

DOI: 10.51790/2712-9942-2021-2-4-1

МАТЕМАТИЧЕСКИЕ ЗАДАЧИ, СВЯЗАННЫЕ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ И ИСКУССТВЕННЫМИ НЕЙРОННЫМИ СЕТЯМИ**В. Б. Бетелин¹, В. А. Галкин²**¹ *Федеральное государственное учреждение «Федеральный научный центр**Научно-исследовательский институт системных исследований Российской академии наук», г. Москва, Российская Федерация, ORCID: <http://orcid.org/0000-0001-6646-2660>, betelin@inbox.ru*² *Сургутский филиал Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», г. Сургут, Российская Федерация, ORCID: <http://orcid.org/0000-0002-9721-4026>, val-gal@yandex.ru*

Аннотация: предложен общий топологический подход для анализа искусственных нейронных сетей на основе симплициальных комплексов и свойств аппроксимации непрерывных отображений их симплициальными приближениями. Выявлены существенные для этого класса задач явления вычислительной неустойчивости, связанной с общими проблемами некорректных задач в гильбертовом пространстве и методами их регуляризации, типичными для обработки Big Data. Сформулированы критерии точности и применимости моделей искусственных нейронных сетей, рассмотрены примеры их реализации на основе теории интерполяции функций. Развитие идей П.Л.Чебышёва о наилучшем приближении служит отправной точкой для широкого класса математических исследований по оптимизации обучающих наборов для построения ИНС.

Ключевые слова: искусственные нейронные сети, методы оптимизации, вычислительная неустойчивость, методы регуляризации.

Благодарности: работа выполнена в рамках государственного задания ФГУ ФНЦ НИИСИ РАН (проведение фундаментальных научных исследований (47 ГП) по теме 0580–2021–0019, FNEF–2021–0019, рег. № 121031300099–5).

Для цитирования: Бетелин В. Б., Галкин В. А. Математические задачи, связанные с искусственным интеллектом и искусственными нейронными сетями. *Успехи кибернетики*. 2021;2(4):6–14. DOI: 10.51790/2712-9942-2021-2-4-1.

MATHEMATICAL PROBLEMS OF ARTIFICIAL INTELLIGENCE AND ARTIFICIAL NEURAL NETWORKS**V. B. Betelin¹, V. A. Galkin²**¹ *Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, Moscow, Russian Federation, ORCID: <http://orcid.org/0000-0001-6646-2660>, betelin@inbox.ru*² *Surgut Branch of Federal State Institute “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, Surgut, Russian Federation, ORCID: <http://orcid.org/0000-0002-9721-4026>, val-gal@yandex.ru*

Abstract: we propose a general topological approach to the analysis of artificial neural networks using simplicial complexes and the approximation of continuous mappings with simplicial ones. The essential properties of numerical instability in such problems were identified. It is associated with ill-posed problems in Hilbert space and regularization methods typically applied to Big Data processing. We formulated the criteria of artificial neural network accuracy and applicability and included some implementation examples based on the interpolation theory. Advancing P.L. Chebyshev’s ideas about the best approximation may be an entry point to various mathematical research on artificial neural network training dataset optimization.

Keywords: artificial neural networks, optimization methods, numerical instability, regularization methods.

Acknowledgements: this study is a part of the Scientific Research Institute for System Analysis (Russian Academy of Sciences) government contract for basic research (47 GP), research area 0580–2021–0019, FNEF–2021–0019, Reg. No. 121031300099–5.

Cite this article: Betelin V. B., Galkin V. A. Mathematical Problems of Artificial Intelligence and Artificial Neural Networks. *Russian Journal of Cybernetics*. 2021;2(4):6–14. DOI: 10.51790/2712-9942-2021-2-4-1.

Во многочисленных вычислительных экспериментах по распознаванию образов, обработке акустической, видео и текстовой информации были найдены подходы к созданию программного обеспечения для практического решения ряда трудно формализуемых задач. Эти успехи породили огромный поток работ и интерес к созданию полуэмпирических методов, носящих название искусственные нейронные сети (ИНС).

Общий подход, лежащий в основе построения ИНС, состоит в принятии гипотезы возможности создания устройства, которое можно обучить на серии примеров принятию решений. Эта размытая формулировка предполагает наличие некоторой связи между выбранными парами объектов $(x, y) \in X \times Y$, определяющей некоторое отношение $R \subset X \times Y$ (гипотетический закон) [1]. Обычно предполагается, что R является функцией (т.е. каждому значению $x \in X$ соответствует ровно одно значение $y \in Y$), при этом аргументы X называются причинами, а множество значений Y — следствиями. Схематически ИНС является параллельной системой распределенной обработки информации в виде ориентированного графа, где вершины графа называются обрабатывающими элементами (искусственные нейроны), а связи между ними называются соединениями. Таким образом, ИНС представляет собой граф — систему узлов, соединенных связями и обменивающихся между собой входными-выходными данными (см., например, рисунок).

