УДК 004.852

ТЕОРЕМИ КОЛМОГОРОВА-АРНОЛЬДА-ЛОРЕНЦА В ЗАДАЧІ ОБҐРУНТУВАННЯ ПРАЦЕЗДАТНОСТІ ШТУЧНИХ НЕЙРОМЕРЕЖ

Т.В.Павлов

Міжнародний науково-навчальний центр інформаційних технологій і систем НАН та МОН

pawlovtaras@gmail.com

В роботі розглянуто серію теорем Колмогорова-Арнольда-Лоренца, з допомогою яких було розв‘язано 13 проблему Гільберта про представлення функцій багатьох змінних. Проаналізовано доцільність використання цих теорем на основі робіт А.Г.Вітушкіна і спільної статті Ф.Джиросі та Т.Поджіо. Окремо розглянуто випадок класу многочленів, за основу взято роботи А.Н. Горбаня.
Ключові слова: представлення функцій багатьох змінних, нейромережі, суперпозиція, теорема Колмогорова-Арнольда

The paper consider a series of theorems of Kolmogorov-Arnold-Lorents related to 13-th Gilbert's problem about representation of function of many variables. Analyzed the feasibility of using these theorems based on collaborative article of F.Girosi and T.Poggio, and on works of A.G.Vitushkin. Separately considered the case of class of polynomials, based on article of A.N.Gorban.
Key words: representation of function of many variables, neural networks, superposition, theorem of Kolmogorov-Lorentz

В работе рассмотрено ряд теорем Колмогорова-Арнольда-Лоренца, с помощью которых была решена 13 проблема Гильберта о представлении фукнций от многих переменных. С помощью работ А.Г.Витушкина и статьи Ф.Джироси и Т.Поджио проанализировано целесообразность использования этих теорем. Отдельно рассмотрен случай с классом полиномов, который решен в работах А.Н.Горбаня.
Ключевые слова: представление функций многих переменных, нейросети, суперпозиция, теорема Колмогорова-Арнольда

1. Вступ

Важливим моментом в теорії апроксимацій є вибір представлення функції. Оскільки будь-яке представлення може бути зображено з допомогою нейромережі, вибір представлення еквівалентний вибору мереж з конкретною архітектурою. Рожевою мрією математиків є отримання певної універсальної архітектури нейромереж, якомога простішої, які представляли б довільну функцію певного класу, знову ж таки, якомога ширшого.

В кінці 80-их, завдяки відкриттю Колмогорова(1957) [2], було зроблено спробу обґрунтувати з його допомогою використання нейромереж з прямими повними звязками і двома "прихованими"шарами для представлення довільної неперервної функції. Хехт-Нільсен [8] довів, що така можливість і справді існує. Однак власне використання цього методу представлення піддали жорсткій критиці Джиросі і Поджіо[3], з допомогою теореми Вітушкіна [12] ними було показано, що представлення, хоч і існує, досить складне і в більшості випадків не відповідає вимогам які ставлять перед ним задачі теорії апроксимацій.

2. Представлення неперервних функцій

1900 рік: На своїй історичній лекції в Парижі Гільберт сформулював 23 основні завдання, які, на його думку, стоять перед математиками в новому столітті. 13 питання стосувалось алгебраїчних рівнянь вигляду:

$$a_n x^n + \ldots + a_1 x + a_0 = 0$$

Відомо, що подібні рівняння до 4 степеня включно розвязуються з допомогою лише алгебраїчних операцій. З теорії Галуа відомо що для рівнянь степеня $n \geq 5$ це не завжди можливо. Оскільки алгебраїчні операції - це операції від максимум двох аргументів, то Гільберт і прийшов до припущення, що:

Твердження 1 (13 проблема Гільберта). *Існує неперервна функція 3 змінних, яку не можна представити як суперпозицію неперервних функцій двох змінних.*

1957 рік: Андрій Миколайович Колмогоров (в співпраці з своїм учнем Володимиром Ігоревичем Арнольдом) спростував припущення Гільберта. Доведено теорему:

Теорема 1 (Колмогорова про суперпозицію функцій). [2] *Довільну неперервну функцію декількох змінних можна представити з допомогою суперпозиції функцій однієї змінної і операції додавання.*

Теорема доводилась поступово, протягом декількох років доведено було серію теорем:

Теорема 2 (Колмогоров). *Довільна неперервна функція більш ніж трьох змінних є суперпозицією неперервних функцій трьох змінних.*

Володимир Арнольд, в свою чергу в [6] довів, що будь-яку функцію трьох змінних можна представити як суперпозицію від функцій двох змінних, тим самим спростувавши припущення Гільберта. Пізніше Колмогоров підсилив теорему, довівши, що:

Теорема 3 (Суперпозиційна теорема Колмогорова). [2] *Для будь-якого $n \geq 1$ існує набір $(2n+1)n$ таких неперервних функцій $u_i j : I \to I (i = 1, \ldots, 2n+1, j = 1, \ldots, n)$ однієї змінної,* *що для будь-якої неперервної функції $f : I^n \to I$ n змінних існують неперервні функції $f_1, \ldots, f^I_{2n+1} \to I$ однієї змінної, для яких:*

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{2n+1} f_i(\sum_{j=1}^{n} u_{ij}(x_j))$$

Оригінальне доведення Колмогорова не було конструктивним.

