

**Математические проблемы создания искусственных
нейронных сетей и искусственный интеллект**

***Mathematical problems of creating artificial
neural networks and artificial intelligence***

創造人工の數學問題 神經網路和人工智慧

Бетелин В.Б.

академик РАН

ФГУ ФНЦ НИИСИ РАН, Москва, Россия

Academician of the Russian Academy of Sciences

FGU FSC NIISI RAS, Moscow, Russia

俄羅斯科學院院士 FGU FSC NIISI RAS, 俄羅斯莫斯科

betelin@niisi.msk.ru

Галкин В.А.

доктор физико-математических наук

ФГУ ФНЦ НИИСИ РАН, Москва, Россия

Doctor of Physical and Mathematical Sciences

FGU FSC NIISI RAS, Moscow, Russia

物理和數學科學博士 FGU FSC NIISI RAS, 俄羅斯莫斯科

val-gal@yandex.ru

Аннотация. Предложен общий подход для анализа искусственных нейронных сетей на основе исследования свойств аппроксимации отношений и отображений их приближениями. Выявлены существенные для этого класса задач явления вычислительной неустойчивости, связанной с общими проблемами некорректных задач в гильбертовом пространстве и методами их регуляризации А.Н.Тихонова, типичными для обработки Big Data. Сформулированы критерии точности и применимости моделей искусственных нейронных сетей, применимость к разработке ИИ, 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, ИИ, методы оптимизации, вычислительная неустойчивость, методы регуляризации

Annotation. A general approach proposed for the analysis of artificial neural networks based on the study of the properties for approximation of relations and mappings by their approximations. The phenomena of computational instability, which are significant for this class of problems and are associated with the general problems of ill-posed problems in the Hilbert space and the methods of their regularization by A.N. Tikhonov, typical for Big Data processing, have been identified. Criteria for the accuracy and applicability of artificial neural network models and applicability to AI development formulated, and examples of their implementation based on the theory of function

interpolation are considered. The development of P.L. Chebyshev's ideas about the best approximation serves as the starting point for a wide class of mathematical studies on optimizing training sets for building ANNs.

Keywords: artificial neural networks (ANN), artificial intelligence (AI), optimization methods, computational instability, regularization methods

註解。 基於對關係近似性質及其近似映射的研究，提出了一種用於分析人工神經網路的通用方法。計算不穩定現象對於此類問題非常重要，並且與希爾伯特空間中不適定問題的一般問題及其大數據處理典型的 A. N. Tikhonov 正則化方法相關，已經被識別。制定了人工神經網路模型的準確性和適用性以及人工智慧開發的適用性標準，並考慮了基於函數插值理論的實施範例。P. L. 切比雪夫關於最佳近似的思想的發展是關於優化構建人工神經網路的訓練集的廣泛數學研究的起點

關鍵字：人工神經網路、人工智慧、最佳化方法、運算不穩定性、正規化方法

1. О необходимости математического обоснования технологий искусственного интеллекта

Технологии искусственного интеллекта, под которыми, как правило, понимаются искусственные нейронные сети, по сути дела, представляют собой технологии построения по конечному набору данных из некоторого множества (конечномерного или бесконечномерного) аппроксимации функции отображения этого множества в заданное конечномерное множество при условии минимизации некоторого функционала. Сходимость и устойчивость таких аппроксимаций для конечномерных множеств – это теорема А.Н. Тихонова о неподвижных точках отображения на упорядоченных ограниченных множествах. Приложения этой теоремы А.Н. Тихонова и являются математическим обоснованием успеха ИИ типа «Цифровой гроссмейстер», который представляет собой, построенную авторами этого ИИ, аппроксимацию функции отображения конечного упорядоченного множества многих тысяч партий, сыгранных реальными гроссмейстерами, в конечное множество из одной партии при условии минимизации ее близости к выигрышу. По сути дела, «цифровой гроссмейстер» – это поисковик Google, на конечном множестве партий, сыгранных реальными гроссмейстерами, и дополненный авторами этого ИИ метрикой близости партии к выигрышу. Полностью аналогично и математическое обоснование успеха ChatGPT, который также представляет собой аналог поисковика Google, на конечном упорядоченном множестве многих десятков и сотен тысяч письменных произведений реальных авторов (тоже BIG DATA), дополненный авторами этого чата метрикой близости различных фрагментов письменных произведений к заданному фрагменту. В случае бесконечномерных пространств, например, при распознавании и классификации изображений, построенная эмпирическим методом аппроксимация компактного отображения обладает принципиальной неустойчивостью, как будет показано в разделе 2.

2. Математическая постановка задачи создания искусственных нейронных сетей (ИНС)

Многочисленные полуэмпирические инженерные описания ИНС, применяемых для решения различных практических задач, в целом сводятся к следующим двум связанным математическим задачам:

- а) По заданному набору экспериментальных точек (X, Y) «придумать закон – оптимальное интерполирующее отображение» $R: X \rightarrow Y$.
- б) Для построенного отображения R найти решения уравнения

$$R(x)=c \quad (1)$$

Множество корней $x(c)=R^{-1}(c)$ уравнения (1) (множество уровня c) используется для «классификации» входных данных ИНС $x(c)$. В большинстве типичных задач уравнение (1) может рассматриваться как задача об отыскании неподвижной точки некоторого отображения $f: M \rightarrow M$

$$\bar{x} = f(\bar{x}) \quad (2)$$

В 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 а), б) и (1), (2). 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, соединённых связями и обменивающихся между собой входными-выходными данными (см., например, рис. 1).

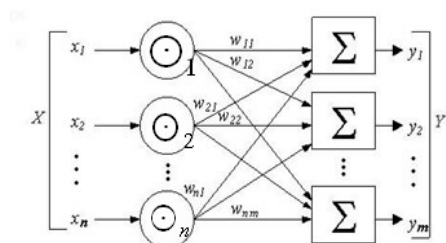


Рис. 1. Пример мнемосхемы структуры элементарной ячейки одного слоя ИНС

Fig. 1. An example of a mnemonic diagram of an ANN elementary cell structure on one layer

圖 1. ANN 一層單元結構助記圖範例

Цель создания ИНС. Настройка ИНС (аналог процедуры автоматизированного программирования структуры ИНС) состоит в подборе семейства искусственных нейронов (преобразований, обозначаемых на рис. 1 значками $\Theta_1, \Theta_2, \dots, \Theta_n$, в вершинах графа) и коэффициентов $w_{i,j}$ умножения значений преобразованных величин вдоль линий соединения вершин на основе дополнительно декларируемых принципов (например, оптимизации). Набор обучающих примеров составляет заданное подмножество $\tilde{R} \subset R$, где отношение R априори неизвестно. Целью создания ИНС является в некотором смысле «оптимальная реконструкция» неизвестного отношения R на основе заданного «обучающего» набора $\tilde{R}$. По своей природе такая постановка задачи является некорректной ввиду существенной множественности её решений («школьная» математически некорректная задача об отыскании кривой, соединяющих заданный набор точек).

