УДК 517.5

АППРОКСИМАЦИЯ ФУНКЦИЙ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ И НЕЧЕТКИХ СИСТЕМ

А.С. Шведов

Дан обзор некоторых результатов, относящихся к аппроксимации функций одного и нескольких действительных переменных. Приведен ряд теорем одного из классических направлений — аппроксимации алгебраическими многочленами. Отмечено, что первые результаты по аппроксимации функций с помощью нейронных сетей и нечетких систем появились как ответ на необходимый для практических задач вопрос о возможности приближенного представления непрерывных функций с помощью таких агрегатов. Затем эти научные направления стали развиваться теми же путями, что и теория приближения функций алгебраическими многочленами. Представлены некоторые результаты, относящиеся к аппроксимации функций нейронными сетями и нечеткими системами.

Ключевые слова: аппроксимация функций, алгебраические многочлены, нейронные сети, нечеткие системы.

ВВЕДЕНИЕ

Нейронные сети и нечеткие системы применяются для решения многих прикладных задач (см., например, книги [1—3]), в том числе и задач управления. Теоремы о возможности аппроксимации непрерывных функций данными агрегатами стали важным шагом на пути превращения и нейронных сетей, и нечетких систем из эвристических методов в математически обоснованные методы.

Основополагающие результаты по теории приближения функций получены в девятнадцатом веке К. Вейерштрассом и П.Л. Чебышевым. Их исследования были продолжены многими авторами. В двадцатом веке большой вклад в теорию приближений внесли академики А.Н. Колмогоров и С.Н. Бернштейн. При создании этой теории в качестве аппроксимирующих агрегатов использовались алгебраические многочлены (для непериодических функций) и тригонометрические полиномы (для периодических функций). Результаты из этих двух разделов теории приближений во многом похожи. Затем стали применяться и другие аппроксимирующие агрегаты — сплайны, нейронные сети, нечеткие системы. Многие интересные результаты связаны с теорией поперечников функциональных классов.

Первые теоремы о возможности аппроксимации непрерывных функций однонаправленными нейронными сетями с одним скрытым слоем получены в работах[1] [4—6]. Первые теоремы о возможности аппроксимации функций нечеткими системами доказаны в работах [7—9]. Исследования и по аппроксимации функций нейронными сетями, и по аппроксимации функций нечеткими системами были продолжены многими авторами.

Существуют и нейро-нечеткие модели, но они в настоящем обзоре не рассматриваются.

В § 1 данной работы приводятся некоторые результаты, относящиеся к приближению функций одного и нескольких действительных переменных алгебраическими многочленами, он носит вспомогательный характер. Его основное назначение — показать те постановки задач, которые затем встречаются в § 2 и 3. Обзор современного состояния теории приближения функций алгебраическими многочленами в настоящей работе не дается. В § 2 дается обзор результатов по приближению функций нейронными сетями, в § 3 — нечеткими системами.

[1] В книге [2] для термина *feedforward neural network* предлагается перевод *нейронная сеть прямого распространения.* Однако более удачным представляется перевод *однонаправленная нейронная сеть*, употребляемый в книге [3].

1. АППРОКСИМАЦИЯ ФУНКЦИЙ АЛГЕБРАИЧЕСКИМИ МНОГОЧЛЕНАМИ

Пусть $C[0, 1]$ — пространство непрерывных функций $f: [0, 1] \to \mathbb{R}$. Норма функции $f \in C[0, 1]$ определяется соотношением

$$\|f\|_\infty = \sup_{x \in [0, 1]} |f(x)|.$$

Через Π_n обозначим пространство алгебраических многочленов степени не выше n, т. е. пространство функций вида

$$P_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n,$$

где $a_0, a_1, \ldots, a_{n-1}, a_n$ — действительные числа. Наилучшим приближением функции $f \in C[0, 1]$ алгебраическими многочленами степени не выше n называется величина

$$E_n(f) = \inf_{P_n \in \Pi_n} \|f - P_n\|_\infty.$$

Следующая теорема принадлежит Вейерштрассу.

Теорема 1. *Для любой функции* $f \in C[0, 1]$

$$E_n(f) \to 0 \text{ при } n \to \infty. \ ♦$$

Доказательство теоремы Вейерштрасса можно найти, например, в книгах [10, 11].

Модуль непрерывности функции $g \in C[0, 1]$ определяется следующим образом. При $\delta > 0$

$$\omega(g, \delta) = \sup_{\substack{x, y \in [0, 1] \\ |x - y| \le \delta}} |g(x) - g(y)|.$$

В терминах модулей непрерывности функции f и ее производных (если они существуют) можно дать ответ на вопрос, насколько быстро последовательность $E_n(f)$ стремится к нулю в зависимости от гладкости функции f.

Следующие две теоремы принадлежат Джексону.

Теорема 2. *Если функция* $f \in C[0, 1]$, *то при любом натуральном* n

$$E_n(f) \le C_0 \omega\left(f, \frac{1}{n}\right),$$

где C_0 *— некоторая абсолютная константа.*

Теорема 3. *Если функция* $f \in C[0, 1]$ *имеет* r *непрерывных производных на отрезке* $[0, 1]$, *то при любом* $n > r$

$$E_n(f) \le \frac{C_r}{n^r} \omega\left(f^{(r)}, \frac{1}{n - r}\right),$$

где константа C_r *зависит только от* r. ♦

Доказательства теорем 2 и 3 можно найти, например, в книге [10]. Формально теорема 1 является следствием теоремы 2, но получена теорема 1 значительно раньше.