1961 - 1965 рр. - Тіхоміров і Остранд розширили теорему Колмогорова до довільного n–вимірного компактного метричного простору.

1966 рік: Лоренц знайшов алгоритм розкладу функції в суперпозицію:

Теорема 4 (Колмогорова-Лоренца)**.** [4] *Для всіх $n \geq 2$ існують неперервні функції $\phi_q : [0,1] \to \mathbb{R}, q = 0, \ldots, 2n$ і константи $\lambda_p \in \mathbb{R}, p = 1, \ldots, n$ такі, що для будь-якої неперервної функції $f : [0,1]^n \to \mathbb{R}$ існує неперервна функція $g : [0,1] \to \mathbb{R}$ така, що:*

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} g(\sum_{p=1}^{n} \lambda_p \phi_q(x_p)).$$

Саме цей варіант теореми(враховуючи уточнення Шпрехера) використовується в сучасній математиці.

1967 рік - Фрідман показав, що внутрішні функції можна вибрати так, щоб вони були неперервними за Ліпшицом. [11]

1971 рік - Хедберг зробив доведення теореми Колмогорова-Лоренца, що базується на теоремі категорій Бера.

1972 рік - Шпрехер помітив, що функції ϕ і константи λ не залежать від вибору функції, а лише від її степеня.

Хоча для багатьох класів функцій теорема Колмогорова-Лоренца не виконується, але правильне наступне твердження:

Теорема 5 (Колмогорова про суперпозицію обчислюваних функцій)**.** [4] *Для всіх $n \geq 2$ існують обчислювані функції $\phi_q : [0,1] \to \mathbb{R}, q = 0, \ldots, 2n$ і обчислювані константи $\lambda_p \in \mathbb{R}, p = 1, \ldots, n$ такі, що для будь-якої неперервної функції $f : [0,1]^n \to \mathbb{R}$ існує неперервна функція $g : [0,1] \to \mathbb{R}$ така, що:*

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} g(\sum_{p=1}^{n} \lambda_p \phi_q(x_p)).$$

Якщо функція f обчислювана то і g буде обчислюваною.

1987 рік: Хехт-Нільсен помітив цікаве застосування цієї теореми в теорії нейромереж. Було доведено теорему:

Теорема 6 (Хехт-Нільсена)**.** [8] *Функцію багатьох змінних $f : \mathbb{R}^n \to \mathbb{R}^m$ можна представити з допомогою двошарової нейромережі з прямими повними звязками з n компонентами вхідного сигналу, $2n + 1$ компонентами прихованого шару з наперед відомими визначеними функціями активації (наприклад, сигмоїдальними) і m компонентами другого шару з невідомими функціями активації.*

Теорема, таким чином, в неконструктивній формі доводить розвязність задачі представлення функції і вказує для кожної задачі мінімальні значення кількості нейронів мережі, необхідних для розв'язку.

Однак з використанням теореми Колмогорова в нейромережах на практиці виникають значні труднощі. В роботі [3] Федеріко Джіросі і Томазо Поджіо піддали нищівній критиці використання мереж, створених на основі теореми Колмогорова. Їх два основні аргументи: По-перше в представленні нейромереж, що будуть використовуватись для навчання деяким функціям необхідний певний порядок гладкості відповідно до вимог нейромережі. Внутрішні функції ϕ_q з теореми Колмогорова дуже негладкі. Це твердження базується на теоремі Вітушкіна:

Наслідок 1 (з теореми Вітушкіна)**.** [15] *Існують* $r(r = 1, 2, \ldots)$ *раз неперервно-диференційовні функції* $n \geq 2$ *змінних, які не можна представити як суперпозицію r раз неперервно диференційовних з менш ніж n змінних. Існують r раз неперервно-диференційовні функції двох змінних, які не можна представити з допомогою додавання і неперервно-диференційовні функції однієї змінної.*

По-друге для задач навчання чи апроксимації зручними є параметричні представлення, що відповідають нейромережам з фіксованими нейронами і параметрами, які можна модифікувати. Мережа ж Колмогорова не є такою - вигляд функцій g_q (що відповідають нейронам з другого прихованого шару) залежить від особливостей функції f.

