*\mathbf{U}\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right|^{2n}_{j,j}\right)^{1/n}. \tag{2}$$

Оценим каждый из множителей по отдельности.

$$\begin{aligned}
\mathbb{E}\left|\mathbf{U}^*\mathbf{U}\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right|^{2n}_{j,j} &= \mathbb{E}\left|\sum_{k=1}^{n}(\mathbf{U}^*\mathbf{U})_{j,k}(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}})_{k,j}\right|^{2n} \\
&= \mathbb{E}\left(\langle \boldsymbol{u}_j, \sum_{k=1}^{n} \boldsymbol{u}_k \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{k,j}\rangle\right)^{2n} \\
&\leqslant \mathbb{E}\left|\boldsymbol{u}_j\right|^{2n} \cdot \left|\sum_{k=1}^{n} \boldsymbol{u}_k \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{k,j}\right|^{2n} \\
&\leqslant \left(\mathbb{E}\left|\boldsymbol{u}_j\right|^{4n}\right)^{1/2} \left(\mathbb{E} \cdot \left|\sum_{k=1}^{n} \boldsymbol{u}_k \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{k,j}\right|^{4n}\right)^{1/2} \\
&= \left(\sum_{k=1}^{n} \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)^2_{k,j}\right)^{n} \mathbb{E}\left|\boldsymbol{u}_j\right|^{4n}.
\end{aligned}$$

В последнем переходе мы воспользовались тем, что $\sum\limits_{k=1}^{n} \boldsymbol{u}_k \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)_{k,j}$ по распределению равно $\sqrt{\sum\limits_{k=1}^{n} \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)^2_{k,j} \boldsymbol{u_j}}$, так как $\mathbf{u}_k \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{D}(\boldsymbol{z}))$.

Подставляя в (2), получаем

$$\mathbb{E}|X|^2 \leqslant \left(\frac{1}{n!}\right)^2 \left(\prod_{j=1}^{n} \left(\sum_{k=1}^{n} \left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)^2_{k,j}\right)\right) \mathbb{E}|\boldsymbol{u}_1|^{4n}. \tag{3}$$

□

Заметим, что

$$|\boldsymbol{u}_1|^2 = \sum_{k=1}^{n} z_k \mathbf{R}^2_{1,k}, \mathbf{R}^2_{1,k} \overset{i.i.d}{\sim} \chi^2_1.$$

Тогда, пользуясь свойствами производящей функции моментов и формулой для $M_{\chi^2_1}(t)$, можно получить явную формулу и для $M_{|\boldsymbol{u}_1|^2}(t)$:

$$M_{|\boldsymbol{u}_1|^2}(t) = \prod_{j=1}^{n} M_{\chi^2_1}(z_j t) = \prod_{j=1}^{n} (1 - 2z_j t)^{-1/2},$$

что позволяет численно найти необходимый момент. Более того, если $\boldsymbol{z} = \mathbf{1}_n$, то оценку можно выписать явно, используя формулу Стирлинга

$$\begin{aligned}
\mathbb{E}\,|\boldsymbol{u}_j|^{4n} &= 2^{2n} \frac{\Gamma\left((n+4n)/2\right)}{\Gamma(n/2)} \\
&= O\left(2^{2n} \frac{\sqrt{5\pi n}\left(\frac{5n}{2e}\right)^{5n/2}}{\sqrt{\pi n}\left(\frac{n}{2e}\right)^{n/2}}\right) \\
&= O\left(2^{2n} \frac{\left(\frac{5n}{2e}\right)^{5n/2}}{\left(\frac{n}{2e}\right)^{n/2}}\right) \\
&= O\left(5^{5n/2}\left(\frac{n}{e}\right)^{2n}\right).
\end{aligned}$$

И следовательно

$$\mathbb{E}|X|^2 \leqslant \left(\frac{1}{n!}\right)^2 \left(\prod_{j=1}^{n}\left(\sum_{k=1}^{n}\left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)^2_{k,j}\right)\right)\mathbb{E}|\boldsymbol{u}_1|^{4n}$$
$$\leqslant O\left(\frac{5^{5n/2}}{n}\prod_{j=1}^{n}\left(\sum_{k=1}^{n}\left(\mathbf{K}^{\sigma,\pi}_{\boldsymbol{x},\boldsymbol{y}}\right)^2_{j,k}\right)\right).$$