Теоремы 2 и 3 относятся к прямым теоремам теории приближений. Существуют и обратные теоремы теории приближений: из достаточно хорошей аппроксимации функции теми или иными агрегатами вытекает соответствующая гладкость этой функции. Про обратные теоремы для приближения алгебраическими многочленами см., например, книгу [11].

Однако теорема 3 не дает ответа на следующий вопрос. Пусть функция $f \in C[0, 1]$ и имеет r непрерывных производных на отрезке $[0, 1]$. Можно ли найти такой алгебраический многочлен P_n, который приближает функцию f, а производные этого многочлена $P'_n, \ldots, P_n^{(r)}$ приближают производные $f', \ldots, f^{(r)}$ соответственно? Такие задачи называются задачами о совместном приближении алгебраическими многочленами функции и ее производных.

Следующие многочлены были введены С.Н. Бернштейном. Для функции $f \in C[0, 1]$ алгебраический многочлен

$$B_n(f; x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k x^k (1 - x)^{n-k}$$

называется многочленом Бернштейна. Доказательство того, что многочлены Бернштейна дают решение задачи о совместном приближении функции и ее производных, можно найти, например, в книге [12]. Однако скорость приближения функции многочленами Бернштейна значительно более медленная, чем та, которая дается теоремой 2 (см., например, работу [13]).

Многочлены Бернштейна дают пример линейного метода приближения. Оператор L_n: $C[0, 1] \to C[0, 1]$ называется линейным, если для любых функций $f, g \in C[0, 1]$ и для любых $a, b \in \mathbb{R}$

$$L_n(af + bg) = aL_n(f) + bL_n(g).$$

Линейный оператор L_n называется положительным, если из условия $f(x) \ge 0$ при $0 \le x \le 1$ вытекает, что $L_n(f; x) \ge 0$ при $0 \le x \le 1$.

Следующая теорема принадлежит П.П. Коровкину [14].

Теорема 4. *Пусть* $\{L_n\}$ *— последовательность линейных положительных операторов из* $C[0, 1]$ *в* $C[0, 1]$. *Рассмотрим функции* $g_k(x) = x^k$. *Тогда если*

$$\|g_k - L_n(g_k)\|_\infty \to 0 \text{ при } n \to \infty,$$

при $k = 0, 1, 2,$ *то для любой функции* $f \in C[0, 1]$

$$\|f - L_n(f)\|_\infty \to 0 \text{ при } n \to \infty. \ \blacklozenge$$

Существуют многочисленные обобщения теоремы Коровкина (см., например, работы [15, 16]).

Задача приближения функций алгебраическими многочленами изучена не только для равномерной метрики, но и для интегральных метрик.

Пространство $L_p[0, 1]$, $1 \le p < \infty$, состоит из измеримых по Лебегу функций $f: [0, 1] \to \mathbb{R}$, для которых

$$\int_0^1 |f(x)|^p dx < \infty.$$

Функции, различающиеся лишь на множестве меры ноль, отождествляются. Норма функции $f \in L^p[0, 1]$ определяется следующим образом:

$$\|f\|_p = \left(\int_0^1 |f(x)|^p dx\right)^{1/p}.$$

Для наилучшего приближения функций $f \in L^p[0, 1]$ алгебраическими многочленами в метрике L^p также имеются теоремы типа Джексона (см., например, работу [17]).

Значительная часть результатов о приближении функций одного действительного переменного в той или иной степени обобщена для функций нескольких действительных переменных (см., например, книгу [18]).

Рассмотрим компактное множество $K \subseteq \mathbb{R}^d$ (т. е. множество K является ограниченным и замкнутым). Пусть $C(K)$ — пространство непрерывных функций $f: K \to \mathbb{R}$ с равномерной нормой

$$\|f\|_\infty = \sup_{x \in K} |f(x)|.$$

Пусть M — линейное подпространство пространства $C(K)$, и выполняются следующие три условия.

1. Подпространство M является алгеброй, т. е. из условия $f \in M$, $g \in M$ следует, что и произведение $fg \in M$.

2. Для любых несовпадающих точек $x, y \in K$ найдется функция $f \in M$ такая, что $f(x) \ne f(y)$.

3. Для любой точки $x \in K$ найдется функция $f \in M$ такая, что $f(x) \ne 0$.

Следующая теорема принадлежит Стоуну.

Теорема 5. *Подпространство* M *является плотным в пространстве* $C(K)$ *относительно равномерной метрики.* ♦

Доказательство теоремы Стоуна можно найти, например, в книге [11].

Отметим, что пространство всех алгебраических многочленов относительно переменных $x_1, ..., x_d$ удовлетворяет условиям 1—3. Поэтому верен многомерный аналог теоремы 1: любую функцию из пространства $C(K)$ можно сколь угодно точно приблизить алгебраическими многочленами в равномерной метрике.

Один из вариантов теоремы Джексона для приближения функций нескольких действительных переменных алгебраическими многочленами дается в работе [19].

Многомерные многочлены Бернштейна имеют вид

$$B_n(f; x_1, ..., x_d) =$$

$$= \sum_{k_1, ..., k_d = 0}^{n} f\left(\frac{k_1}{n}, \frac{k_2}{n}, ..., \frac{k_d}{n}\right) C_n^{k_1} C_n^{k_2} \ldots C_n^{k_d} \times$$

$$\times x_1^{k_1}(1 - x_1)^{n - k_1} x_2^{k_2}(1 - x_2)^{n - k_2} \ldots x_d^{k_d}(1 - x_d)^{n - k_d}.$$

Результаты о совместном приближении функции нескольких действительных переменных и ее частных производных многомерными многочленами Бернштейна устанавливаются в работе [20].