3. Представлення многочленів

В випадку функцій-многочленів ситуація значно простіша і питання про представлення повністю розв'язане А.Н.Горбанем.

Теорема 7. *Нехай R - кільце многочленів над полем характеристики* 0*. Нехай p - многочлен від одної змінної,* E_p *- множина многочленів, які можна отримати з всіх многочленів першого степеня з R, суперпозиції, додавання та p. Тоді якщо степінь p більше першого то* $E_p = R$*.*

Тобто з допомогою додавання, множення та суперпозиції отримуємо довільний елемент кільця многочленів. Враховуючи теорему Стоуна-Вейєрштрасса таким чином маємо можливість побудувати наближене з довільною точністю представлення будь-якої неперервної функції. Оскільки багато алгоритмів МГУА базуються на поліноміальному представленні планується в майбутньому розглянути їх з цієї точки зору.

4. Висновки

Клас диференційовних функцій: Згідно теоремі Вітушкіна точне представлення з допомогою суперпозиції функцій меншої кількості змінних неможливе без втрати класу гладкості.
Клас многочленів: Маємо точне представлення для довільного многочлена від декількох змінних.
Клас неперервних функцій: З допомогою теорем Колмогорова-Арнольда-Лоренца маємо точне представлення довільної неперервної функцій. Але враховуючи сказане вище гладкості в функціях, через які виражається добитись неможливо. В той же час, завдяки теоремі Стоуна-Вейєрштрасса наближене з довільною точністю представлення через поліноми можливе.

Тож в задачах, де гладкість функції не вимагається і необхідне будь-яке представлення в вигляді нейромережі - маємо універсальний метод побудови представлення неперервних функцій. Водночас, теорема Вітушкіна є теоремою існування а не загальності, тож, імовірно, є певні класи функцій, в яких проблем з негладкістю внутрішніх функцій не буде, що зберігає його корисність як надійного методу в часткових випадках.

Окрім того там, де це можливо, результат можна отримати знизивши вимоги до точності представлення функції.

Література

[1] Колмогоров А.Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных, Докл. АН СССР, том 108, с. 2, 1956.

[2] Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // Докл. АН СССР, том 114, с. 953-956, 1957

[3] F. Girosi and T. Poggio Representation properties of networks: Kolmogorov's theorem is irrelevant. Neural Comp., 1:465 – 469, 1989.

[4] Vasco Brattka, A computable Kolmogorov Superposition Theorem

[5] V. Arnold. On the representation of functions of several variables by superpositions of functions of fewer variables. Mat. Prosveshchenie, 3:41 – 61, 1958.

[6] V. Arnold. On functions of three variables. Dokl. Akad. Nauk SSSR, 114:679 – 681, 1957. English translation: American Math. Soc. Transl (2), 28, 51 - 54, 1963.

[7] G. Lorentz. Approximation of functions. Holt, Rinehart, Winston, 1966

[8] Hecht-Nielsen R. Kolmogorov's Mapping Neural Network Existence Theorem // IEEE First Annual Int. Conf. on Neural Networks, San Diego, 1987, Vol. 3, pp. 11-13

[9] V. Kurkova. Kolmogorov's theorem is relevant. Neural Computation, 3:617 – 622, 1991

[10] Muller B., Reinhart J. Neural Networks: an introduction, Springer-Verlag, Berlin Heidelberg, 1990

[11] B. Fridman An improvement on the smoothness of the functions in Kolmogorov's theorem on superpositions. Dokl. Akad. Nauk SSSR, 177:1019–1022, 1967. English translation: Soviet Math. Dokl. (8), 1550-1553, 1967

[12] A.G.Vitushkin On Hilbert's thirteenth problem. Dokl. Akad. Nauk SSSR 95, 701-704, 1954

[13] A.G.Vitushkin Some properties of linear superposition of smooth functions. Dokl. Akad. Nauk SSSR 156: 1003-1006, 1964

[14] A.G.Vitushkin On Representation of functions by means of Superpositions and related topics. L'Enseignement Matematique, 1977

[15] A.G.Vitushkin, G.M.Henkin Linear superposition of functions. Russian Math.Surveys 22, 77-125, 1967

[16] А.Скопенков 13-я проблема Гильберта и базисные вложения

[17] А.Н.Горбань "Обобщенная апроксимационная теорема и точное представление многочленов от нескольких переменных суперпозициями многочленов от одного переменного Известия высших учебных заведений (математика), 1998

[18] А.Н.Горбань "Обобщенная апроксимационная теорема и вычислительные возможности нейронных сетей"