Конференция, посвященная 120-летию А.Н. Колмогорова

Москва, МГУ

25 апреля 2023 г.

О работах А.Н. Колмогорова по теории функций и их влиянии на развитие математики XX-XXI веков

Б. С. Кашин

I) $I^n = [0,1]^n$ – n-мерный куб, $n = 1, 2, \ldots$
T – единичная окружность
$C(I^n)$, $C(\mathbb{R}^n)$, $C(T)$ – соответствующее пространство непрерывных функций

Теорема (А.Н. Колмогоров, ДАН СССР, Т. 114, № 5 (1957) 953-956)

При любом $n = 2, 3 \ldots$ существуют такие непрерывные на $[0,1]$ функции
$\psi^{p,q}(x)$, $1 \leqslant p \leqslant n$, $1 \leqslant q \leqslant 2n+1$,
что каждая функция n переменных $f(x_1, \ldots, x_n) \in C(I^n)$ представима в виде

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \chi_q \left(\sum_{p=1}^{n} \psi^{p,q}(x_p) \right),$$

где $\chi_q \in C(\mathbb{R})$, $1 \leqslant q \leqslant 2n+1$.

Kolmogorov's Mapping Neural Network Existence Theorem

Robert Hecht-Nielsen
Hecht-Nielsen Neurocomputer Corporation
5893 Oberlin Drive
San Diego, CA 92121
619-546-8877

Dedicated to Andrei Nikolaevic Kolmogorov

Abstract

An improved version of Kolmogorov's powerful 1957 theorem concerning the representation of arbitrary continuous functions from the n-dimensional cube to the real numbers in terms of one dimensional continuous functions is reinterpreted to yield an existence theorem for mapping neural networks.

1 Introduction

In 1957 Soviet mathematician A.N. Kolmogorov published an astounding theorem concerning the representation of arbitrary continuous functions from the n-dimensional cube to the real numbers in terms of one dimensional functions [2]. This theorem intrigued a number of mathematicians and over the next twenty years several improvements to it were discovered, notably by G. G. Lorentz [3,5].

The Kolmogorov theorem was discovered during a friendly mathematical duel between Kolmogorov and fellow Soviet mathematician V. I. Arnol'd in which they each tried to be the first to put to rest the remaining questions surrounding the 13th problem of Hilbert (a prominent mathematician who, at the turn of the century, announced a list of difficult problems for 20th century mathematicians to solve). In a series of papers in the mid to late 1950's Kolmogorov and Arnol'd fought their battle, each trying to one-up the other in successive papers. Kolmogorov won. His result was a mathematical supernova.

Although Kolmogorov's theorem was both powerful and shocking (many mathematicians do not believe it can be true when they first see it), it has not been found to be of much utility in terms of its use in proving other important theorems. In mathematical terms, no one has found a significant use for it. The point of this paper is that this is *not* the case in neurocomputing!

3 Kolmogorov's Theorem

In this section an improved version of Kolmogorov's theorem due to Sprecher is reexpressed as a result concerning the existence of mapping neural networks. The theorem follows:

Kolmogorov's Mapping Neural Network Existence Theorem: Given any continuous function
$\phi : I^n \longrightarrow \mathbf{R}^m$, $\phi(\mathbf{x}) = \mathbf{y}$, where I is the closed unit interval $[0,1]$ (and therefore I^n is the n-dimensional unit cube), ϕ can be implemented *exactly* by a three-layer neural network having n processing elements in the first (x - input) layer, (2n+1) processing elements in the middle layer, and m processing elements in the top (y - output) layer (see Figure 1).

The processing elements on the bottom layer are fanout units that simply distribute the input x-vector components to the processing elements of the second layer.

The processing elements of the second layer implement the following transfer function:

$$z_k = \sum_{j=1}^{n} \lambda^k \psi(x_j + \epsilon k) + k$$

where the real constant λ and the continuous real monotonic increasing function ψ are independent of ϕ (although they do depend on n) and the constant ϵ is a rational number $0 < \epsilon \leq \delta$, where δ is an arbitrarily chosen positive constant. Further, it can be shown that ψ can be chosen to satisfy a Lipschitz condition $|\psi(x) - \psi(y)| \leq c|x - y|^\alpha$ for any $0 < \alpha \leq 1$.