§3. Частичные статистики

На практике часто нас может интересовать не вся перестановка, а лишь первые несколько элементов. Например, если речь идёт о задаче ранжирования, то нам могут быть нужны только первые k наиболее релевантных объектов, то есть значения $\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_k^{-1}$. При этом, как и раньше, обучение происходит на парах $(\boldsymbol{x}, \sigma) \in \mathcal{X}^n \times \mathrm{S}_n$.

Определение. *Значение ядра для* $A, B \subseteq \mathrm{S}_n$*, где* $A, B \neq \varnothing$*, естественно доопределяется* [3] *как*

$$k((\boldsymbol{x}, B), (\boldsymbol{y}, A)) = \frac{1}{|A||B|}\sum_{\pi\in B}\sum_{\sigma\in A} k((\boldsymbol{x}, \pi), (\boldsymbol{y}, \sigma)).$$

В контексте описанной выше задачи нас интересует случай, когда $B = \{\pi\}$ и $A = \{\sigma \,|\, \sigma_1^{-1} = a_1, \ldots, \sigma_k^{-1} = a_k\}$. Используя инвариантность

$$k((\boldsymbol{x}, \pi), (\boldsymbol{y}, \sigma)) = k((\boldsymbol{x}_{\tau_1}, \tau_1^{-1}\pi), (\boldsymbol{y}_{\tau_2}, {\tau_2}^{-1}\sigma)),$$

достаточно научиться работать со случаем, когда $A = \{\sigma \,|\, \sigma_1 = 1, \ldots, \sigma_k = k\}$, а $B = \{e\}$.

Имеем

$$k((\boldsymbol{x}, e), (\boldsymbol{y}, A)) = \frac{1}{|A|}\sum_{\sigma\in A} k((\boldsymbol{x}, e), (\boldsymbol{y}, \sigma))$$
$$\propto \sum_{\sigma\in A}\sum_{\tau\in \mathrm{S}_n} k_{\mathrm{S}_n}(\tau, e)\prod_{j=1}^{n} k_{\mathcal{X}}(x_j, y_{\tau\sigma^{-1}(j)})$$
$$= \sum_{\tau\in \mathrm{S}_n} k_{\mathrm{S}_n}(\tau, e)\prod_{j=1}^{k} k_{\mathcal{X}}(x_j, y_{\tau(j)}) \sum_{\sigma\in\mathrm{Sym}([k+1,n])}\prod_{j>k} k_{\mathcal{X}}(x_j, y_{\tau\sigma^{-1}(j)})$$
$$= \sum_{\tau\in \mathrm{S}_n} k_{\mathrm{S}_n}(\tau, e)\prod_{j=1}^{k} k_{\mathcal{X}}(x_j, y_{\tau(j)})\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[k+1,n],\tau([k+1,n])}),$$

где $[p,q] = \{p, p+1, \dots, q-1, q\}$ и $\propto$ означает существование такой ненулевой константы, зависящей лишь от n и k, что левая часть равна правой, умноженной на эту константу. Так как перманент не зависит от перестановок столбцов, то получаем выражение

$$\sum_{\tau': [k] \overset{inj.}{\mapsto} [n]} \mathcal{S}(\tau') \prod_{j=1}^{k} k_{\mathcal{X}}(x_j, y_{\tau'(j)}) \mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[k+1,n],[n]\setminus\tau'([k])}),$$

где $[n] = \{1, 2, \dots, n\}$ и

$$\mathcal{S}(\tau') = \sum_{\substack{\tau \in \mathrm{S}_n: \\ \tau|_{[k]} = \tau'}} k_{\mathrm{S}_n}(\tau, e).$$