По направлениям, обозначенным в данном параграфе, развиваются и теория приближения нейронными сетями, и теория приближения нечеткими системами. Преимущества этих подходов заключаются в большей однородности аппроксимирующих агрегатов, чем та, которая есть у одночленов $x_1^{k_1}, ..., x_d^{k_d}$ при различных $k_1, ..., k_d$, а также в меньшей зависимости от конкретной функциональной формы.

Отметим, что в теории приближения алгебраическими многочленами задаче численного нахождения коэффициентов аппроксимирующего многочлена уделяется относительно небольшое внимание. Наиболее известен алгоритм Ремеза (см., например, книгу [21]). Интерес к этим вопросам оказывается значительно бóльшим, когда в качестве аппроксимирующих агрегатов применяются нейронные сети или нечеткие системы.

2. АППРОКСИМАЦИЯ ФУНКЦИЙ НЕЙРОННЫМИ СЕТЯМИ

Рассмотрим однонаправленную нейронную сеть типа MISO (multiple input single output) с одним скрытым слоем, состоящим из n нейронов. Входом является $x = (x_1, ..., x_d)$, выходом является

$$f_n(x) = \sum_{j=1}^{n} \beta_j \varphi(w_j x + \theta_j). \qquad (1)$$

Здесь $\varphi: \mathbb{R} \to \mathbb{R}$ — функция активации; $w_j \in \mathbb{R}^d$, $\theta_j \in \mathbb{R}$ — внутренние параметры для j-го нейрона; $\beta_j \in \mathbb{R}$; $w_j x$ — скалярное произведение векторов w_j и x.

В качестве функций активации часто используются ступенчатая функции $\varphi(t) = 0$ при $t < 0$, $\varphi(t) = 1$ при $t \geq 0$; функция $\varphi(t) = \dfrac{1}{1+e^{-t}}$; гиперболический тангенс $\varphi(t) = \dfrac{e^t - e^{-t}}{e^t + e^{-t}}$.

Функция φ называется сигмоидальной, если $\varphi(t) \to 0$ при $t \to -\infty$, $\varphi(t) \to 1$ при $t \to \infty$. Первоначально к этим двум условиям добавлялось условие монотонного неубывания функции φ; тогда график этой функции напоминает букву S, откуда и происходит название «сигмоидальная». Однако во многих теоретических результатах условие, что функция φ монотонная, оказывается лишним.

Следующая теорема принадлежит авторам работ [4—6].

Теорема 6. *Пусть сигмоидальная функция* φ *непрерывна и ограничена на* $\mathbb{R}$. *Тогда множество сумм вида* (1) *плотно в пространстве* $C([0, 1]^d)$. *Другими словами, для любой функции* $f \in C([0, 1]^d)$ *и для любого* $\varepsilon > 0$ *при некотором натуральном* n *существует функция* f_n *вида* (1) *такая, что*

$$|f(x) - f_n(x)| < \varepsilon$$

при всех $x \in [0, 1]^d$. ♦

Обобщения теоремы 6 для различных классов функций активации даются в ряде работ (см., например, работу [22]). Но очевидно, что результат, аналогичный теореме 6, не может быть верен, если φ — алгебраический многочлен.

Может быть рассмотрена однонаправленная нейронная сеть с одним скрытым слоем, для которой

$$f_n(x) = \sum_{j=1}^{n} \beta_j \Phi\left(\frac{x - z_j}{\sigma_j}\right). \tag{2}$$

Функция активации $\Phi: \mathbb{R}^d \to \mathbb{R}$; $z_j \in \mathbb{R}^d$, $\sigma_j > 0$, $\beta_j \in \mathbb{R}$. Если имеет место соотношение

$$\Phi(z) = g(\|z\|), \quad z \in \mathbb{R}^d,$$

где $g: \mathbb{R}_+ \to \mathbb{R}$, $\|\cdot\|$ — евклидова норма в $\mathbb{R}^d$, то нейронную сеть называют сетью с радиальными базисными функциями.

Следующая теорема доказана в работе [23].

Теорема 7. *Пусть функция* Φ *является интегрируемой, ограниченной, почти всюду непрерывной на* $\mathbb{R}^d$ *и*

$$\int_{\mathbb{R}^d} \Phi(x)dx \neq 0.$$

Тогда при любом $p \in [1, \infty)$ *семейство функций вида* (2)*, где* n *может быть произвольным натуральным числом, плотно в* $L^p(\mathbb{R}^d)$. ♦

Теорема 7 остается верной (см. [23]) и если вместо семейства функций (2) рассмотреть более узкое семейство, для которого $\sigma_1 = \sigma_2 = \ldots = \sigma_n$. Это замечание существенно, поскольку при обучении нейронной сети расчет внутренних параметров оказывается более трудной задачей, чем расчет весов $\beta_1, \beta_2, \ldots, \beta_n$.

В работах [24, 25] вопрос об уменьшении числа внутренних параметров рассматривается для сигмоидальной функции активации. Доказывается теорема того же вида, что и теорема 6, но при этом внутренние веса не зависят от аппроксимируемой функции.

Разумеется, представляет интерес вопрос, как зависит точность приближения функции от числа нейронов и от свойств самой функции (сравните теоремы 2 и 3). Такие задачи рассматриваются, например, в работе [26].

Установлено, что для некоторых задач аппроксимации разрывных функций нескольких переменных в метриках L^p, $1 \leq p < \infty$, нейронные сети с двумя скрытыми слоями могут обладать лучшими аппроксимационными свойствами, чем нейронные сети с одним скрытым слоем [27].