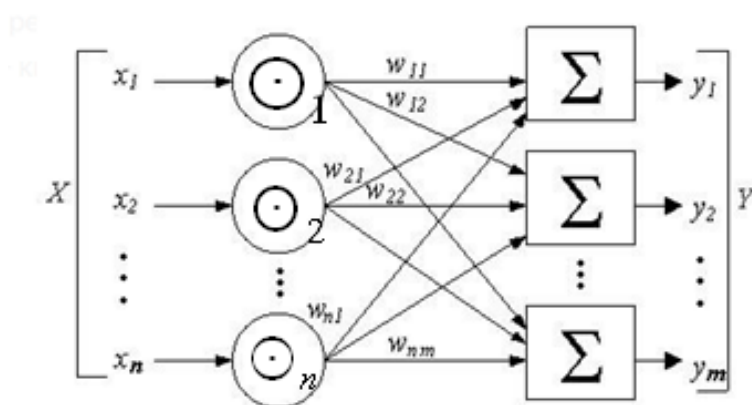


Рис. Пример блок-схемы структуры элементарной ячейки одного слоя ИНС

Каждый узел в сети получает и передает сигналы, которые он преобразовывает через некоторую систему заданных отображений \odot (т.е. искусственные нейроны, вообще говоря, различные для разных узлов). Далее преобразованные сигналы умножаются на весовые коэффициенты $w_{i,j}$ в процессе передачи данных по линиям соединения между вершинами i и j графа, а затем полученные значения суммируются при поступлении на сумматоры Σ , формируя выходную информацию описанного слоя нейронной сети. Полученные данные могут быть повторно подвергнуты аналогичной обработке и т.д. до формирования конечного выходного сигнала $y = (y_1, y_2, \dots, y_m) \in \mathbb{R}_m$.

Такие графы с заданными функциями преобразования сигналов при соответствующей настройке способны реализовывать различные классы математических алгоритмов.

Базовая элементарная ячейка графа ИНС (так называемый один слой ИНС) приведена на рис., где входные данные $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}_n$ некоторым образом преобразуются в выходные данные $y = (y_1, y_2, \dots, y_m) \in \mathbb{R}_m$. Такие ячейки могут комбинироваться, вообще говоря, произвольным образом, когда выходные данные одного слоя передаются соединениями графа на другой слой (включая организацию обратных связей). Граф, составленный из таких ячеек, с дополнительной надстройкой в виде алгоритма преобразования данных в вершинах и на их связях назовем реализацией ИНС.

Настройка ИНС (аналог процедуры автоматизированного программирования структуры ИНС) состоит в подборе семейства искусственных нейронов (преобразований, обозначаемых на рис. значками $\odot_1, \odot_2, \dots, \odot_n$ в вершинах графа) и коэффициентов $w_{i,j}$ умножения сигналов вдоль линий соеди-

нения вершин на основе дополнительно декларируемых принципов (например, оптимизации). Набор обучающих примеров составляет заданное подмножество $\bar{R} \subset R$, где отношение R априори неизвестно. Целью создания ИНС является в некотором смысле «оптимальная реконструкция» неизвестного отношения R на основе заданного «обучающего» набора \bar{R} . По своей природе такая постановка задачи является некорректной ввиду существенной множественности ее решений («школьная» задача об отыскании кривой, соединяющей заданный набор точек). Таким образом, необходимым элементом построения ИНС служит введение целевого правила (критерия, функционала и т.п.) F , на основе применения которого выполняется построение «наилучшего» продолжения заданного отношения \bar{R} до отношения R_F , которое аппроксимирует R (например, посредством минимизации функционала F на некотором множестве параметров). Назовем R_F математической моделью реализации ИНС. Описанная схема обычно применяется для построения продолжения функций, в частности, в задачах интерполяции.

Типичная постановка теории приближенного вычисления значений функции R на множестве аргументов $x = (x_1, x_2, \dots, x_n) \in D \subset \mathbb{R}_n$ заключается в выборе оптимального многопараметрического приближения из заданного множества функций $\bar{R}(x, w_1, w_2, \dots, w_k) \approx R(x)$, где (w_1, w_2, \dots, w_k) — множество параметров, по которым проводится вышеупомянутая оптимизация. Например, для заданных значений аргументов $\bar{x}^{(l)} \in D$, $l = 1, 2, \dots, N$, требуется выполнение минимальности уклонения нормы $F(w_1, w_2, \dots, w_k) = \max_{1 \leq l \leq N} \|\bar{R}(\bar{x}^{(l)}, w_1, w_2, \dots, w_k) - \bar{y}^{(l)}\|$ в предположении, что пары $(\bar{x}^{(l)}, \bar{y}^{(l)})$, $1 \leq l \leq N$, находятся в заданном отношении \bar{R} . Такой подход типичен для задач прогноза, регрессии, построения интерполяционных многочленов [2]. Если условие минимума уклонения означает совпадение значений $\bar{y}^{(l)} = R(\bar{x}^{(l)}) = \bar{R}(\bar{x}^{(l)}, w_1, w_2, \dots, w_k)$, $1 \leq l \leq N$, то такой способ построения приближения называется интерполяцией (в этом случае целевой функционал $F(w_1, w_2, \dots, w_k)$ в точке оптимума принимает наименьшее значение, равное нулю). Отметим, что определение функционала F может быть осуществлено разными способами на основе экспертных оценок выбора метрик (расстояний) на множестве значений Y , определяющих цели решаемых задач. Такого сорта проблемы возникают при планировании экспериментов в естественных науках, в экономике, при построении алгоритмов распознавания образов, при решении обратных задач классификации объектов, при проектировании систем управления, наведения, прицеливания, навигации и т. п. [2].