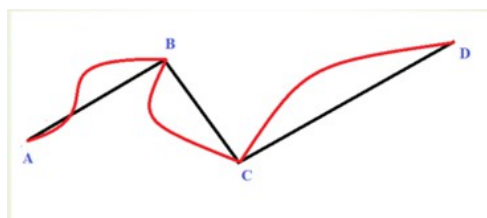


Рис. 2. «Школьная некорректная задача» об отыскании закона, связывающего экспериментальные точки A, B, C, D, имеющая бесконечное множество решений

Fig. 2. “School ill-posed problem” about finding a law connecting experimental points A, B, C, D, which has an infinite number of solutions

圖 2. “學校不適定問題”，尋找連接實驗點 A、B、C、D 的規律，該規律有無限多個解

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 $\tilde{R}$ до отношения R_F , которое аппроксимирует R (например, посредством минимизации функционала F на некотором множестве параметров). Назовём R_F математической моделью реализации ИНС. Описанная схема обычно применяется для построения продолжения функций, в частности, в задачах интерполяции.

Что математически означает термин «обучение»? Процесс «обучения» является процедурой настройки весов w_k функций активации Θ_k с

целью уменьшения «расстояния» в некоторой заданной метрике между желаемыми (целевыми) и получаемыми векторами на выходе. В конечном счете – это решение задачи синтеза преобразования R , для которого на заданном множестве «обучающих» пар (x, y) выполняется соотношение $y=R(x)$. Отысканию подлежат веса w_k и функции активации Θ_k относительно априори выбранной процедуры оптимизации для заданного функционала (метрики).

Типичная постановка теории приближенного вычисления значений функции F 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 $\tilde{R}(x, w_1, w_2, \dots, w_k) \approx R(x)$, где $(w_1, w_2, \dots, w_k)$ - множество параметров, по которым проводится вышеупомянутая оптимизация. Например, для заданных значений аргументов $\tilde{x}^{(l)} \in D, l=1, 2, \dots, N$ требуется выполнение минимальности уклонения нормы

$F(w_1, w_2, \dots, w_k) = \max_{1 \leq l \leq N} \|\tilde{R}(\tilde{x}^{(l)}, w_1, w_2, \dots, w_k) - \tilde{y}^{(l)}\|$ в предположении, что пары $(\tilde{x}^{(l)}, \tilde{y}^{(l)})$, $1 \leq l \leq N$ находятся в заданном отношении $\tilde{R}$.

Такой подход типичен для задач прогноза, регрессии, построения интерполяционных многочленов [2]. Если условие минимума уклонения означает совпадение значений $\tilde{y}^{(l)} = R(\tilde{x}^{(l)}) = \tilde{R}(\tilde{x}^{(l)}, w_1, w_2, \dots, w_k), 1 \leq l \leq N$, 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].

Задачи построения ИНС на основе многочленов многих переменных существенным образом связаны с приближением достаточно гладких отображений частичными суммами соответствующего ряда Тейлора, что требует вычисления оценок производных неизвестной функции на обучающем множестве. Это приводит к необходимости решения интегральных уравнений Вольтерра (Фредгольма) 1-го рода типа свёртки с приближённо заданной правой частью. Собственно говоря, этот подход положен в основу популярного названия «свёрточные ИНС». Отметим, что уравнение Фредгольма 1-го рода на пространстве суммируемых функций является типичной некорректной задачей, обладающей выраженной вычислительной неустойчивостью, которая преодолевается на основе методов регуляризации А.Н.Тихонова [17].

Фундаментальные проблемы устойчивости и сходимости. 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 её структуры (а именно, к динамическому изменению набора функций искусственных нейронов Θ_i , правила выбора весовых коэффициентов w_i , изменению 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 надёжность.
8. Зависимость точности и быстродействия ИНС от её размерности.
9. Возможность автоматического понижения размерности ИНС в процессе эксплуатации с целью оптимизации её 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). Соответственно, в дальнейшем под технологией ИИ мы понимаем упорядоченную иерархию связанных между собой ИНС на основе априори введённого Человеческим интеллектом отношения порядка на этом множестве. Очевидно, что при таком подходе ИИ является производным работы естественного (человеческого) интеллекта.

Согласно теореме Гёделя в любой непротиворечивой формальной системе на языке этой системы можно сформулировать утверждение, не выводимое вместе со своим отрицанием, то есть формальная система является неполной. Для разрешения неопределенности необходимо добавить это утверждение или его отрицание в состав аксиом рассматриваемой формальной системы. Для машины Тьюринга это означает наличие внешней по отношению к ней более интеллектуально развитой системы, которой каждый раз необходимо выбирать,

какую новую аксиому добавить в формальную систему и в соответствии с этим выбором исправить алгоритм работы машины Тьюринга. Такие интеллектуальные системы могут образовывать иерархию все более сложных и всеобъемлющих систем, на вершине которой находится Человек. Необходимость такой иерархии следует из теоремы Левенгейма–Сколема, которая утверждает, что любая система аксиом допускает намного больше существенно различных интерпретаций, чем предполагалось при ее создании. Аксиомы не устанавливают пределов для интерпретаций или моделей, и поэтому необходимо существование корректора — Человека. Аксиоматические системы, к которым применима теорема Левенгейма–Сколема, предназначаются для задания одной вполне конкретной интерпретации, и, будучи применимыми к совершенно различным моделям, они тем самым не соответствуют своему назначению. Подводя итог сказанному, можно отметить, что прорыв, произведенный ИНС и методами глубокого обучения, обеспечил решение задач, отражающих периферийные функции человека. Сам процесс обдумывания, планирования своих действий («поведенческое» управление), логического «рассуждения» и абстрактного мышления пока не воспроизводится работой искусственных нейронных сетей. Эти процессы так или иначе будут воспроизводиться на основе объединения различных методов ИИ (не только и не столько нейросетевых) с имитационным моделированием [18].

Естественным развитием упомянутых выше интерполяционных идей построения ИНС и сопутствующей иерархической системы ИИ является интерполяция в пространствах логических переменных и применение оптимизационных процедур на упорядоченной структуре логического вывода, связанного с теоремами Гёделя и Левенгейма–Сколема.