3.1. Случай $k = 1$. В случае $A = \{\sigma | \sigma_1 = 1\}$ получившуюся формулу можно упростить. Рассмотрим, какой длины цикл может соответствовать 1 в перестановке τ. Группируя по длине этого цикла, получаем:

$$\begin{aligned} &k((\boldsymbol{x}, e), (\boldsymbol{y}, A)) \\ &\quad \propto \left(p_1(\boldsymbol{z}) \sum_{\alpha \in \mathrm{S}_{n-1}} p_\alpha(\boldsymbol{z}) \right) k_{\mathcal{X}}(x_1, y_1) \mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[2,n],[n]\setminus\{1\}}) \\ &\quad + \sum_{s=2}^{n} \left(\sum_{l=2}^{n} \frac{(n-2)!}{(n-l)!} p_l \sum_{\alpha \in \mathrm{S}_{n-l}} p_\alpha \right) k_{\mathcal{X}}(x_1, y_s) \mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[2,n],[n]\setminus\{s\}}). \end{aligned}$$

Полученное выражение можно упростить, используя классические соотношения для симметрических полиномов [4]:

$$\sum_{\alpha \in \mathrm{S}_l} p_\alpha(\boldsymbol{z}) = l!\, h_l(\boldsymbol{z}), \sum_{j=1}^{l} p_i(\boldsymbol{z}) h_{l-j}(\boldsymbol{z}) = l h_l(\boldsymbol{z}), \tag{4}$$

где $h_l(\boldsymbol{z})$ – полный однородный симметрический полином:

$$h_l(\boldsymbol{z}) = \sum_{1 \leqslant j_1 \leqslant j_2 \leqslant \dots \leqslant j_l \leqslant n} \prod_{s=1}^{l} z_{j_s}.$$

Тогда

$$\sum_{l=2}^{n} \frac{(n-2)!}{(n-l)!} p_l \sum_{\alpha \in \mathrm{S}_{n-l}} p_\alpha = (n-2)! \sum_{l=2}^{n} p_l h_{n-l}$$
$$= (n-2)!\,(n h_n - p_1 h_{n-1}),$$

и

$$p_1 \sum_{\alpha \in \mathrm{S}_{n-1}} p_\alpha = (n-1)!\, p_1 h_{n-1}.$$

Таким образом,

$$\begin{aligned} &k((\boldsymbol{x}, e), (\boldsymbol{y}, A)) \\ &\quad \propto (n-1)!\, p_1 h_{n-1} k_{\mathcal{X}}(x_1, y_1) \mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[2,n],[n]\setminus\{1\}} \right) \\ &\quad + \sum_{s=2}^{n} (n-2)!\, (n h_n - p_1 h_{n-1})\, k_{\mathcal{X}}(x_1, y_s) \mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[2,n],[n]\setminus\{s\}} \right) \\ &\quad = (n-2)! \Bigg((n h_n - p_1 h_{n-1})\, \mathrm{Perm}\,(\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}}) + \\ &\quad + n\,(p_1 h_{n-1} - h_n)\, k_{\mathcal{X}}(x_1, y_1) \mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[2,n],[2,n]} \right) \Bigg). \end{aligned}$$

Задача свелась к вычислению перманентов и симметрических многочленов p_j, h_j. Многочлен p_j может быть легко вычислен по определению, а для вычисления h_j за полиномиальное время можно воспользоваться соотношением (4).

Несмотря на то, что наилучший известный алгоритм [6] для вычисления перманента работает за время $O(n2^n)$, существует полиномиальный по времени алгоритм [1] аппроксимации для матриц с неотрицательными элементами, который будет применим в большинстве случаев, так как значения ядер, используемые на практике, обычно неотрицательны.

3.2. Случай $k = 2$. Аналогичный подход может быть применён для $k = 2$. Разобьём сумму

$$\sum_{\tau' : [2] \overset{inj}{\mapsto} [n]} \mathcal{S}(\tau') \prod_{j=1}^{2} k_{\mathcal{X}}(x_j, y_{\tau'(j)}) \mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[n]\setminus\tau'([2])}) \qquad (5)$$

в зависимости от вида τ'.