Остановимся на наиболее изученном примере построения интерполяционных многочленов как прообразе математических проблем, возникающих при создании ИНС.

Пусть на отрезке числовой прямой $x \in X = [a, b]$ заданы точки $a \leq \bar{x}^{(1)} < \bar{x}^{(2)} < \dots < \bar{x}^{(N)} \leq b$ (узлы интерполяции), которым соответствуют значения $\bar{y}^{(l)}$, $l = 1, 2, \dots, N$, так что $\bar{R} = \{(\bar{x}^{(1)}, \bar{y}^{(1)}), (\bar{x}^{(2)}, \bar{y}^{(2)}), \dots, (\bar{x}^{(N)}, \bar{y}^{(N)})\}$. Положим, что аппроксимирующее семейство задается многочленами:

$$\bar{R}(x, w_1, w_2, \dots, w_N) = \sum_{i=1}^N w_i x^{i-1} \quad (1)$$

с вещественными коэффициентами w_i . Оптимизационная задача об отыскании неопределенных коэффициентов w_i на основе обращения в нуль значений функционала:

$$F(w_1, w_2, \dots, w_k) = \max_{1 \leq l \leq N} |\bar{R}(\bar{x}^{(l)}, w_1, w_2, \dots, w_k) - \bar{y}^{(l)}|$$

в этом случае приводит к известной системе линейных алгебраических уравнений с определителем Вандермонда, а соответствующий интерполяционный многочлен (1) в форме Лагранжа имеет следующий стандартный вид [2]:

$$R_F(x) = \sum_{i=1}^N \bar{y}^{(i)} \frac{\omega_N(x)}{\omega'_N(\bar{x}^{(i)})(x - \bar{x}^{(i)})}, \quad \omega_N(x) = \prod_{i=1}^N (x - \bar{x}^{(i)}). \quad (2)$$

Таким образом, в соответствии с блок-схемой на рис. интерполяционный многочлен (2) можно считать реализацией ИНС с одним входным скалярным входом $x \in \mathbb{R}$ и скалярным выходом $y = R_F(x) \in \mathbb{R}$. При этом искусственные нейроны в N вершинах однослойной ИНС получают входной сигнал x , преобразовывают его по формулам, которые назовем нейронами Лагранжа:

$$\odot_i(x) = \frac{\omega_N(x)}{\omega'_N(\bar{x}^{(i)})(x - \bar{x}^{(i)})}, \quad 1 \leq i \leq N.$$

Выходные сигналы с каждого искусственного нейрона Лагранжа $\odot_i(x)$ умножаются на коэффициенты $w_i = \tilde{y}^{(i)}$ и поступают на сумматор \sum , формируя по формуле (2) выходной сигнал y . (Альтернативной реализацией такой ИНС может считаться искусственный нейрон, напрямую заданный формулой (2).)

Отметим важные понятия для более общих ИНС, которые ассоциируются с рассмотренным примером:

1. Емкость памяти ИНС, определяемая объемом обучающего набора:

$$\tilde{R}_N = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}.$$

2. Точность ИНС по приближению заданного класса аппроксимируемых отношений R . Достижимость поставленных целей.

3. Скорость обучения ИНС на основе заданного алгоритма F .

4. «Гибкость» ИНС по отношению к перенастройке ее структуры (а именно, к динамическому изменению набора функций искусственных нейронов \odot_i , правил выбора весовых коэффициентов w_i , обучающего набора \tilde{R}_N и т.п.).

5. Быстродействие ИНС.

6. Устойчивость получаемых выходных данных при изменениях входных данных.

7. Реализуемость вычислительных процедур на искусственных нейронах, их надежность.

8. Зависимость точности и быстродействия ИНС от ее размерности.

9. Возможность автоматического понижения размерности ИНС в процессе эксплуатации с целью оптимизации ее характеристик.

10. Отказоустойчивость, резервирование функций и помехозащищенность ИНС.