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

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 её 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 объёмов обучающих наборов и как следствие — границ реализуемости и безопасной

применимости систем ИИ, построенных на комплексах взаимодействующих ИНС.

Подчеркнём, что классическая теорема Вейерштрасса [5] о плотности множества многочленов в пространстве непрерывных функций на компакте позволяет утверждать существование сходящихся ИНС, построенных на базе многочленов, 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 основе многочленов, удовлетворяющих условию (3) и обучающих множеств $\tilde{R}_N$ при $N \gg 1$.

Важно отметить, что симплициальные аппроксимации многомерных непрерывных отображений на симплициальных компактных комплексах имеют чётко выраженную структуру ИНС (которую назовём симплициальной ИНС), представленную на рис.1. 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:

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

Подчеркнём, 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

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

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

объема симплицеального комплекса K . Отметим, что в сочетании с заданной точностью ИНС ε , выраженной через модуль непрерывности $\omega_R(\delta)$, получаем условие

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

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 симплицеальной ИНС на заданном объеме N обучающего набора $\tilde{R}_N$. Разумной является постановка задачи о точности симплицеальных 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 (6) 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 видеоинформации 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 счёт линейной интерполяции распространяются на точку x . При этом её барицентрические координаты $\{\mu_i\}$ в точности являются весовыми коэффициентами связей в симплицеальной ИНС, соответствующей рисунку 1. 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, выполненных Человеком.

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 счёт процедуры усреднения превратить слабо сходящиеся последовательности в сильно сходящиеся. Близкое явление имеет место в методе Ричардсона ускорения сходимости разностных схем на грубых сетках [11].

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 является, 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 мнемосхеме (рис. 1), 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 многих переменных может быть представлена в виде суперпозиции операций сложения и конечного набора функций одной переменной.

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 объёмов входных данных с целью их классификации, распознавания объектов, видео и акустической информации и т.п.

Процедура сжатия больших объёмов 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 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*

$$R(x)=y \quad (4)$$

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 неплотных подмножеств. Центральной проблемой для отыскания решений обратной задачи (4) с компактным непрерывным оператором R на гильбертовом пространстве H является либо отсутствие решения уравнения (4) (т.е. не при всех значениях $y \in H$ существует решение уравнения (4)), либо отсутствие свойства непрерывности у обратного оператора R^{-1} на образе $R(H) \subset H$ - принципиальная вычислительная неустойчивость решений задачи (4).

Теорема 1, [10]. 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 (4), 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 $R(H_1)$.

Отметим, что для линейных операторов R , удовлетворяющих условиям этой теоремы, отсутствие непрерывности у обратного оператора R^{-1} имеет место в каждой точке образа $R(H_1) \subset H_2$ [11]. 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 (4) 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 с подавлением высокочастотной компоненты входного сигнала $x \in H_1 = L_2[a, b]$.

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.

Модели ИНС и задачи о неподвижной точке. Построение модели ИНС может рассматриваться как задача (2) об отыскании неподвижной точки $\bar{x} (\bar{x} = f(\bar{x}))$ для некоторого отображения $f: M \rightarrow M$ на множестве данных $M \supset X \cup Y$. Выбор отображения f назовём слоем ИНС. По существу, построение слоя ИНС является этапом построения многослойной ИНС, состоящей из последовательно применяемых преобразований $f_n: M \rightarrow M$ – слоёв, которые последовательно модифицируются пользователем для достижения некоторого «оптимального, идеального для пользователя» результата. Искомый результат представляет собой неподвижную точку некоторого предельного преобразования $f = \lim_{n \rightarrow \infty} f_n$. В многих случаях процесс отыскания $\bar{x}$ основывается на рассмотрении последовательности итераций

$$x_{n+1} = f_n(x_n) \quad (5)$$

Часто при конкретных реализациях многослойных ИНС слои выбирают одинаковыми, т.е. итерации $f_n \equiv f$. Такой подход к созданию многослойных ИНС, вообще говоря, может не сходиться. А.Н.Шарковским для одномерных итераций $x_{n+1} = f_n(x_n)$ был обнаружен динамический хаос [19].

Таким образом, построение цепочки слоёв ИНС должно подчиняться модификациям, гарантирующих сходимость итераций.

Ниже рассмотрены алгоритмы, обладающие такими свойствами.

Теорема 2, [22]. Пусть (M, ρ) – полное метрическое пространство, на котором действует последовательность равностепенно сжимающих операторов

$$f_n: M \rightarrow M, n \in \mathbb{N} \quad (6)$$

т. е. постоянные сжатия для f_n не превосходят величины $0 \leq q < 1$. Пусть последовательность неподвижных точек $\bar{x}_n, n \in \mathbb{N}$, отображений f_n сходится к некоторой точке $\bar{x} \in M$. Тогда точка $\bar{x}$ может быть найдена как предел итераций

$$x_{n+1} = f_{n+1}(x_n), n = 0, 1, \dots, \quad (7)$$

где начальная точка $x_0 \in M$ выбирается произвольно.

Замечание. Достаточным условием сходимости точек $\bar{x}_n$ является поточечная сходимость отображений f_n на M .

Приведённая выше теорема связана с построением систем с единственным решением. Практические задачи классификации и распознавания образов

связаны с установлением соответствия поступающих данных одному из множества объектов, каждому из которых сопоставляется неподвижная точка многослойной ИНС. То есть для таких отображений неподвижная точка не единственна, поэтому одним из важнейших аспектов создания устойчивых многослойных ИНС является построение алгоритмов, стабилизирующих предельное поведение итераций (7), при отказе от требования теоремы 2, связанного со свойством равномерного сжатия. В частности, хаотическое поведение последовательностей (5), (6) обусловлено наличием подмножества, где отображение (6) является достаточно сильно растягивающим. Аналогичными свойством обладает достаточно широкий класс отображений, в который, естественно, попадают экспериментально создаваемые ИНС. Такое поведение типично для гиперболических динамических систем [21].

С указанными явлениями связано общее свойство неустойчивости поведения высокочастотных гармоник разложения отображений в ряд Фурье, а именно, высоким номерам базисных функций соответствуют большие коэффициенты растяжения, порождающие хаотическое поведение итераций (5).

Теорема 3, [22]. Пусть непрерывное отображение f на $\mathbb{R}_n$ оставляет инвариантным ограниченный замкнутый шар M . Пусть последовательность итераций (7) определяется формулами

$$x_n = f\left(\frac{1}{n} \sum_{i=0}^{n-1} x_i\right) \quad (8)$$

Предположим, что почти везде существует предел средних арифметических $\bar{x} = \frac{1}{n} \sum_{i=0}^{n-1} x_i$. Тогда последовательность итераций (8) сходится к неподвижной точке $\bar{x}$ отображения $f: M \rightarrow M$ почти для всех начальных точек $x_0 \in M$.