- Если $\tau'(1) = 2, \tau'(2) = 1$, то
$$\mathcal{S}(\tau') = \sum_{\alpha \in \mathrm{S}_{n-2}} p_2 p_\alpha = (n-2)!\, p_2 h_{n-2}.$$
Поэтому вклад в (5) будет равен
$$(n-2)!\, p_2 h_{n-2} k_{\mathcal{X}}(x_1, y_2) k_{\mathcal{X}}(x_2, y_1) \mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]}).$$
- Если $\tau'(1) = 1, \tau'(2) = 2$, то
$$\mathcal{S}(\tau') = \sum_{\alpha \in \mathrm{S}_{n-2}} p_1^2 p_\alpha = (n-2)!\, p_1^2 h_{n-2}.$$
Поэтому вклад в (5) будет равен
$$(n-2)!\, p_1^2 h_{n-2} k_{\mathcal{X}}(x_1, y_1) k_{\mathcal{X}}(x_2, y_2) \mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]}).$$
- Если $\tau'(1)=1$, $\tau'(2)=s>2$, то сведём к случаю $k=1$, $\tau'(1)>1$
$$\mathcal{S}(\tau') = p_1 (n-3)!\, \left((n-1) h_{n-1} - p_1 h_{n-2}\right).$$
Поэтому вклад в (5) всех τ' данного вида будет равен
$$\begin{aligned} &(n-3)!\, p_1 \left((n-1) h_{n-1} - p_1 h_{n-2}\right) k_{\mathcal{X}}(x_1, y_1) \\ &\times \sum_{s=3}^{n} k_{\mathcal{X}}(x_2, y_s) \mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[2,n]\setminus\{s\}}\right) \\ &= (n-3)!\, p_1 \left((n-1) h_{n-1} - p_1 h_{n-2}\right) k_{\mathcal{X}}(x_1, y_1) \\ &\times \left(\mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[2,n],[2,n]}\right) - k_{\mathcal{X}}(x_2, y_2) \mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]}\right)\right). \end{aligned}$$
- Если $\tau'(1) = s > 2, \tau'(2) = 2$, то из симметрии получаем, что вклад всех τ' данного вида равен

$$\begin{aligned} &(n-3)!\, p_1 \left((n-1) h_{n-1} - p_1 h_{n-2}\right) k_{\mathcal{X}}(x_2, y_2) \\ &\times \left(\mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[1,n]\setminus\{2\},[1,n]\setminus\{2\}}\right) - k_{\mathcal{X}}(x_1, y_1) \mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]}\right)\right). \end{aligned}$$

- Если $\tau'(1) = 2, \tau'(2) = s > 2$, то аналогично случаю $k=1$, $\tau'(1) > 1$ распишем сумму по длине l цикла, содержащего 1,

$$\mathcal{S}(\tau') = \sum_{l=3}^{n} \frac{(n-3)!}{(n-l)!} (n-l)!\, h_{n-l} = (n-3)!\, (n h_n - p_1 h_{n-1} - p_2 h_{n-2}).$$

Поэтому вклад в (5) всех τ' данного вида будет равен

$$\begin{aligned}
&(n-3)!\,(nh_n - p_1h_{n-1} - p_2h_{n-2})k_{\mathcal{X}}(x_1,y_2)\\
&\times \sum_{s=3}^{n} k_{\mathcal{X}}(x_2,y_s)\mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[1,n]\setminus\{2,s\}}\right)\\
&= (n-3)!\,(nh_n - p_1h_{n-1} - p_2h_{n-2})k_{\mathcal{X}}(x_1,y_2)\\
&\times \left(\mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[2,n],[1,n]\setminus\{2\}}\right) - k_{\mathcal{X}}(x_2,y_1)\mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]}\right)\right).
\end{aligned}$$

- Если $\tau'(1) = s > 2, \tau'(2) = 1$, то из симметрии получаем, что вклад всех τ' данного вида равен

$$\begin{aligned}
&(n-3)!\,(nh_n - p_1h_{n-1} - p_2h_{n-2})k_{\mathcal{X}}(x_2,y_1)\\
&\times \left(\mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[1,n]\setminus\{2\},[2,n]}\right) - k_{\mathcal{X}}(x_1,y_2)\mathrm{Perm}\left((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]}\right)\right).
\end{aligned}$$

- Если $\tau'(1) = s_1 > 2, \tau'(2) = s_2 > 2$, то есть два подслучая, которые надо просуммировать.