The m top layer processing elements have the following transfer functions:

$$y_i = \sum_{k=1}^{2n+1} g_i(z_k)$$

where the functions g_i $i = 1, 2, \ldots, m$ are real and continuous (and depend on ϕ and ϵ.

Robert Hecht-Nielsen (1947-2019), профессор университета Калифорнии (Сан-Диего), один из пионеров практического использования нейросетей для вычислений, автор первого учебника по теме «нейрокомпьютинг», лауреат премий IEEE Neural Networks Pioneer Award и International Gabor Award of the International Neural Network Society

Теорема Колмогорова
повлекла ряд исследований
как чисто теоретических,
так и прикладных

(G.G. Lorentz, D. Sprecher,
V. Kurkova, V. Maiorov,
A. Pinkus …)

Теорема (V. Maiorov, A. Pinkus)

Существует сигмоидная, действительно-аналитическая функция $\sigma(x)$, $x \in \mathbb{R}$, такая, что для каждой $f \in C(I^n)$, $n = 2, 3, \dots$, и каждого $\varepsilon > 0$ найдутся постоянные $d_i,\ c_{ij},\ \theta_{ij},\ \gamma_i$ и векторы $w^{ij} \in \mathbb{R}^n$, для которых

$$\left| f(x) - \sum_{i=1}^{6d+3} d_i \sigma \left(\sum_{j=1}^{3d} c_{ij} \sigma(w^{ij} \cdot x + \theta_{ij}) \right) + \gamma_i \right| \leqslant \varepsilon$$

для любой точки $x = (x_1, \dots , x_n) \in I^n$.

Теорема (G. Kozma, A. Olevskii, Труды МИАН, Т. 319 (2022) 134-181)

Для любой непрерывной функции $f \in C(T)$ существует абсолютно непрерывный гомеоморфизм $\varphi\colon T \to T$ такой, ряд Фурье суперпозиции $f(\varphi)$ сходится равномерно.

II) В самом начале XX-века Лебегом была создана теория интеграла и введено пространство *L(K)* -- функций, заданных на некотором множестве *K*, интегрируемых по Лебегу.

Сразу же встал вопрос об основных свойствах рядов Фурье функций из *L(T)*, в первую очередь об их сходимости.

При этом интересно отметить, что еще в 70-е годы XIX века была поставлена задача о существовании точек сходимости ряда Фурье произвольной функции из *C(T)*.

Итак, пусть $f \in L(T)$,

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx \qquad (1)$$

– ряд Фурье функции f,

$$S_n(f, x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kx + b_k \sin kx, \quad n = 1, 2, \ldots,$$

– частные суммы.

Лебег (1909) показал,
что для любой функции $f \in L(T)$

$$\frac{S_1(f,x)+\cdots+S_n(f,x)}{n} \xrightarrow[n\to\infty]{} f(x)$$

для п.в. $x \in T$.

Харди (1913): если $f \in L(T)$, то

$$|S_n(f,x)| = o(\ln n) \quad \text{для п.в. } x \in T.$$

Первая опубликованная работа А.Н. Колмогорова (рукопись подписана автором: Москва, 2 июня 1922 г.) внесла фундаментальный вклад в теорию тригонометрических рядов.

Теорема (А.Н. Колмогоров, Fundam. Math., 4 (1923) 324-328)

Существует функция $f \in L(T)$ такая, что

$$\overline{\lim_{n\to\infty}} |S_n(f,x)| = \infty \quad \text{для п.в.} \quad x \in T.$$

Важные результаты в связи с теоремой Колмогорова

были получены С.В. Бочкаревым (1975, 2000)

и С.В. Конягиным (2000)

Теорема (С.В. Конягин, 2000)

Если $\lambda_n = o\left(\frac{\ln^{1/2} n}{\log\log n}\right)$ при $n \to \infty$,

то существует $f \in L(T)$ такая, что

$$\varlimsup_{n\to\infty} \frac{|S_n(f,x)|}{\lambda_n} = \infty \quad \text{для п.в.} \quad x \in T.$$

III) Сопряженным к ряду (1) называется ряд

$$\sum_{k=1}^{\infty} -b_k \cos kx + a_k \sin kx \qquad (2)$$

Ряды (1) и (2) – действительная и мнимая части степенного ряда

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k - ib_k) z^k \qquad (3)$$

И. И. Привалов (1918) доказал, что в случае, когда действительная часть ряда (3) на T – есть функция $f \in L(T)$, его мнимая часть восстанавливается на T интегралом

$$g(\theta) = -\frac{1}{2\pi} \int_T \frac{f(\theta + t)}{\operatorname{tg} \frac{t}{2}}\, dt, \qquad (4)$$

т.е., в частности, интеграл (4) существует (в смысле главного значения) для п.в. $\theta \in T$.