Таким образом, алгоритмы (8) могут быть положены в основу создания классов устойчивых сходящихся ИНС.

Открытым является фундаментальный вопрос о сходимости средних арифметических $\frac{1}{n} \sum_{i=0}^{n-1} x_i$. Он содержательно решён в рамках эргодической теоремы Биркгофа–Хинчина [23] для отображений $f: M \rightarrow M$, сохраняющих меру. Вычислительные эксперименты авторов демонстрируют надёжность и вычислительную устойчивость алгоритма (8) для широкого класса непрерывных отображений шара в себя.

Неподвижные точки отображений на упорядоченных пространствах

(монотонная сходимостъ итераций). Перспективным направлением построения и обоснования сходимости алгоритмов (2) для иерархии ИНС, которые составляют структуру ИИ является применение теорем о неподвижных точках на локально выпуклых и упорядоченных пространствах, связанных с работами А.Н.Тихонова, Ю.Шаудера и А.Тарского [1]. Эти теоремы позволяют во многих практических задачах для ИНС, основанных на сохраняющих порядок непрерывных отображениях, построить конструктивные сходящиеся алгоритмы к наименьшей (наибольшей) неподвижной точке. В частности, это утверждение имеет глубокие приложения в формальной семантике языков программирования (в абстрактной интерпретации – теории аппроксимации семантики компьютерных языков, связанной с именами П. и Р.Кузо [25]).

Методы погружения и применение ИНС для решения задач механики сплошной среды. Следует подчеркнуть, что в некотором смысле наиболее адаптированной к технологии ИНС, разработанной и универсальной является теория вычислительных методов для линейных задач. Поэтому с этой точки зрения важными являются процедуры «погружения» нелинейных задач в общие классы линейных за счет изменения исходной размерности описания и расширения функциональных пространств [26].

Отмеченное выше погружение нелинейных систем в линейную динамику может быть универсальным образом реализовано в рамках концепции функциональных решений для квазилинейных систем на основе вложения Янга соболевских обобщенных решений в пространство линейных функционалов, снабженное тихоновской топологией [27, 28].

Заметим, что условие слабой устойчивости [27, 28] по своей природе значительно проще понятия устойчивости в теории разностных схем для нелинейных задач, оно отражает «правильное» воспроизведение основных законов сохранения. Поэтому предложенный метод погружения нелинейных законов сохранения в линейные задачи на пространстве функциональных решений, снабжённом топологией А.Н.Тихонова, имеет универсальный вычислительный характер, реализуемый в рамках локально выпуклых ИНС, соответствующих мнемосхеме на рис. 1. Это обусловлено тем, что определение функционального решения естественным образом продолжается на минимальную выпуклую оболочку, содержащую функциональные решения для заданных начальных данных интегральной формы задачи Коши.

1. On the need for mathematical justification of artificial intelligence

technologies.

Artificial intelligence (AI) technologies, which, as a rule, are understood as artificial neural networks (ANN), in fact, are technologies for constructing approximation of the mapping from a finite set of data (belonging to finite-dimensional or infinite-dimensional spaces) into a given finite-dimensional set, provided that some functional is minimized. The convergence and stability of such approximations is A.N. Tikhonov's Theorem on fixed points of mappings on ordered bounded sets. The applications of this theorem are the mathematical justification for the success of AI of the "Digital Grandmaster" type. It was constructed by the authors of this AI as approximation for the mapping from a finite ordered set of many thousands of games played by real grandmasters into a finite set of one game, provided its distance to winning is minimized. In essence, the "digital grandmaster" is a Google search engine based on a finite set of games played by real grandmasters, and supplemented by the authors of this AI with a metric of the game's proximity to winning. The mathematical justification of the ChatGPT success, which is also an analogue of the Google search engine, is completely analogous to the finite ordered set of many tens and hundreds of thousands of written works by real authors (they are also Big Data). It is supplemented by the authors of this chat with the metric for evaluation the proximity of various fragments of written works to a given fragment. As will be shown in Section 2, in the case of infinite-dimensional spaces, for example, related to the problems of image recognition and classification, the approximation of compact mapping, constructed by the empirical methods, is fundamentally unstable.

2. Mathematical formulation of the problem of creating artificial neural networks.

Numerous semi-empirical engineering descriptions of ANN used to solve various practical problems are generally reduced to the following two related mathematical problems:

a) For a given set of experimental points "come up with a law – optimal interpolation mapping";

b) For the constructed mapping find solutions to the Equation

$$R(x)=c \quad (1)$$

The set of roots $x(c)=R^{-1}(c)$ of Equation (1) (the set of level c) is used to "classify" the input data of the ANN $x(c)$. In most typical problems, Equation (1) considered as a problem of finding a fixed point of some mapping $f:M \rightarrow M$

$$\bar{x}=f(\bar{x}) \quad (2)$$

In numerous computational experiments on pattern recognition, processing of acoustic, video and text information, approaches have been found to create software for the practical solution of a number of difficult-to-formalize problems (a), (b) and (1), (2). These successes have generated a huge flow of work and interest in the

creation of semi-empirical methods called artificial neural networks (ANN). The general approach underlying the construction of the ANN consists of accepting the hypothesis of the possibility of creating a device that can be trained on a series of examples of decision-making. This vague formulation assumes that there is some connection between the selected pairs of objects $(x, y) \in X \times Y$, which determines some relation $R \subset X \times Y$ (hypothetical law) [1]. It is usually assumed that R is a function (i.e., exactly one value $x \in X$ corresponds to each value $y \in Y$), while the arguments X are called causes, and the set of values are called Y consequences. Schematically, the ANN is a parallel system of distributed information processing in the form of a directed graph, where the vertices of the graph are called processing elements (artificial neurons), and the sluices between them are called connections. Thus, the ANN is a graph — a system of nodes connected by sluices, exchanging input and output data among themselves (see, for example, Fig. 1).

The purpose of creating an ANN. Setting up an ANN (analogous to the procedure of automated programming of the ANN structure) consists in selecting a family of artificial neurons (transformations indicated in Fig. 1 by icons $\Theta_1, \Theta_2, \dots, \Theta_n$, in the vertices of the graph) and multiplication coefficients $w_{i,j}$ to the values of the transformed quantities along the vertex connection lines based on additionally declared principles (for example, optimization). The collection of training examples is a given subset $\tilde{R} \subset R$, where the relation R is a priori unknown. The purpose of creating an ANN is (in a some sense) an "optimal reconstruction" of an unknown relation R based on a given "training" set $\tilde{R}$. By its nature, this formulation of the problem is ill-posed one due to the significant multiplicity of its solutions (for example, it is the "school problem" of finding a curve connecting a given set of points).