11. Совместимость с другими ИНС за счет расширения структуры на основе сращивания их графов и взаимного обмена данными.

12. Топологические инварианты структуры ИНС и гомотопическая эквивалентность различных реализаций ИНС.

Отметим, что содержание п. 11 относится к комплексу, составленному из ИНС различной природы, который можно охарактеризовать как приближение систем искусственного интеллекта (ИИ).

Что касается возникающих проблем с точностью ИНС при неограниченном наращивании объема обучающей совокупности \tilde{R} , то здесь показательными являются примеры К.Д.Т. Рунге [3] и С.Н. Бернштейна [4] для интерполяционных многочленов с неограниченным увеличением количества узлов интерполяции.

Пример 1. (К.Д.Т. Рунге, 1901 г.) Рассмотрим на отрезке $X = [-1, 1]$ интерполяцию функции (построение ИНС):

$$R(x) = \frac{1}{1 + 25x^2}$$

по обучающему набору $\tilde{R} = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}$, где $\tilde{x}^{(i)}$, $1 \leq i \leq N$ образуют равномерную сетку на X , а $\tilde{y}^{(i)} = R(\tilde{x}^{(i)})$. Тогда при неограниченном увеличении числа узлов интерполяции (число искусственных нейронов в ИНС $N \rightarrow \infty$) последовательность аппроксимаций (2) расходится на интервале $0.726 \dots \leq |x| \leq 1$. Таким образом, имеем:

$$\lim_{N \rightarrow \infty} \max_{-1 \leq x \leq 1} |R(x) - R_F(x)| \rightarrow \infty.$$

Следует подчеркнуть, что аппроксимируемая функция R является аналитической в окрестности действительной оси. Особенности типа полюса первого порядка у аналитического продолжения функции R на комплексную плоскость находятся в точках $\pm \frac{1}{5}i$.

Пример 2. (С.Н. Бернштейн, 1912 г.) На отрезке $X = [-1, 1]$ рассмотрим интерполяцию функции (построение ИНС):

$$R(x) = |x|.$$

Тогда последовательность интерполяционных многочленов Лагранжа R_F (2), построенных на равномерных сетках при условии, что $\tilde{x}^{(1)} = -1$, $\tilde{x}^{(N)} = 1$, не стремится при неограниченном возрастании объема $N \rightarrow \infty$ обучающего набора $\tilde{R}_N = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}$ к значениям функции $R(x) = |x|$ ни в одной точке отрезка $X = [-1, 1]$, отличной от точек x , равных трем значениям $-1; 0; 1$.

Более того, справедлива следующая теорема.

Теорема 1. (Г. Фабер [5].) Для любой интерполяционной таблицы размерности N найдутся непрерывная функция и точка $x \in X = [a, b]$ такие, что интерполяционные многочлены (2) не сходятся к указанной функции в точке x при $N \rightarrow \infty$.

Примеры Рунге, Бернштейна и общий результат Фабера указывают на возможность глубокого заблуждения, что неограниченное наращивание нейронов в ИНС и большие объемы данных $N \gg 1$ (Big Data), положенные в основу обучающих наборов \tilde{R}_N , могут обеспечить ее эффективность в рамках ИИ. То же относится ко многим утверждениям о построении так называемых «цифровых двойников» за счет безграничного наращивания обучающих наборов данных $\tilde{R}_N = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}$.

Указанные замечания относятся к общей фундаментальной проблеме анализа вычислительной устойчивости ИНС при наращивании объемов обучающих наборов и как следствие — границ реализуемости и безопасной применимости систем ИИ, построенных на комплексах взаимодействующих ИНС.

Подчеркнем, что классическая теорема Вейерштрасса [5] о плотности множества многочленов в пространстве непрерывных функций на компакте позволяет утверждать существование сходящихся ИНС, построенных на базе многочленов (1), но теорема Фабера, а также примеры К.Д.Т. Рунге и С.Н. Бернштейна указывают на риски возникновения значительных ошибок при применении ИНС, построенных на базе обучающих множеств \tilde{R}_N при $N \gg 1$.

Отметим, что на практике при построении ИНС не используются нейроны Лагранжа. В рамках вычислительных экспериментов подобран ряд искусственных нейронов, однако обоснованность их применения требует будущего детального исследования и ответов на вопросы относительно свойств ИНС, указанные в пп. 1–12. Примером популярных функций инициализации искусственного нейрона в вычислительных экспериментах является $\odot_a(x) = x\Theta(x - a)$, $a \in \mathbb{R}$, где $\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0, \end{cases}$ — функция Хевисайда (индикатор-функция множества неотрицательных чисел). По сути, это напрямую связано с реализацией простейшего интерполяционного кусочно-линейного сплайна R_F [6] на заданных значениях $\tilde{R}_N = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}$ функции $R(\tilde{x}^{(i)}) = \tilde{y}^{(i)}$, где на каждом участке линейности R_F справедливо соотношение:

$$R_F(x) = \tilde{y}^{(i)} + \frac{\tilde{y}^{(i+1)} - \tilde{y}^{(i)}}{\tilde{x}^{(i+1)} - \tilde{x}^{(i)}}(x - \tilde{x}^{(i)}), \quad \tilde{x}^{(i)} \leq x \leq \tilde{x}^{(i+1)}.$$