  Во-первых, эти элементы могут быть в одном цикле длины l, тогда вклад в $\mathcal{S}(\tau')$ равен

$$\sum_{l=4}^{n} \frac{(n-4)!}{(n-l)!} p_l(l-3)(n-l)!\,h_{n-l} = (n-4)!\,\sum_{l=4}^{n}(l-3)p_l h_{n-l}.$$

  Во-вторых, эти элементы могут быть в разных циклах

$$\begin{aligned}
&\sum_{l=2}^{n-2}\sum_{r=2}^{n-l} \frac{(n-4)!}{(n-(l+r))!} p_l p_r (n-(l+r))!\,h_{n-(l+r)}\\
&\qquad\qquad = (n-4)!\,\sum_{l=2}^{n-2}\sum_{r=2}^{n-l} p_l p_r h_{n-(l+r)}.
\end{aligned}$$

  Причём последнее выражение можно ещё упростить, заметив, что

$$\sum_{r=2}^{n-l} p_r h_{(n-l)-r} = (n-l)h_{n-l} - p_1 h_{n-l-1}.$$

Сложим результаты для двух подслучаев и получим

$$\begin{aligned}
\mathcal{S}(\tau') &= (n-4)!\sum_{l=4}^{n}(l-3)p_l h_{n-l} + (n-4)!\sum_{l=2}^{n-2} p_l\left((n-l)h_{n-l} - p_1 h_{n-l-1}\right)\\
&= (n-4)!\left(\sum_{l=1}^{n}(l-3)p_l h_{n-l} + 2p_1 h_{n-1} + p_2 h_{n-2}\right.\\
&+ \sum_{l=1}^{n}(n-l)p_l h_{n-l} - (n-1)p_1 h_{n-1} - p_{n-1}h_1\\
&\left.- p_1\sum_{l=1}^{n-1} p_l h_{n-l-1} + p_1 p_{n-1} + p_1^2 h_{n-2}\right)\\
&= (n-4)!\left((n-3)nh_n + 2p_1 h_{n-1} + p_2 h_{n-2} - (n-1)p_1 h_{n-1}\right.\\
&\left.- p_{n-1}h_1 - (n-1)p_1 h_{n-1} + p_1 p_{n-1} + p_1^2 h_{n-2}\right)\\
&= (n-4)!\left((n-3)nh_n - (2n-4)p_1 h_{n-1} + p_2 h_{n-2} + p_1^2 h_{n-2}\right).
\end{aligned}$$