Thus, a necessary element of building an ANN is the introduction of a target rule (criterion, functional, etc.) F , based on the application of which the construction of the "best" continuation of a given relationship $\tilde{R}$ to a relationship R_F is performed. It approximates R (for example, by minimizing the functional F on a certain set of parameters). Let us call the relation R_F as mathematical model of the ANN implementation. The described scheme is usually applied to construct continuation of functions, in particular, in interpolation problems.

What does the term "learning" mean mathematically? The "learning" process is a procedure for adjusting the weights w_k of activation functions Θ_k in order to reduce the "distance" in some given metric between the desired (target) and the resulting vectors at the output. Ultimately, this is a solution to the problem of synthesis the transformation R , for which a relation $y=R(x)$ is performed on a given set of "training" pairs (x, y) . Weights w_k and activation functions Θ_k are to be found with respect to the a priori selected optimization procedure for a given functional

(metric). A typical formulation of approximate calculation of function F values on a set of arguments $x = (x_1, x_2, \dots, x_n) \in D \subset \mathbb{R}_n$ consists in choosing the optimal multi-parametric approximation from a given set of functions $\tilde{R}(x, w_1, w_2, \dots, w_k) \approx R(x)$, where $(w_1, w_2, \dots, w_k)$ is the set of parameters for which the above optimization is carried out. For example, for the given values of the arguments $\tilde{x}^{(l)} \in D, l = 1, 2, \dots, N$ the minimum of the deviation of the norm $F(w_1, w_2, \dots, w_k) = \max_{1 \leq l \leq N} \|\tilde{R}(\tilde{x}^{(l)}, w_1, w_2, \dots, w_k) - \tilde{y}^{(l)}\|$ is required, assuming that the pairs $(\tilde{x}^{(l)}, \tilde{y}^{(l)}), 1 \leq l \leq N$ are in a given relationship $\tilde{R}$. This approach is typical for forecasting, regression, and the construction of interpolation polynomials [2]. If the minimum deviation condition means that, the values coincide

$$\tilde{y}^{(l)} = R(\tilde{x}^{(l)}) = \tilde{R}(\tilde{x}^{(l)}, w_1, w_2, \dots, w_k), 1 \leq l \leq N,$$

then this method of constructing an approximation is called interpolation (in this case, the target functional $F(w_1, w_2, \dots, w_k)$ at the optimum point takes the smallest value equal to zero). Note that the functional definition F can be carried out in different ways based on expert assessments of the choice of metrics (distances) on a set of values Y that determine the goals of the tasks being solved. This kind of problems arise when planning experiments in the natural sciences, in economics, when construct pattern recognition algorithms, when solving inverse problems of object classification, when designing control systems, guidance, aiming, navigation, etc. [2].

The problems of constructing an ANN based on polynomials of many variables are essentially related to the approximation of sufficiently smooth mappings by partial sums of the corresponding Taylor series, which requires calculating estimates of derivatives of an unknown function on the training set. This leads to the need to solve Volterra (Fredholm) integral equations of the 1st kind of convolution type with an approximately given right-hand side. Strictly speaking, this approach is the basis of the popular name "convolutional ANN". Note that the Fredholm equation of the 1st kind on the space of summable functions is a typical ill-posed problem with pronounced computational instability, which is overcome on the basis of A.N. Tikhonov regularization methods [17].

Fundamental problems of stability and convergence. Note the important concepts for more general ANN that are associated with the considered example.

1. The memory capacity of the ANN, determined by the volume of the training set $\tilde{R}_N = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}$.
2. The accuracy of the ANN according to a given class of approximated relations R . Achievability of the set goals.
3. The learning rate of the ANN based on a given algorithm F .
4. The "flexibility" of the ANN in relation to the reconfiguration of its structure (namely, to a dynamic change in the set of functions of artificial neurons O_i , the rules

for choosing weight coefficients w_i , changing the training set $\tilde{R}_N$, etc.).

5. Performance of ANN.
6. Stability of the received output data when the input data changes.
7. Feasibility of computational procedures on artificial neurons, their reliability.
8. The dependence of the accuracy and speed of the ANN on its dimension.
9. The possibility of automatic reduction of the dimension of the ANN during operation in order to optimize its characteristics.
10. Fault tolerance, redundancy of functions and noise immunity of ANN.
11. Compatibility with other ANN due to the expansion of the structure based on the splicing of their graphs and mutual data exchange.
12. Topological invariants of the ANN structure and homotopy equivalence of various implementations of ANN.

Note that the content of the item 11 refers to a complex composed of ANNs of various nature, which can be characterized as an approximation of Artificial Intelligence systems. Accordingly, in the future, by AI technology we mean an ordered hierarchy of interconnected ANNs based on the a priori introduced by Human intelligence order relation on this set. Obviously, with this approach, AI is a derivative of the work of natural (human) intelligence. According to Gödel's Theorem, in any consistent formal system, in the language of this system, it is possible to formulate a statement that is not deducible together with its negation, i.e. the formal system is incomplete. To resolve the uncertainty, it is necessary to add this statement or its negation to the axioms of the formal system under consideration. For a Turing Machine, this means that there is a more intellectually developed system external to it, which each time needs to choose which new axiom to add to the formal system and, in accordance with this choice, correct the algorithm of the Turing Machine. Such intelligent systems can form a hierarchy of increasingly complex and comprehensive systems, at the top of which is a Person. The need for such a hierarchy follows from the Levenheim–Skolem Theorem, which states that any system of axioms allows for much more significantly different interpretations than was assumed when it was created. Axioms do not set limits for interpretations or models, and therefore the existence of a Human corrector is necessary. Axiomatic systems, to which the Levenheim–Skolem Theorem applies, are intended to set one very specific interpretation, and, being applicable to completely different models, they thereby do not correspond to their purpose. Summing up the above, it can be noted that the breakthrough made by ANNs applications and deep learning methods provided the solution of problems reflecting the peripheral functions of a Person. The very process of thinking, planning one's actions ("behavioral" management), logical "reasoning" and abstract thinking is not yet reproduced by the work of the Artificial Neural Networks. These processes will be reproduced one way or another on the basis of combining various methods and (not only and not so much Neural Network)

with simulation modeling [18].