Очевидно, что имеет место формула:

$$R_F(x) = \sum_{i=1}^{N-1} \left(\tilde{y}^{(i)} + \frac{\tilde{y}^{(i+1)} - \tilde{y}^{(i)}}{\tilde{x}^{(i+1)} - \tilde{x}^{(i)}} \odot_{\tilde{x}^{(i)}}(x) \right) \left(1 - \Theta(x - \tilde{x}^{(i+1)}) \right) \Theta(x - \tilde{x}^{(i)}), \quad (3)$$

определяющая симплициальную (непрерывную кусочно-линейную) аппроксимацию $R_F(x)$ отображения R на симплициальном комплексе $(\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(N)})$. Формулы (3) имеют естественное обобщение на многомерные симплициальные комплексы [7].

Важно отметить, что симплициальные аппроксимации многомерных непрерывных отображений на симплициальных компактных комплексах имеют четко выраженную структуру ИНС (которую назовем симплициальной ИНС), представленную на рис. Процедура барицентрического подразделения симплициальных комплексов [7] на компактах в сочетании с равномерной непрерывностью отображений на них позволяет утверждать, что таким образом построенные симплициальные ИНС R_F равномерно сходятся к R на компактах при диаметре $\delta_N > 0$ барицентрических подразделений комплекса, стремящемся к нулю:

$$\|R_F - R\|_C \rightarrow 0, \quad N \rightarrow \infty, \quad \delta_N \rightarrow 0.$$

Подчеркнем, что скорость сходимости симплициальной ИНС на комплексе K для произвольных непрерывных отображений R определяется модулем непрерывности $\omega_R(\delta) = \sup_{\|u-v\| \leq \delta} \|R(u) - R(v)\|_{C(K)}$,

$\delta > 0$. Нетрудно убедиться, что погрешность приближения $\|R_F - R\|_C$ на симплициальном комплексе K , являющемся компактом в \mathbb{R}_n , оценивается сверху величиной $\omega_R(\delta)$, где $\delta > 0$ — точная верхняя

грань диаметра симплексов в комплексе $K = D$, на котором определено отображение R со значениями в \mathbb{R}_m . Действительно, на каждом симплексе T , входящем в симплициальный комплекс K , справедлива формула:

$$R_F(x) = \sum_{i=0}^n \mu_i R(a_i)x = \sum_{i=0}^n \mu_i a_i \in T, \mu_i \geq 0, \quad \sum_{i=0}^n \mu_i = 1,$$

где μ_i — барицентрические координаты точки x в симплексе T , натянутом на вершины a_i , $0 \leq i \leq n$, n — размерность симплекса T . Таким образом, поскольку диаметр симплекса T не превосходит $\delta > 0$, то:

$$\|R(x) - R_F(x)\|_{\mathbb{R}_m} \leq \sum_{i=0}^n \mu_i \|R(x) - R_F(a_i)\|_{\mathbb{R}_m} \leq \omega_R(\delta).$$

Следовательно, получается оценка сверху для объема N обучающего набора \tilde{R}_N симплициальной ИНС:

$$N \geq (n+1)\text{vol}(K)\delta^{-1},$$

где $\text{vol}(K)$ — оценка объема симплициального комплекса K . Отметим, что в сочетании с заданной точностью ИНС ε , выраженной через модуль непрерывности $\omega_R(\delta)$, получаем условие:

$$\varepsilon \approx \omega_R \left(\frac{(n+1)\text{vol}(K)}{N} \right). \quad (4)$$

В частности, для гладкого отображения R погрешность приближения R_F на комплексе K пропорциональна величине N^{-1} . Однако в большинстве практических задач разумная априорная информация о модуле непрерывности ω_R отсутствует, что приводит к возможности грубых ошибок аппроксимации при интуитивном выборе объема N обучающего множества \tilde{R}_N . Ясно, что в силу этих причин не существует универсальной по точности симплициальной ИНС на заданном объеме N обучающего набора \tilde{R}_N . Разумной является постановка задачи о точности симплициальных ИНС на классах непрерывных отображений в зависимости от априорной информации о модуле непрерывности $\omega_R(\delta)$.

Отметим также, что в силу монотонного возрастания $\omega_R(\delta)$ существенное влияние на величину погрешности аппроксимации в формуле (4) оказывает размерность n пространства аргументов $x \in K$. Это имеет значение для входных данных большой размерности $n \gg 1$, в частности, эта ситуация типична при обработке видеoinформации и распознавании образов.