Теорема (А.Н. Колмогоров, Fundam. Math., 7 (1925) 24-29)

Для любого $y > 0$

$$\operatorname{meas}\{\theta \in T\colon |g(\theta)| \geqslant y\} \leqslant \frac{C\|f\|_{L(\pi)}}{y} \qquad (5)$$

Из (5) А. Н. Колмогоров вывел, что ряды (1), (2) сходятся в $L^p(T)$ при любом $p < 1$, если $f \in L(T)$.

Оценка «слабого типа» (5) –
центральный результат о сходимости
тригонометрических рядов
в интегральных метриках,
непосредственно повлиявший на
развитие теории функций
и функционального анализа в XX веке:
классические теоремы об операторах
слабого типа, интерполяционная
теорема Марцинкевича, теоремы
факторизации (Гротендик, Стейн,
Никишин …)

IV) **Поперечник по Колмогорову**

Пусть G – линейное метрическое пространство с метрикой ρ, F – подмножество в G.

Определение (А.Н. Колмогоров, Annals of Math., 37 (1936) 107-110)

Для $n = 0, 1, \ldots$ *n-поперечником* множества F в G называется величина

$$d_n(F, G) = \inf_{L_n} \sup_{f \in F} \rho(f, L_n),$$

где внешний inf берется по всем подпространствам G размерности $\leqslant n$.

Понятие поперечника в последние десятилетия стало одним из важнейших в теории приближений.

Причина: расширение возможностей компьютеров, что сделало целесообразным использование линейных подпространств общего вида в качестве аппарата приближения.

Важные практические приложения – «сжатые измерения» (compressed sensing), восстановлений функций по дискретной информации, …

В последние годы выяснилось,
что ряд важных задач
дискретной математики
– это задачи об оценках
поперечников, в том числе
в пространствах Хемминга.

Пример

ε-ранг $N \times N$ матрицы $A = \{a_{ij}\}_{i,j=1}^{N}$:

$$\operatorname{rank}_\varepsilon = \min\{r\colon \exists B = \{b_{ij}\}_{i,j=1}^{N}, \operatorname{rank} B \leqslant r, \\ |b_{ij} - a_{ij}| \leqslant \varepsilon, 1 \leqslant i, j \leqslant N\}.$$

Если $W = \{w_i\}_{i=1}^{N} \subset \mathbb{R}^N$ – набор строк матрицы A, то легко видеть, что

$$\operatorname{rank}_\varepsilon(A) = \min\{r\colon d_r(W, l_\infty^N) \leqslant \varepsilon\}.$$

Пусть $A_N = \{a_{ij}\}_{i,j=1}^N$ таково, что

$$a_{ij} = \begin{cases} 1, & \text{если } i \leqslant j, \\ 0, & \text{иначе.} \end{cases}$$

Известно, что

$$C_1 \log^2 N \leqslant \mathrm{rank}_{1/3}(A_N) \leqslant C_2 \log^3 N.$$

Задача

Определить порядок величины $\mathrm{rank}_{1/3}(A_N)$ при $N \to \infty$.

V) **«Chaining» Колмогорова**

Результат, содержащий основную идею метода, был получен А.Н. Колмогоровым в 1934-м году, но опубликован, с его согласия, Е.Е. Слуцким только в 1937-м году

(E. Slutsky, Giorn. Inst. Ital. degli Attuovi, VIII (1937))

Теорема (А.Н. Колмогоров)

Пусть $y(t)$ есть случайная функция, определенная в замкнутом интервале $[a,b]$, и пусть для $a \leqslant t \leqslant b$ имеет место условие

$$\mathrm{E}|y(t) - y(t')|^m \leqslant C|t - t'|^\alpha \qquad (m > 0, \quad \alpha > 1).$$

Тогда $y(t)$ стохастически эквивалентна некоторой функции $Z(t)$ непрерывной в том же интервале $[a,b]$.

Простая идея доказательства этой теоремы лежит в основе важнейших современных результатов об оценках максимума случайных процессов.