В работе [28] рассматривается понятие почти равномерной сходимости для разрывных функций, определенных на отрезке. Доказывается теорема типа Вейерштрасса для нейронных сетей вида (1).

В работе [29] рассматривается задача интерполяции нейронными сетями вида (2) с радиальными базисными функциями. Доказывается существование решения у этой задачи.

Аппроксимация периодических функций d переменных нейронными сетями вида (1) изучается в работе [30], а также в ряде других работ.

Оказывается, что вейвлет нейронные сети обладают некоторыми преимуществами с точки зрения аппроксимации функций (см., например, работу [31]).

В работе [32] проводится сравнение аппроксимации функций нейронными сетями и аппроксимации функций алгебраическими многочленами.

Во многих публикациях изучается совместное приближение функции и ее производных нейронными сетями. Некоторый обзор этого направления дается в работе [33].

Упомянем также обзоры [34, 35] по аппроксимации функций с помощью нейронных сетей.

Существует большое число работ практической направленности и работ, в которых рассматриваются различные подходы к обучению нейронных сетей. Под обучением нейронной сети понимается нахождение параметров в формулах (1) или (2). В работе [36] сравнивается аппроксимация с помощью нейронных сетей с другими применяемыми на практике методами приближения функций. В работе [37] описаны численные эксперименты с различными функциями активации. В работе [38] представлены некоторые приемы, позволяющие улучшить аппроксимацию кусочно-непрерывных функций одного переменного. В работе [39] для обучения нейронной сети используются методы робастной статистики.

Для обучения нейронной сети применяются генетические алгоритмы (см., например, книгу [3]). Применяются и другие современные методы нахождения экстремумов функций многих переменных. Однако в работе [40] присутствует полемика с таким подходом. Там утверждается, что традиционные методы, основанные на решении алгебраических уравнений, могут дать лучшие результаты, если при обучении нейронной сети используются не только информация о значениях аппроксимируемой функции, но и информация об ее градиентах.

Теорема Колмогорова о представлении непрерывных функций нескольких переменных с помощью композиций непрерывных функций одного переменного и сложения [41] хотя и не относится прямо к нейронным сетям, представляет собой математический результат из близкой области и также находит практические применения (см., например, работу [42]). Работа [41] продолжает работы [43, 44], в которых дается решение 13-й проблемы Гильберта.

3. АППРОКСИМАЦИЯ ФУНКЦИЙ НЕЧЕТКИМИ СИСТЕМАМИ

Пусть A_j^l, $l = 1, ..., d, j = 1, ..., n$ — нечеткие множества, универсальным множеством является $\mathbb{R}$ (определение нечеткого множества и исторические ссылки см., например, в работе [45]). Через $\mu_{A_j^l}$ обозначим функцию принадлежности нечеткого множества A_j^l. Соответственно, $\mu_{A_j^l}: \mathbb{R} \to [0, 1]$. Например, в качестве функций принадлежности могут использоваться кусочно-линейные функции.

Как и в предыдущих параграфах, рассматривается задача построения для некоторой функции аппроксимирующего агрегата из определенного класса. Этот агрегат должен каждому значению $x = (x_1, ..., x_d)$ из некоторого множества, принадлежащего $\mathbb{R}^d$, ставить в соответствие действительное число y. В предыдущих параграфах такими агрегатами были алгебраические многочлены и нейронные сети[2].

Нечеткая система состоит из n нечетких правил вида:

$$R_j: \text{ЕСЛИ } (x_1 = A_j^1) \text{ И ... И } (x_d = A_j^d),$$
$$\text{ТО } y = a_j^0;\ j = 1, ..., n. \quad (3)$$

Здесь $a_j^0 \in \mathbb{R}$. Если бы консеквент имел вид $y = A_j^0$, где A_j^0 — некоторое нечеткое множество, то нечеткая система (3) называлась бы нечеткой системой Мамдани. Если бы консеквент имел вид $y = a_j^0 + a_j^1 x_1 + ... + a_j^d x_d$, где $a_j^0, a_j^1, ..., a_j^d \in \mathbb{R}$, то нечеткая система (3) называлась бы нечеткой системой Такаги — Сугено. В приведенном виде нечеткая система (3) может рассматриваться и как упрощенная форма нечеткой системы Мамдани, и как упрощенная форма нечеткой системы Такаги — Сугено. Подробнее о нечетких системах Мамдани и Такаги — Сугено см., например, в книге [1].

Пусть

$$\xi_j(x) = \frac{\prod_{l=1}^{d} \mu_{A_j^l}(x_l)}{\sum_{k=1}^{n} \prod_{l=1}^{d} \mu_{A_k^l}(x_l)}, \quad j = 1, ..., n,$$

и

$$f_n(x) = \sum_{j=1}^{n} a_j^0 \xi_j(x). \quad (4)$$

Если будет доказано, что универсальным аппроксиматором (в некотором смысле) является нечеткая система (3), (4), это будет означать, что универсальными аппроксиматорами (в этом же смысле) являются и соответствующая нечеткая система Мамдани, и соответствующая нечеткая система Такаги — Сугено, поскольку они включа-

[2] Разумеется, эти агрегаты являются функциями. Но мы все же употребляем слово *агрегат*, поскольку слово *функция* неоднократно употребляется для обозначения других объектов: *функция активации*, *функция принадлежности*, *аппроксимируемая функция*.

ют в себя большее число аппроксимирующих агрегатов, чем упрощенная нечеткая система (3), (4).