Следует подчеркнуть, что эффективность рассматриваемых ИНС обусловлена тем, что «предсказываемое» значение функции $R(x)$ обеспечивается «окружением» точки x внутри симплекса $T \in K$ достаточно малого диаметра δ , в вершинах которого заданы обучающие значения из R_N , которые за счет линейной интерполяции распространяются на точку x . При этом ее барицентрические координаты $\{\mu_i\}$ в точности являются весовыми коэффициентами связей в симплициальной ИНС, соответствующей рис. Естественно, что по построению ИНС в силу приведенных конструктивных особенностей принципиально не могут быть использованы для разумных с точки зрения точности постановок задач прогноза, выходящих за границы комплекса K (экстраполяции). Для этих целей необходимо рассмотрение обучающих наборов на более широких комплексах $K' \supset K$, что, несомненно, требует экспертных оценок Человека.

П.Л. Чебышёвым при рассмотрении задачи о многочлене, наименее уклоняющемся от нуля, была решена задача о минимизации ошибки интерполяции заданного класса функций за счет расположения узлов интерполирования по корням многочленов Чебышёва [8]. Очевидно, что это служит отправной точкой для широкого класса исследований по оптимизации выбора обучающих наборов для построения ИНС.

Важнейшим аспектом практического применения ИНС является существенное снижение объема N обучающего множества. На этом пути эффективные решения связаны с применением процедуры последовательного усреднения аппроксимаций на основе ЗБЧ (закона больших чисел П.Л. Чебышёва), дающего существенное ускорение сходимости частичных сумм в теории тригонометрических рядов Фурье [9]. Аналогичный эффект связан с теоремой Мазура [1], позволяющей за счет процедуры усреднения превратить слабо сходящиеся последовательности в сильно сходящиеся. Близкое явление имеет место в методе Ричардсона ускорения сходимости разностных схем на грубых сетках [10].

Важный пример ускорения сходимости демонстрирует усреднение, примененное к бесконечной последовательности $\{e_i\}_{i=1}^{\infty}$ ортонормированных векторов в гильбертовом пространстве H со скалярным произведением, равным символу Кронекера:

$$(e_i, e_j)_H = \delta_{i,j}.$$

Очевидно, что в силу равенства Парсеваля последовательность e_i слабо сходится к нулю при $i \rightarrow \infty$, но при этом она не имеет предела в норме пространства H . В то же время процедура вычисления среднего арифметического значений e_i :

$$S_n = \frac{1}{n} \sum_{i=1}^n e_i, \quad n \in \mathbb{N},$$

обеспечивает сильную сходимость последовательности $S_n \rightarrow 0$, $n \rightarrow \infty$ (в норме пространства H). По сути, именно этот прием используется для обоснования плотности множества тригонометрических многочленов на компактах (и, соответственно, алгебраических многочленов) в пространстве непрерывных функций в норме C [9]). Очевидно, что такой подход может быть положен в основу создания малоразмерных симплициальных ИНС¹.

В частности, на этой базе можно рассматривать подход к построению приближенных решений системы Навье–Стокса для несжимаемой жидкости на основе серии точных решений [11]. Очевидно, что точные решения являются в этом случае обучающим множеством для соответствующей ИНС.

Отметим, что так называемые многослойные ИНС представляют собой суперпозицию нескольких симплициальных ИНС, в которых могут реализовываться также обратные связи и т.п. Вопросы их декомпозиции, точности, эффективности в настоящее время являются, по большей мере, областью инженерных вычислительных экспериментов.

Важным моментом для будущих исследований математической природы ИНС является отыскание связанных с ними топологических инвариантов (гомотопических групп) и отыскание условий эквивалентности структуры различных математических моделей, реализующих ИНС.

Предложенная конструкция построения вышеупомянутых симплициальных ИНС, как и большинство публикаций по разработке ИНС, основанных на блок-схеме (рис.), идейно связана с работами А.Н. Колмогорова и В.И. Арнольда [13] о представлении непрерывных функций. В 1957 г. ими была доказана теорема (напрямую относящаяся к 13-й проблеме Д. Гильберта), что любая непрерывная функция R многих переменных может быть представлена в виде суперпозиции операций сложения и конечного набора функций одной переменной.

Теорема 2. (А.Н. Колмогоров, В.И. Арнольд.) Пусть $R : [0,1]^n \rightarrow \mathbb{R}$ является произвольной непрерывной функцией n переменных x_1, x_2, \dots, x_n . Тогда она может быть представлена в следующем виде:

$$R(x_1, x_2, \dots, x_n) = \sum_{i=1}^{2n+1} \Phi_i \left(\sum_{k=1}^n \psi_{i,k}(x_k) \right),$$

где все $\Phi_i : \mathbb{R} \rightarrow \mathbb{R}$ и $\psi_{i,k} : \mathbb{R} \rightarrow \mathbb{R}$ являются непрерывными функциями одной переменной, при этом внутренние функции $\psi_{i,k}$ выбираются независимо от вида R .