The natural development of the above-mentioned interpolation ideas of constructing the ANNs and the accompanying hierarchical AI system is the interpolation in the spaces of logical variables and the application of optimization procedures on the ordered structure of logical inference associated with the Gödel and Levenheim–Skolem Theorems.

Big Data and the effectiveness of ANN training based on them. As for the emerging problems with the accuracy of the ANN with an unbounded increase in the volume of the training set $\tilde{R}_N$, the examples of K.D.T.Runge [3] and S.N.Bernstein [4] for interpolation polynomials with an unrestricted increase in the number of interpolation nodes are indicative here. The Runge's, Bernstein's examples and Faber's general result indicate the possibility of a deep misconception that the unlimited growth of neurons quantity in the ANN and large amounts of data $N \gg 1$ (Big Data), which form the basis of training sets $\tilde{R}_N$, can ensure its effectiveness within the framework of AI. The same applies to many statements about the construction of so-called "digital twins" due to the limitless expansion of training data sets $\tilde{R}_N = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}$.

These remarks relate to the general fundamental problem of analyzing the computational stability of the ANN when increasing the volume of training sets and, consequently, the boundaries of the feasibility and safe applicability of AI systems built on complexes of interacting ANNs.

We emphasize that the classical Weierstrass Theorem [5] on the density of the set of polynomials in the space of continuous functions on a compact allows us to assert the existence of convergent ANNs constructed based on polynomials. However, Faber's Theorem, as well as K.D.T.Runge's and S.N.Bernstein's examples indicate the risks of significant errors when using ANN with polynomials satisfying condition (3) and training sets $\tilde{R}_N$ at $N \gg 1$.

It is important to note that simplicial approximations of multidimensional continuous mappings on simplicial compact complexes have a well-defined ANN structure (which we will call simplicial ANN), shown in Fig.1. The procedure of the barycentric subdivision of simplicial complexes [7] on compacts in combination with the uniform continuity of the mappings on them allows us to assert that the simplicial ANN constructed in this way uniformly converge to on compacts with the diameter $\delta_N > 0$ of the barycentric subdivisions of the complex tending to zero:

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

We emphasize that the convergence rate of a simplicial ANN on a complex K for arbitrary continuous mappings R determined by the modulus of continuity $\omega_R(\delta) = \sup_{\|u-v\| \leq \delta} \|R(u) - R(v)\|_{C(K)}$, $\delta > 0$. It is not difficult to make sure that the approximation error $\|R_F - R\|_C$ on a simplicial complex K , which is compact set in $\mathbb{R}_n$, is estimated from above by value $\omega_R(\delta)$, where $\delta > 0$ is supremum of the

diameter of the simplexes in the complex $K=D$ on which the mapping R with values in $\mathbb{R}_m$ is defined. Indeed, on every simplex T included in the simplicial complex K , the formula is valid

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

where μ_i are the barycentric coordinates of a point x in a simplex T stretched over vertices $a_i, 0 \leq i \leq n$, n is the dimension of the simplex T . Thus, since the diameter of the simplex T does not exceed $\delta > 0$, then

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

Consequently, an upper bound is obtained for the volume N of the training set $\tilde{R}_N$ of the simplicial ANN $N \leq (n+1) \text{vol}(K) \delta^{-1}$, where $\text{vol}(K)$ is an estimate of the volume of the simplicial complex K . Note that in combination with a given accuracy of the ANN ε expressed in terms of the modulus of continuity $\omega_R(\delta)$, we obtain the condition

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

In particular, for a smooth mapping R the error of approximation function R_F on the complex is proportional to the value N^{-1} . However, in most practical problems there is no reasonable a priori information about the modulus of continuity ω_R , which leads to the possibility of gross approximation errors when intuitively choosing the volume N of the training set $\tilde{R}_N$. It is clear that for these reasons there is no universal simplicial ANN for a given volume N of the training set $\tilde{R}_N$. It is reasonable to formulate the problem of the accuracy of simplicial ANNs on classes of continuous mappings depending on a priori information about the $\omega_R(\delta)$.

Note also that, due to the monotonic increase $\omega_R(\delta)$, the dimension n of the argument space $x \in K$ has a significant impact on the approximation error. This is important for big dimension input data $n \gg 1$, in particular, this situation is typical when processing video information and pattern recognition. It should be emphasized that the effectiveness of the considered ANN is due to the fact that the "predicted" value of the function $R(x)$ is provided by the "environment" of a point x inside a simplex $T \in K$ of a sufficiently small diameter δ , at the vertices of which training values from R_N are set, which, due to linear interpolation, extend to the point. At the same time, its barycentric coordinates $\{\mu_i\}$ are exactly the weighting coefficients of the connections in the simplicial ANN corresponding to Fig. 1. Naturally, due to the above design features, the construction of the ANN of such type cannot be applicable with reasonable accuracy to problems of forecast in points, which go beyond the boundaries of the complex (extrapolation). For these purposes, it is

necessary to consider training sets on broader complexes $K' \supset K$, which undoubtedly requires expert assessments performed by a Person. Academician P.L.Chebyshev, when considering the problem of the polynomial least deviating from zero, solved the problem of minimizing the interpolation error of a given class of functions due to the location of interpolation nodes among the roots of Chebyshev's polynomials [8]. Obviously, this serves as a starting point for a wide class of studies on optimizing the selection of training sets for building an ANNs. The most important aspect of the practical application of ANNs is a significant reduction in the volume of the training set. In this way, effective solutions are associated with the application of the procedure of sequential averaging of approximations based on Great Numbers Law, which gives a significant acceleration of the convergence of partial sums in the theory of trigonometric Fourier series [9]. A similar effect is associated with Mazur's Theorem [1], which allows using the averaging procedure to turn weakly converging sequences into strongly converging ones. A similar phenomenon occurs in Richardson's method of accelerating convergence of difference schemes on coarse grids [11].

Note that the so-called multilayer ANNs are a superposition of several simplicial ANNs, in which feedbacks can also be implemented, etc. The issues of their decomposition, accuracy, efficiency are currently, at least, the field of engineering computational experiments.

An important point for future studies of the mathematical nature of ANN is the search for topological invariants (homotopy groups) associated with them and the search for equivalence conditions of the structure of various mathematical models implementing ANNs.

The proposed construction of the above-mentioned simplicial ANN, like most publications on the development of ANNs based on a mnemonic scheme (Fig. 1), is ideologically related to the works of A.N.Kolmogorov and V.I.Arnold [13] on the representation of continuous functions. In 1957, they proved a Theorem (directly related to D. Hilbert's 13th problem) that any continuous function of many variables can be represented as a superposition of addition operations and a finite set of functions of one variable.