Таким образом, вклад в (5) всех τ' данного вида будет равен

$$\begin{aligned}
&(n-4)!\left((n-3)nh_n - (2n-4)p_1 h_{n-1} + p_2 h_{n-2} + p_1^2 h_{n-2}\right)\\
&\quad\times \sum_{\substack{s_1,s_2\in[3,n]\\ s_1\neq s_2}} k_{\mathcal{X}}(x_1,y_{s_1})k_{\mathcal{X}}(x_2,y_{s_2})\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[n]\setminus\{s_1,s_2\}})\\
&= (n-4)!\left((n-3)nh_n - (2n-4)p_1 h_{n-1} + p_2 h_{n-2} + p_1^2 h_{n-2}\right)\Bigg(\mathrm{Perm}\,(\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})\\
&- \sum_{\substack{s_1\in[1,2]\\ s_2\in[3,n]}} k_{\mathcal{X}}(x_1,y_{s_1})k_{\mathcal{X}}(x_2,y_{s_2})\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[n]\setminus\{s_1,s_2\}})\\
&- \sum_{\substack{s_1\in[3,n]\\ s_2\in[1,2]}} k_{\mathcal{X}}(x_1,y_{s_1})k_{\mathcal{X}}(x_2,y_{s_2})\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[n]\setminus\{s_1,s_2\}})\\
&+ \sum_{\substack{s_1,s_2\in[1,2]\\ s_1\neq s_2}} k_{\mathcal{X}}(x_1,y_{s_1})k_{\mathcal{X}}(x_2,y_{s_2})\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[n]\setminus\{s_1,s_2\}})\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= (n-4)!\left((n-3)nh_n - (2n-4)p_1h_{n-1} + p_2h_{n-2} + p_1^2h_{n-2}\right)\Bigg(\mathrm{Perm}\,(\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}}) \\
&- k_{\mathcal{X}}(x_1,y_1)(\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[n]\setminus\{1\},[n]\setminus\{1\}}) - (k_{\mathcal{X}}(x_2,y_2))\,\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]})) \\
&- k_{\mathcal{X}}(x_1,y_2)(\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[n]\setminus\{1\},[n]\setminus\{2\}}) - (k_{\mathcal{X}}(x_2,y_1))\,\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]})) \\
&- k_{\mathcal{X}}(x_2,y_1)(\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[n]\setminus\{2\},[n]\setminus\{1\}}) - (k_{\mathcal{X}}(x_1,y_2))\,\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]})) \\
&- k_{\mathcal{X}}(x_2,y_2)(\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[n]\setminus\{2\},[n]\setminus\{2\}}) - (k_{\mathcal{X}}(x_1,y_1))\,\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]})) \\
&+ k_{\mathcal{X}}(x_1,y_1)k_{\mathcal{X}}(x_2,y_2)\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]}) \\
&+ k_{\mathcal{X}}(x_1,y_2)k_{\mathcal{X}}(x_2,y_1)\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]})\Bigg) \\
&= (n-4)!\left((n-3)nh_n - (2n-4)p_1h_{n-1} + p_2h_{n-2} + p_1^2h_{n-2}\right)\Bigg(\mathrm{Perm}\,(\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}}) \\
&+ 3\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[3,n],[3,n]})\,(k_{\mathcal{X}}(x_1,y_1)k_{\mathcal{X}}(x_2,y_2) + k_{\mathcal{X}}(x_1,y_2)k_{\mathcal{X}}(x_2,y_1)) \\
&- \sum_{s_1\in[1,2]}\sum_{s_2\in[1,2]} k_{\mathcal{X}}(x_{s_1},y_{s_2})\mathrm{Perm}((\mathbf{K}_{\boldsymbol{x},\boldsymbol{y}})_{[n]\setminus\{s_1\},[n]\setminus\{s_2\}})\Bigg).
\end{aligned}
$$

Складывая вышеуказанные слагаемые, в итоге получаем формулу с не зависящим от n числом слагаемых, каждое из которых может быть приближенно вычислено за полиномиальное время.

Список литературы

1. M. Jerrum, A. Sinclair, E. Vigoda, *A polynomial-time approximation algorithms for permanent of a matrix with non-negative entries.* — J. ACM **51** (2004), 671–697.
2. R. Kondor, *Group Theoretical Methods in Machine Learning.* PhD Thesis, 2008.
3. R. Kondor, M. Barbosa, *Ranking with kernels in Fourier space.* — In: COLT 2010 – The 23rd Conference on Learning Theory (2010), 451–463.
4. I. G. Macdonald, *Symmetric Functions and Hall Polynomials.* Oxford classic texts in the physical sciences, Clarendon Press, (1998).
5. C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning.* MIT Press, 2006.
6. H. J. Ryser, *Combinatorial Mathematics.* The Carus Mathematical Monographs, AMS, 1963.
7. V. N. Vapnik, *The nature of statistical learning theory.* Springer-Verlag, New York, 1995.
8. И. Ф. Азангулов, В. А. Боровицкий, А. В.Смоленский, *Об одном классе гауссовских процессов на симметрической группе.* — Зап. научн. семин. ПОМИ **515** (2022), 19–29.

Azangulov I. F., Eremeev D. A. Power sum kernels in permutation learning.

In this paper, we consider the use of power sum kernels in solving the problem of permutation learning. We present a way to approximate a symmetrized kernel that naturally arises in this problem using the Monte Carlo method and estimate the convergence rate. We also touch on the problem of partial rankings and present some results for the case when the number of fixed elements is 1 or 2.

С.-Петербургский государственный университет,
Университетская наб. 7/9,
С.-Петербург, Россия
E-mail: iska.azn@gmail.com
E-mail: dimaeremeev2002@gmail.com

Поступило 7 октября 2023 г.