В различных формах и при различных ограничениях теоремы о том, что нечеткие системы являются универсальными аппроксиматорами для непрерывных функций (т. е. имеет место теорема, аналогичная теореме 5), доказываются во многих работах. В ряде работ не формулируются явно все используемые при доказательствах условия. Первые доказательства приведены в работах[3] [7—9, 46, 47].

Приводимое ниже доказательство теоремы о том, что нечеткие системы являются универсальными аппроксиматорами для непрерывных функций, является обработкой доказательства из работы [47]. Однако содержит, на наш взгляд, определенные упрощения и дается в более строгой форме.

Пусть компактное множество $K \subseteq \mathbb{R}^d$. Без ограничения общности будем считать, что $K \subseteq [0, 1]^d$. Выберем произвольную функцию $f \in C(K)$ и произвольное $\varepsilon > 0$.

Воспользовавшись теоремой 5, найдем алгебраический многочлен P, аппроксимирующий функцию f в равномерной метрике на K с точностью $\varepsilon/2$, т. е.

$$\sup_{x \in K} |f(x) - P(x)| < \frac{\varepsilon}{2}.$$

Пусть m — натуральное число, и каждое из целых чисел $q_1, ..., q_m$ может принимать значения $0, 1, ..., m$. Через Q_m обозначим множество точек $q = (q_1, ..., q_d)$. Таким образом, множество Q_m содержит $(m + 1)^d$ точек. При $q \in Q_m$ через q/m обозначим точку $\left(\frac{q_1}{m}, ..., \frac{q_d}{m}\right)$, принадлежащую множеству $[0, 1]^d$.

При каждом $q \in Q_m$ будем считать, что $A_q^1, ..., A_q^d$ — нечеткие множества. Функцию принадлежности нечеткого множества A_q^l обозначим $\mu_{A_q^l}$. Как и раньше, универсальным множеством является $\mathbb{R}$, т. е. каждая функция $\mu_{A_q^l} : \mathbb{R} \to [0, 1]$.

Нечеткая система состоит из $n = (m + 1)^d$ нечетких правил.

$$R_q : \text{ЕСЛИ } (x_1 = A_q^1) \text{ И ... И } (x_d = A_q^d),$$
$$\text{ТО } y = P\left(\frac{q}{m}\right), \; q \in Q_m.$$

[3] В работе [47] содержится критическая оценка результата из работы [46]. С этой критикой можно согласиться.

Тогда

$$\xi_q(x) = \frac{\prod_{l=1}^{d} \mu_{A_q^l}(x_l)}{\sum_{q' \in Q_m} \prod_{l=1}^{d} \mu_{A_{q'}^l}(x_l)}, \tag{5}$$

$$f_n(x) = \sum_{q \in Q_m} P\left(\frac{q}{m}\right) \xi_q(x). \tag{6}$$

Теперь необходимо наложить условия на нечеткие множества A_q^l. Будем считать, что знаменатель дроби (5) положителен при любом $x \in [0, 1]^d$. При этом условии система функций $\xi_q(x)$ является разбиением единицы для $[0, 1]^d$. Второе условие состоит в том, что существуют последовательности неотрицательных чисел $\alpha_m \to 0$ и $\beta_m \to 0$ при $m \to \infty$ такие, что при любом $x \in [0, 1]^d$

$$\sum_{q \in M_2(x)} \xi_q(x) < \alpha_m,$$

где

$$M_1(x) = \left\{ q \in Q_m : \left\| x - \frac{q}{m} \right\| \le \beta_m \right\},$$

$$M_2(x) = \left\{ q \in Q_m : \left\| x - \frac{q}{m} \right\| > \beta_m \right\}.$$

Очевидно, что функции принадлежности $\mu_{A_q^l}$ можно построить так, чтобы эти два условия выполнялись, если в качестве класса функций принадлежности рассматривается класс всех кусочно-линейных функций[4]. Доказательство приводимой ниже теоремы основано на стандартных рассуждениях, относящихся к разбиению единицы.

Теорема 8. *Существует функция f_n вида* (6) *такая, что*

$$\sup_{x \in K} |f(x) - f_n(x)| < \varepsilon.$$

Доказательство теоремы 8. Существуют константы $C_1 > 0$, $C_2 > 0$ такие, что

$$|P(x)| < C_1, \quad \|\mathrm{grad}\, P(x)\| < C_2$$

при любом $x \in [0, 1]^d$.

[4] Более того, последовательность $\{\alpha_m\}$ можно было бы взять нулевой. Но тогда были бы исключены из рассмотрения некоторые другие классы функций принадлежности, используемые при решении практических задач. Отметим также, что выбор системы нечетких множеств не связан с функцией f.

Зафиксировав натуральное число m и $x \in [0, 1]^d$, положим

$$\gamma(x) = \sum_{q \in M_2(x)} \xi_q(x).$$

Тогда на основании выражения (6)

$$|P(x) - f_n(x)| = \left|P(x) - \sum_{q \in Q_m} P\left(\frac{q}{m}\right)\xi_q(x)\right| \le$$

$$\le \left|P(x) - \sum_{q \in M_1(x)} P\left(\frac{q}{m}\right)\xi_q(x)\right| + \left|\sum_{q \in M_2(x)} P\left(\frac{q}{m}\right)\xi_q(x)\right|.$$

Второе слагаемое в правой части неравенства не превосходит $C_1\alpha_m$. Поэтому

$$|P(x) - f_n(x)| \le$$

$$\le \left|\gamma(x)P(x) + \sum_{q \in M_1(x)}\left(P(x) - P\left(\frac{q}{m}\right)\right)\xi_q(x)\right| + C_1\alpha_m \le$$

$$\le \sum_{q \in M_1(x)}\left|P(x) - P\left(\frac{q}{m}\right)\right|\xi_q(x) + 2C_1\alpha_m.$$