ANNs based on data interpolation are exceptional and convenient from the point of view of their detailed research, theoretical analysis, since a minimum of optimizing functional leads to precisely solved problems on which fundamentally important problems of construction and functioning of ANNs can be investigated.

In practice, the dimension of the input data space is significantly larger than the dimension of the output data. Typical use of ANN is aimed at "compressing" the processed stream of large volumes of input data in order to classify them, recognize objects, video and acoustic information, etc.

The procedure for compressing large amounts of data we consider as the action of a compact continuous operator R in a Hilbert space H , i.e. $R:H \rightarrow H$.

Accordingly, the procedure of "compressing" input information under the action of the operator R interpreted as the fact that bounded sets of input Big Data x from an infinite-dimensional Hilbert space H transformed by the ANN into pre-compact subsets of output data $y \in H$ (almost finite sets, i.e. well approximated by finite ε -networks). The tasks of pattern recognition and their classification can be formulated as finding a solution to the inverse problem

$$R(x)=y \quad (4)$$

with given data $y \in H$ lying in the pre-compact set. Essential for this class of problems is Baire category Theorem [9]. The central problem for finding solutions to the inverse problem (4) with a compact continuous operator R on a Hilbert space H is either the absence of a solution to Equation (4) (i.e., there exists $y \in H$ when we have no solutions to Equation (4)), or the absence of the continuity property of the inverse operator R^{-1} on the image $R(H) \subset H$ - it is the fundamental computational instability of solutions to problem (4).

Theorem 1, [10]. *Let $R: H_1 \rightarrow H_2$ be a continuous compact mapping on Hilbert spaces H_1, H_2 . Let the inverse mapping $R^{-1}: R(H_1) \rightarrow H_1$ defined on the image $R(H_1) \subset H_2$. Then the set of points $y \in H_2$ for which there is no solution to the Equation is everywhere dense in H_2 , and the operator $R^{-1}: R(H_1) \rightarrow H_1$ is not continuous in any open ball in the induced metric by the space H_2 on the image $R(H_1)$.*

Note that for linear operators R satisfying the conditions of this Theorem, the absence of continuity in the inverse operator R^{-1} takes place at every point of the image $R(H_1) \subset H_2$ [11]. Since this phenomenon is true for everywhere dense in the set of output data $y \in R(H_1)$, processing of Big Data streams based on the ANN for the purpose of classification and recognition of input data, based on solving problem (4) with a compact operator, is always an ill-posed problem. Therefore, stable approximation of the solution additionally requires the use of regularization methods [17]. In particular, this phenomenon is typical for hardware digital data processing based on frequency filtering of signals with suppression of the high-frequency component of the input signal $x \in H_1 = L_2[a, b]$.

Thus, classes of problems based on the use of ANN as a means of approximating mappings and constructing solutions to inverse problems with a compact operator reveal a common mathematical property – computational instability when processing large data arrays. These problems are also similar for the tasks of "restoring" images by "digital twins", since they have the same mathematical nature.

ANN models and fixed-point problems. The construction of the ANN model can

be considered as the problem (2) of finding a fixed-point $\bar{x}(\bar{x}=f(\bar{x}))$ for some mapping $f:M \rightarrow M$ on a data set $M \supset X \cup Y$. The choice of mapping f is called the ANN layer. In essence, the construction of an ANN layer is a stage in the construction of a multilayer ANN consisting of consistently applied transformations $f_n:M \rightarrow M$ – layers that are consistently modified by the user to achieve some "optimal, ideal for the user" result. The desired result is a fixed-point of some limit transformation $f=\lim_{n \rightarrow \infty} f_n$. In many cases, the search process of $\bar{x}$ is based on considering a sequence of iterations

$$x_{n+1}=f_n(x_n) \quad (5)$$

Often, with specific implementations of multilayer ANN, the layers are chosen to be the same, i.e. iterations $f_n \equiv f$. This approach to the creation of multilayer ANN, generally speaking, may not converge. A.N.Sharkovsky discovered dynamic chaos for one-dimensional iterations $x_{n+1}=f_n(x_n)$ [19].

Thus, the construction of a chain of ANN layers should be subject to modifications that guarantee the convergence of iterations. Algorithms with such properties are considered below.

Theorem 2, [22]. *Let (M, ρ) be a complete metric space on which a sequence of uniformly compressing operators acts*

$$f_n:M \rightarrow M, n \in \mathbb{N} \quad (6)$$

that is, contraction constant for f_n do not exceed the value $0 \leq q < 1$. Let the sequence of fixed points $\bar{x}_n, n \in \mathbb{N}$, for mappings f_n converge to some point $\bar{x} \in M$. Then the point $\bar{x}$ can be found as the limit of iterations

$$x_{n+1}=f_{n+1}(x_n), n=0,1, \dots, \quad (7)$$

where the starting point $x_0 \in M$ is chosen arbitrarily.

Remark. A sufficient condition for the convergence of points $\bar{x}_n$ is the pointwise convergence of mappings f_n on M .

The above Theorem is related to the construction of systems with an unique solution. Practical tasks of classification and pattern recognition are associated with establishing the correspondence of incoming data to one of the many objects, each of which is associated with a fixed-point of a multilayer ANN. That is, for such mappings, a fixed-point is not the only one, therefore, one of the most important aspects of creating stable multilayer ANN is the construction of algorithms that stabilize the limiting behavior of iterations (7), while rejecting the requirement of Theorem 2 associated with the property of uniform contraction. In particular, the chaotic behavior of sequences (5), (6) is due to the presence of a subset where the mapping (6) is quite strongly stretching. A fairly wide class of mappings has a similar property, which, of course, includes experimentally created ANNs. This behavior is typical for hyperbolic dynamical systems [21].

These phenomena are associated with the general property the behavior instability of high-frequency harmonics for the decomposition of mappings into a

Fourier series, namely, to high-frequency basis functions corresponds large stretching coefficients that generate chaotic behavior of iterations (5).

Theorem 3, [22]. *Let a continuous mapping f on $\mathbb{R}_n$ maps a bounded closed ball M into itself. Let the sequence of iterations (7) be determined by the formulas*

$$x_n = f\left(\frac{1}{n} \sum_{i=0}^{n-1} x_i\right) . \quad (8)$$

Assume that there exists a limit of arithmetic averages $\bar{x} = \frac{1}{n} \sum_{i=0}^{n-1} x_i$ almost everywhere. Then the sequence of iterations (8) converges to a fixed point $\bar{x}$ of the mapping $f : M \rightarrow M$ for almost all initial points.

Thus, algorithms (8) can be used as a basis for creating classes of stable convergent ANNs.