Этот результат послужил отправной точкой для исследований [14–16] по разработке методов глубокого обучения на базе ИНС.

ИНС, основанные на интерполяции данных, являются исключительными и удобными с точки зрения их детального исследования, теоретического анализа, поскольку минимум оптимизирующего функционала приводит к точно решаемым задачам, на которых можно исследовать принципиально важные проблемы построения и функционирования ИНС.

На практике размерность пространства входных данных существенно больше размерности выходных данных. Типичное использование ИНС нацелено на «сжатие» обрабатываемого потока больших объемов входных данных с целью их классификации, распознавания объектов, видео и акустической информации и т.п.

¹ Общая идея существенного снижения размерности вычислительных методов предложена в докладе акад. В.Б. Бетелина и проф. В.А. Галкина «О проблеме снижения размерности сеточных аппроксимаций» на заседании Президиума РАН 16.02.2021, см. [12].

Процедура сжатия больших объемов данных может быть рассмотрена как действие компактного непрерывного оператора R в гильбертовом пространстве H , т.е. $R : H \rightarrow H$. Соответственно, процедуру «сжатия» входной информации под действием оператора R можно трактовать как то, что ограниченные множества входных данных Big Data x в бесконечномерном гильбертовом пространстве H преобразуются ИНС в предкомпактные подмножества выходных данных $y \in H$ (почти конечные множества, т.е. хорошо аппроксимируемые конечными ε -сетями). Задачи распознавания образов, задачи их классификации можно сформулировать как отыскание решения $x \in H$ обратной задачи:

$$R(x) = y \quad (5)$$

на основе полученных данных $y \in H$, лежащих в предкомпактном множестве. Существенной для этого класса задач является теорема Бэра [9] о невозможности представления полного метрического пространства в виде счетного объединения нигде не плотных подмножеств. Центральной проблемой для отыскания решений обратной задачи (5) с компактным непрерывным оператором R на гильбертовом пространстве H является либо отсутствие решения уравнения (5) (т.е. не при всех значениях $y \in H$ существует решение уравнения (5)), либо отсутствие свойства непрерывности у обратного оператора R^{-1} на образе $R(H) \subset H$ — принципиальная вычислительная неустойчивость решений задачи (5).

Теорема 3. Пусть $R : H_1 \rightarrow H_2$ является непрерывным компактным отображением на гильбертовых пространствах H_1, H_2 . Пусть на образе $R(H_1) \subset H_2$ определено обратное отображение $R^{-1} : R(H_1) \rightarrow H_1$. Тогда множество точек $y \in H_2$, для которых не существует решение уравнения (5), является всюду плотным в H_2 , а оператор $R^{-1} : R(H_1) \rightarrow H_1$ не является непрерывным ни в каком открытом шаре в индуцированной метрике пространством H_2 на образе $R(H_1)$.

Доказательство. Выберем произвольную точку $y_0 \in R(H_1)$ и для произвольного положительного числа ε рассмотрим шар $\bar{B}_\varepsilon(y_0) = \{y \in H_1 : \|y - y_0\|_{H_1} \leq \varepsilon\}$. Следуя рассуждениям [9] для обобщенного уравнения Фредгольма 1-го рода, покажем наличие в этом шаре точек y , в которых уравнение (5) не имеет решения. Отметим, что пересечение $\bar{B}_\varepsilon(y_0) \cap R(H_1) \neq \bar{B}_\varepsilon(y_0)$. Действительно, предполагая противное, получаем равенство $\bar{B}_\varepsilon(y_0) = \bigcup_{n=1}^{\infty} \{R(\bar{B}_n(R^{-1}(y_0))) \cap \bar{B}_\varepsilon(y_0)\}$. В силу компактности отображения $R : H_1 \rightarrow H_2$ получаем компактность каждого множества под знаком этого объединения. Но компактные подмножества являются нигде не плотными в гильбертовом пространстве H_2 . Поскольку $\bar{B}_\varepsilon(y_0)$ является полным метрическим пространством в метрике, индуцированной объемлющим гильбертовым пространством H_2 , то в силу теоремы Бэра [9] $\bar{B}_\varepsilon(y_0)$ не может быть представлено в виде объединения указанных компактных множеств. Таким образом, в произвольном шаре из H_2 с центром в каждой точке образа $y_0 \in R(H_1)$ находятся точки, для которых не существует решение уравнения (5). Следовательно, множество точек неразрешимости этого уравнения всюду плотное в H_2 .