Воспользовавшись тем, что

$$\left|P(x) - P\left(\frac{q}{m}\right)\right| \le \left(\sup_{y \in [0, 1]^d}\|\mathrm{grad}P(y)\|\right)\left\|x - \frac{q}{m}\right\|,$$

получаем

$$|P(x) - f_n(x)| \le C_2\beta_m + 2C_1\alpha_m.$$

Откуда

$$\|P - f_n\|_\infty \le C_2\beta_m + 2C_1\alpha_m.$$

При достаточно большом m

$$C_2\beta_m + 2C_1\alpha_m < \frac{\varepsilon}{2}.$$

Далее,

$$\sup_{x \in K}|f(x) - f_n(x)| \le \sup_{x \in K}|f(x) - P(x)| +$$
$$+ \sup_{x \in K}|P(x) - f_n(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Теорема 8 доказана. ♦

В доказательстве теоремы 8 при больших d число нечетких правил может быть очень большим. Представляет интерес вопрос об уменьшении числа нечетких правил при сохранении точности аппроксимации (можно соотнести данный вопрос с теоремами 2 и 3). Такие задачи рассматриваются во многих работах (см., например, [48, 49].) Число нечетких правил может быть уменьшено благодаря применению многосеточных методов или кластеризации [50].

В ряде работ рассматривается аппроксимация нечеткими системами в пространствах Соболева, т. е. совместная аппроксимация функции и ее частных производных, см., например, работы [51, 52].

Доказательство теоремы 8 дословно повторяется, если взять в качестве t-нормы не произведение, а минимум. (Достаточно воспользоваться не совсем привычным обозначением $\prod_{j=1}^{n} a_j = \min(a_1, ..., a_n)$.) Существуют результаты об аппроксимации функций нечеткими системами, где используются другие t-нормы (см., например, работу [53]) или вместо t-норм используются нормы другого вида [54, 55]. Могут рассматриваться и системы нечеткого вывода отличные от систем Мамдани и систем Такаги — Сугено (см. работу [56]).

В работе [57] доказывается теорема об универсальной аппроксимации для иерархических нечетких систем.

В работе [58] рассматривается задача аппроксимации нечеткими системами, которые являются монотонными функциями по каждой переменной при фиксированных остальных переменных.

Аппроксимация системами, в которых используются нечеткие множества типа 2, изучается в работе [59].

В работе [60] рассматривается нечеткая система с последующим сглаживанием.

Идентификации нечетких систем, т. е. определению параметров антецедентов и консеквентов всех нечетких правил, уделяется большое внимание (см., например, книги [61, 62]). В ряде работ идентификация нечетких систем рассматривается и в связи с задачей аппроксимации функций (см., например, работу [49]).

ЗАКЛЮЧЕНИЕ

Теория приближения функций — это большое и важное научное направление, появление которого связано с именами К. Вейерштрасса и П.Л. Чебышева. Результаты и методы данной теории находят применения в различных прикладных исследованиях.

Сначала в теории приближения функций изучались задачи, где в качестве аппроксимирующих агрегатов рассматривались либо алгебраические многочлены (для непериодических функций), либо тригонометрические полиномы (для периодических функций). Однако потребности практики привели к значительному расширению данного научного направления, поскольку оказалось, что во многих приложениях вместо алгебраических многочленов или тригонометрических полиномов лучше использовать другие аппроксиматоры. К таким новым аппроксиматорам, применение и изучение которых активно начались в последние десятилетия, относятся нейронные сети и нечеткие системы.

Влияние классических разделов теории приближения функций на развитие этих новых разделов оказалось очень сильным. Доказываются те же теоремы, что и в классических разделах, об увеличении скорости сходимости при повышении гладкости функции, о совместном приближении функции и ее производных, о формосохраняющих приближениях, о приближении в интегральных метриках. Однако результаты об аппроксимации функций с помощью нейронных сетей или с помощью нечетких систем значительно теснее связаны с практикой, чем результаты об аппроксимации функций алгебраическими многочленами или тригонометрическими полиномами.

Рассматриваемые теоремы об аппроксимации важны, например, при изучении временных рядов. Зависимость будущих значений от прошлых может выражаться (или достаточно хорошо приближенно выражаться) сложной нелинейной функцией. Сама эта функция неизвестна. Однако знание того, что ее можно аппроксимировать, например, с помощью нечеткой системы, позволяет уйти от вопроса, какого вида нелинейные функции лучше включать в математическую модель для временного ряда, и в то же время не ограничиваться линейными моделями. Широкое применение нейронных сетей при прогнозировании на финансовых рынках также во многом объясняется известными теоремами об аппроксимационных свойствах этих сетей.

ЛИТЕРАТУРА

1. *Пегат А.* Нечеткое моделирование и управление. 2-е изд. — М.: БИНОМ. Лаборатория знаний, 2013. — 798 с.
2. *Хайкин С.* Нейронные сети. Полный курс: 2-е изд. — М.: Изд. дом «Вильямс», 2017. — 1104 с.
3. *Рутковская Д., Пилиньский М., Рутковский Л.* Нейронные сети, генетические алгоритмы и нечеткие системы: 2-е изд. — М.: Горячая линия — Телеком, 2013. — 384 с.
4. *Cybenko G.* Approximation by superpositions of sigmoidal function // Mathematics of Control, Signals, and Systems. — 1989. — Vol. 2. — P. 303—314.
5. *Funahashi K.I.* On the approximate realization of continuous mappings by neural networks // Neural Networks. — 1989. — Vol. 2. — P. 183—192.
6. *Hornik K., Stinchcombe M., White H.* Multilayer feedforward networks are universal approximators // Neural Networks. — 1989. — Vol. 2. — P. 359—366.
7. *Kosko B.* Fuzzy systems as universal approximators // Proc. of IEEE 1992 Intern. Conf. Fuzzy Systems (San Diego, CA) Mar. 1992. — P. 1153—1162.
8. *Wang L.-X.* Fuzzy systems are universal approximators // in: Proc. IEEE 1992 Intern. Conf. Fuzzy Systems (San Diego, CA) Mar. 1992. — P. 1163—1170.
9. *Wang L.-X., Mendel J.M.* Fuzzy basis functions, universal approximation, and orthogonal least-squares learning // IEEE Trans. on Neural Networks. — 1992. — Vol. 3. — P. 807—814.
10. *Натансон И.П.* Конструктивная теория функций. — М.: Государственное технико-теоретическое изд-во, 1949. — 688 с.
11. *Дзядык В.К.* Введение в теорию равномерного приближения функций полиномами. — М.: Наука, 1977. — 512 с.
12. *Гончаров В.Л.* Теория интерполирования и приближения функций. — М.: Гос. техн.-теорет. изд-во, 1934. — 316 с.
13. *Lorentz G.G.* Bernstein polynomials. — Toronto: University of Toronto Press, 1953. — 130 p.
14. *Коровкин П.П.* Линейные операторы и теория приближений. — М.: Гос. изд-во физ.-мат. лит., 1959. — 211 с.
15. *Altomare F.* Korovkin-type theorems and approximation by positive linear operators // Surveys in Approximation Theory. — 2010. — Vol. 5. — P. 92—164.
16. *Anastassiou G.A.* Fuzzy mathematics: Approximation theory. — Berlin: Springer, 2010. — 444 p.
17. *Потапов М.К.* О приближении алгебраическими многочленами в интегральной метрике с весом Якоби // Вестник Московского университета. Сер. Математика и механика. — 1983. — Вып. 4. — С. 43—52.
18. *Никольский С.М.* Приближение функций многих переменных и теоремы вложения. — М.: Наука, 1977. — 456 с.
19. *Шведов А.С.* Приближение функций многих переменных многочленами и поперечники некоторых функциональных классов // Analysis Mathematica. — 1982. — Vol. 8. — P. 135—150.
20. *Veretennikov A.Yu., Veretennikova E.V.* On partial derivatives of multivariate Bernstein polynomials // Siberian Advances in Mathematics. — 2016. — Vol. 26. — P. 294—305.
21. *Лоран П.-Ж.* Аппроксимация и оптимизация. — М.: Мир, 1975. — 496 с.
22. *Mhaskar H.N., Micchelli C.A.* Approximation by superposition of sigmoidal and radial basis functions // Advances in Applied Mathematics. — 1992. — Vol. 13. — P. 350—373.
23. *Park J., Sandberg I.W.* Universal approximation using radial-basis-function networks // Neural Computation. — 1991. — Vol. 3. — P. 246—257.
24. *Chen T., Chen H.* Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks // IEEE Trans. on Neural Networks. — 1995. — Vol. 6. — P. 904—910.
25. *Chen T., Chen H.* Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems // Ibid. — 1995. — Vol. 6. — P. 911—917.
26. *Barron A.R.* Universal approximation bounds for superpositions of a sigmoidal function // IEEE Trans. on Information Theory. — 1993. — Vol. 39. — P. 930—945.
27. *Chui C.K., Li X., Mhaskar H.N.* Neural networks for localized approximation // Mathematics of Computation. — 1994. — Vol. 63. — P. 607—623.
28. *Llanas B., Lantaron S., Sainz F.J.* Constructive approximation of discontinuous functions by neural networks // Neural Process Lett. — 2008. — Vol. 27. — P. 209—226.
29. *Hou M., Han X.* Constructive approximation to multivariate function by decay RBF neural network // IEEE Trans. on Neural Networks. — 2010. — Vol. 21. — P. 1517—1523.
30. *Wang J.-J., Chen B.-L., Yang C.-Y.* Approximation of algebraic and trigonometric polynomials by feedforward neural networks // Neural Comput & Applic. — 2012. — Vol. 21. — P. 73—80.
31. *Fang Y., Chow T.W.S.* Wavelets based neural network for function approximation // in: J. Wang, et al. (eds.): ISNN 2006, LNCS 3971. — Berlin: Springer, 2006 — P. 80—85.
32. *Xie T.-F., Zhou X.-L.* Neural networks for optimal approximation of continuous functions in $\mathbb{R}^d$ // Appl. Math. Journal Chinese Univ. — 2012. — Vol. 27. — P. 335—344.
33. *Ito Y.* Simultaneous approximations of polynomials and derivatives and their applications to neural networks // Neural Computation. — 2008. — Vol. 20. — P. 2757—2791.
34. *Anastassiou G.A.* Intelligent systems: Approximation by artificial neural networks. — Berlin: Springer, 2011. — 107 p.
35. *Iatan I.F.* Modern neural methods for function approximation // in: Issues in the Use of Neural Networks in Information Retrieval, Studies in Computational Intelligence, vol. 661, Cham, Springer, 2017. — P. 107—121.