Покажем теперь, что сужение $R^{-1} : R(H_1) \rightarrow H_1$ не может быть непрерывным в каком-либо открытом шаре в индуцированной метрике. Предполагая противное, получаем, что существуют такая точка $y_0 \in R(H_1)$ и открытый шар в индуцированной метрике из H_2 на $R(H_1)$, где отображение $R^{-1} : R(H_1) \rightarrow H_1$ является непрерывным. Но тогда в гильбертовом пространстве H_1 можно указать открытый шар B_δ , $\delta > 0$, с центром в точке $R^{-1}(y_0)$, который гомеоморфно отображается на компактную окрестность точки y_0 . Открытый шар в гильбертовом пространстве не является компактным множеством [9]. Следовательно, оператор $R^{-1} : R(H_1) \rightarrow H_1$ не является непрерывным на всюду плотном подмножестве в индуцированной метрике из H_2 на $R(H_1)$. Теорема доказана.

Отметим, что для линейных операторов R , удовлетворяющих условиям этой теоремы, отсутствие непрерывности у обратного оператора R^{-1} имеет место в каждой точке образа $R(H_1) \subset H_2$ [9]. Поскольку это явление справедливо на всюду плотном во множестве выходных данных $y \in R(H_1)$, то обработка больших потоков данных на основе ИНС с целью классификации и распознавания входных данных x , основанная на решении задачи (5) с компактным оператором R , является всегда некорректной задачей и устойчивое получение аппроксимации решения x дополнительно требует применения методов регуляризации [17]. В частности, это явление типично для аппаратной цифровой обработки данных на основе частотной фильтрации сигналов R с подавлением высокочастотной компоненты входного сигнала $x \in H_1 = L_2[a, b]$. Положим $R(x) = (k_1 c_1, k_2 c_2, \dots, k_n c_n, \dots)$, где вектор $(c_1, c_2, \dots, c_n, \dots) \in l_2 = H_2$ состоит из коэффициентов разложения в ряд Фурье вектора x по некоторой

полной ортонормированной системе в $L_2[a, b]$, а заданная последовательность $k_n \rightarrow 0$, $k_n \neq 0$ определяет частотный фильтр входного сигнала x . Очевидно, фильтр $R : L_2[a, b] \rightarrow l_2$ является компактным оператором и, соответственно, восстановление исходного образа (сигнала) x на основе решения задачи (5) для приближенных (зашумленных) данных $y \in l_2$ является некорректной задачей, у которой либо отсутствует решение, либо малые погрешности данных приводят к значительным ошибкам решения.

Таким образом, классы задач, основанные на применении ИНС как средства аппроксимации отображений и построения решений обратных задач с компактным оператором, обнаруживают общее математическое свойство — вычислительную неустойчивость при обработке больших массивов данных. Эти проблемы аналогичны также для задач «восстановления» образов по «цифровым двойникам», поскольку имеют ту же математическую природу.

ЛИТЕРАТУРА

1. Данфорд Н., Шварц Дж.Т. *Линейные операторы. Общая теория*. Т. 1. М: ИЛ; 1962.
2. Бахвалов Н. С. *Численные методы*. М: Наука; 1973.
3. Runge K. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. *Zeitschrift für Mathematik und Physik*. 1901;46:224–243.
4. Бернштейн С. Н. *Собрание сочинений*. Т. 1–4. М., 1952–1964.
5. Дзядык В. К. *Введение в теорию равномерного приближения функций полиномами*. М.: Наука; 1977.
6. Калиткин Н. Н. *Численные методы*. М.: Наука; 1988.
7. Александров П. С., Пасынков Б. А. *Введение в теорию размерности*. М.: Наука; 1973.
8. Березин И. С., Жидков Н. П. *Методы вычислений*. М.: ГИФМЛ; 1962.
9. Колмогоров А. Н., Фомин С. В. *Элементы теории функций и функционального анализа*. М.: Наука; 1972.
10. Николаев Е. С., Самарский А. А. Выбор итерационных параметров в методе Рундсона. *ЖВМ и МФ*. 1972;12:960–973.
11. Бетелин В. Б., Галкин В. А., Дубовик А. О. Точные решения системы Навье–Стокса для несжимаемой жидкости в случае задач, связанных с нефтегазовой отраслью. *Доклады Российской академии наук. Математика, информатика, процессы управления*. 2020;495:13–16.
12. *Новости ФГУ ФНЦ НИИСИ РАН*. Режим доступа: <https://www.niisi.ru/news.htm>.
13. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения. *ДАН СССР*. 1957;114(5):953–956.
14. Hecht-Nielsen R. Kolmogorov's Mapping Neural Network Existence Theorem. *Proceedings of the International Conference on Neural Networks*. New York: IEEE Press; 1987;III:11–14.
15. Funahashi K. On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks*. 1989;2(3):183–192.
16. Cybenko G. Approximation by Superposition of a Sigmoidal Function. *Math. Control, Signals and Systems*. 1989;2:303–314.
17. Тихонов А. Н., Арсенин В. Я. *Методы решения некорректных задач*. М.: Наука; 1986.