36. *Cherkassky V., Gehring D., Mulier F.* Comparison of adaptive methods for function estimation from samples // IEEE Trans. on Neural Networks. — 1996. — Vol. 7. — P. 969—984.
37. *Gomes G.S. da S., Ludermir T.B., Almeida L.M.* Neural networks with asymmetric activation function for function approximation // Intern. Joint Conf. on Neural Networks (Atlanta, GA) June 2009.
38. *Selmic R.R., Lewis F.L.* Neural-network approximation of piecewise continuous functions: Application to friction compensation // IEEE Trans. on Neural Networks. — 2002. — Vol. 13. — P. 745—751.
39. *Rudenko O., Bezsonov O.* Function approximation using robust radial basis function networks // Journal of Intelligent Learning Systems and Applications. — 2011. — Vol. 3. — P. 17—25.
40. *Ferrari S., Stengel R.F.* Smooth function approximation using neural networks // IEEE Trans. on Neural Networks. — 2005. — Vol. 16. — P. 24—38.
41. *Колмогоров А.Н.* О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // Доклады АН СССР. — 1957. — Т. 114. — С. 953—956.
42. *de Figueiredo R.J.P.* Implications and applications of Kolmogorov's superposition theorem // IEEE Trans. on Automatic Control. — 1980. — Vol. 25. — P. 1227—1231.
43. *Колмогоров А.Н.* О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных // Доклады АН СССР. — 1956. — Т. 108. — С. 179—182.
44. *Арнольд В.И.* О функциях трех переменных // Доклады АН СССР. — 1957. — Т. 114. — С. 679—681.
45. *Шведов А.С.* Нечеткое математическое программирование: краткий обзор // Проблемы управления. — 2017. — № 3. — С. 2—10.
46. *Buckley J.* Sugeno type controllers are universal controllers // Fuzzy Sets and Systems. — 1993. — Vol. 53. — P. 299—303.
47. *Ying H.* General Takagi-Sugeno fuzzy systems are universal approximators // in: Proc. IEEE Intern. Conf. on Fuzzy Systems (Anchorage, AK) May 1998. — Vol. 1. — P. 1819—823.
48. *Yan S.Y., Sun Z.Q.* Universal approximation for Takagi-Sugeno fuzzy systems combining statically and dynamically constructive methods — MISO cases // Intern. Conf. on Information, Management and Engineering (Kuala Lumpur, Malaysia) April 2009.
49. *Sonbol A.H., Fadali M.S., Jafarzadeh S.* TSK fuzzy function approximators: Design and accuracy analysis // IEEE Trans. on Systems, Man and Cybernetics — Part B: Cybernetics. — 2012. — Vol. 42. — P. 702—712.
50. *Herrera L.J., Pomares H., Rojas I.,* et al. Multigrid-based fuzzy systems for function approximation // in: R. Monroy, et al. (eds.): MICAI 2004, LNAI 2972. — Berlin: Springer, 2004. — P. 252—261.
51. *Landajo M., Rio M.J., Perez R.* A note on smooth approximation capabilities of fuzzy systems // IEEE Trans. on Fuzzy Systems. — 2001. — Vol. 9. — P. 229—237.
52. *Salgado P.* Approximation of functions and its derivatives by fuzzy systems // Intern. Conf. on Computational Intelligence for Modelling, Control and Automation (Vienna, Austria) Nov. 2005.
53. *Bede B., Rudas I.J., Schwab E.D., Schwab G.* Approximation properties of Łukasiewicz fuzzy systems // Annual Conf. of the North American Fuzzy Information Processing Society (Redmond, WA) Aug. 2015.
54. *Yager R.R., Kreinovich V.* Universal approximation theorem for uninorm-based fuzzy systems modeling // Fuzzy Sets and Systems. — 2003. — Vol. 140. — P. 331—339.
55. *Rickard J.T., Aisbett J., Mendel J.M.* Rule-based fuzzy systems with weighted power mean firing operator as universal approximators // IEEE Intern. Conf. on Fuzzy Systems (Brisbane, Australia) June 2012.
56. *Mandal S., Jayaram B.* Similarity-based reasoning fuzzy systems and universal approximation // in: Mathematics and Computing 2013 R.N. Mohapatra, et al. (eds.), Springer Proceedings in Mathematics and Statistics 91, Springer India, 2014. — P. 215—230.
57. *Joo M.G., Lee J.S.* Universal approximation by hierarchical fuzzy system with constraints on the fuzzy rule // Fuzzy Sets and Systems. — 2002. — Vol. 130. — P. 175—188.
58. *Hušek P.* Recursive least square approximation using monotonic fuzzy systems // IEEE Intern. Symposium on Intelligent Control (Buenos Aires, Argentina) September, 2016.
59. *Ying H.* A sufficient condition on a general class of interval type-2 Takagi-Sugeno fuzzy systems with linear rule consequent as universal approximators // Journal of Intelligent and Fuzzy Systems. — 2015. — Vol. 29. — P. 1219—1227.
60. *Richardson J., Korniak J., Reiner P.D., Wiliamowski B.M.* Nearest-neighbor spline approximation (NNSA) improvement to TSK fuzzy systems // IEEE Trans. on Industrial Informatics. — 2016. — Vol. 12. — P. 169—178.
61. *Lilly J.H.* Fuzzy control and identification. — Hoboken (NJ): Wiley, 2010. — 231 p.
62. *Zhang H., Liu D.* Fuzzy Modeling and Fuzzy Control. — Boston: Birkhauser, 2006. — 416 p.

Статья представлена к публикации членом редколлегии А.С. Манделем.

Шведов Алексей Сергеевич — д-р физ.-мат. наук, профессор, Национальный исследовательский университет «Высшая школа экономики», ✉ ashvedov@hse.ru.