The fundamental question of the convergence of arithmetic $\frac{1}{n} \sum_{i=0}^{n-1} x_i$ means is open. It is meaningfully solved within the framework of the Birkhoff–Khinchin ergodic theorem [23] for measure-preserving mappings $f : M \rightarrow M$. The authors' computational experiments demonstrate the reliability and computational stability of algorithm (8) for a wide class of continuous mappings of a ball into itself.

Fixed points of maps on ordered spaces (monotonic convergence of iterations). A promising direction for constructing and substantiating the convergence of algorithms (2) for the hierarchy of ANN that make up the structure of AI is the application of fixed-point theorems on locally convex and ordered spaces related to the works of A.N. Tikhonov, Ju. P. Schauder and A. Tarski [1]. These theorems make it possible to built constructive convergent algorithms to the smallest (largest) fixed-point in many practical problems for ANN based on order-preserving continuous mappings. In particular, this statement has deep applications in the formal semantics of programming languages (in an abstract interpretation – the theory of approximation of the semantics of computer languages associated with the names P. and R. Kuzo [25]).

Immersion methods and the use of ANNs for solving problems of continuum mechanics.

It should be emphasized that, in a sense, the theory of computational methods for linear problems is the most adapted to the ANN technology, developed and universal. Therefore, from this point of view, the procedures of "immersion" of nonlinear problems into general linear classes by changing the initial dimension of the description and expanding the functional spaces are important [26].

The above-mentioned immersion of nonlinear systems into linear dynamics can be implemented in a universal way within the framework of the concept of functional

solutions for quasi-linear systems based on the Young's embedding of Sobolev generalized solutions into the space of linear functionals equipped with Tikhonov topology [27, 28].

Note that the condition of weak stability [27, 28] is by its nature much simpler than the concept of stability in the theory of difference schemes for nonlinear problems, it reflects the "correct" reproduction of the basic conservation laws. Therefore, the proposed method of immersion of nonlinear conservation laws into linear problems on the space of functional solutions equipped with the topology of A.N.Tikhonov has a universal computational character, implemented within the framework of locally convex ANNs corresponding to the mnemonic diagram in Fig. 1. This is due to the fact that the definition of a functional solution naturally extends to a minimal convex set containing functional solutions for the given initial data of the integral form of the Cauchy problem.

1. 關於人工智慧技術數學論證的必要性

人工智慧技術通常被理解為人工神經網絡，本質上是從某個集合（有限維或無限維）中的有限數據集構造映射函數的近似值的技術將這個集合轉化為給定的有限維集合，服從某個泛函的最小化。有限維集的這種近似的收斂性和穩定性是 A. N. 的定理。吉洪諾夫論有序有界集映射的不動點。A. N. 對該定理的應用 Tikhonov 和是「數位大師」類型人工智慧成功的數學理由，該人工智慧由該人工智慧的作者構建，是映射數千個遊戲的有限有序集的函數的近似值由真正的大師將一場比賽分成有限的一組，但盡量減少與獲勝的距離。從本質上講，「數位特級大師」是一個谷歌搜尋引擎，基於真正的特級大師所玩的有限遊戲集，並由該人工智慧的作者補充，並提供了遊戲接近獲勝的指標。ChatGPT（也是 Google 搜尋引擎的類似物）在由真實作者（也是大數據）的數以萬計、數十萬篇書面作品組成的有限有序集合上取得成功的數學論證，並由該作者的作者進行了補充用書面作品的各個片段與給定片段的接近程度來度量聊天是完全相似的。在無限維空間的情況下，例如在影像辨識和分類中，經驗方法構造的緊映射近似從根本上來說是不穩定的，如第 2 節所示。

2. 創建人工神經網路 (ANN) 問題的數學表述

用於解決各種實際問題的人工神經網路的大量半經驗工程描述通常可歸結為以下兩個相關的數學問題：

- a) 對於給定的一組實驗點 (X, Y) ，「得出一個定律—最優插值映射」 $R: X \rightarrow Y$ 。
- b) 對於建構的映射 R ，求方程式的解

$$R(x)=c \quad (1)$$

方程式(1)的根集 $x(c)=R^{-1}(c)$ (水平集 c) 用於對 ANN 輸入資料 $x(c)$ 進行「分類」。在大多數典型問題中，方程式 (1) 可以被視為尋找某個映射 $f:M \rightarrow M$ 的不動點的問題

$$\bar{x} = f(\bar{x}) \quad . \quad (2)$$

在關於模式識別、聲音、視訊和文字資訊處理的大量計算實驗中，找到了創建軟體的方法，以實際解決許多難以形式化的問題 a)、b) 和 (1)、(2)。這些成功引發了大量的工作和興趣，人們對創建稱為人工神經網路 (ANN) 的半經驗方法產生了興趣。

建立人工神經網路的一般方法是接受這樣的假設：創建一個可以透過一系列範例進行訓練以做出決策的設備。這種模糊的表述假設選定的對象對之間存在某種聯繫 $(x, y) \in X \times Y$ ，這些聯繫定義了某種關係 $R \subset X \times Y$ (假設定律) [1]。通常假設 R 是函數 ($x \in X$ 即每個值恰好對應一個值 $y \in Y$)，參數 X 稱為原因，值集合 Y 稱為結果。簡單地說，人工神經網路是一個有向圖形式的分散式資訊處理的平行系統，其中圖的頂點稱為處理元素 (人工神經元)，它們之間的連接稱為連接。因此，人工神經網路是一個圖——一個透過連結連接並相互交換輸入輸出資料的節點系統 (例如，請參見圖 1)。

建立 ANN 的目的。設定 ANN (類似 ANN 結構自動編程的過程) 包括選擇一系列人工神經元 (轉換，如圖 1 所示，在頂點處以圖示

$\Theta_1, \Theta_2, \dots, \Theta_n$ 表示) 圖表的) 和係數 $w_{i,j}$ 根據附加聲明的原則 (例如，最佳化) 沿著連接頂點的線乘以變換量的值。訓練範例集構成給定的子集

$\tilde{R} \subset R$ ，其中比率 R 是先驗未知的。從某種意義上說，創建 ANN 的目標是基於給定的「訓練」集「最優地重建」未知關係 R 。從本質上講，這種問題的表述是不正確的，因為其解決方案具有顯著的多樣性 (「學校」數學上不正確的問題，即找到連接給定點集的曲線)。

因此，建構 ANN 的必要元素是引入目標規則 (準則、函數等) F ，基於該規則的應用，給定關係與關係 R_F 的「最佳」延續，該關係近似於 R (例如，透過一組參數最小化函數 F)。我們將 R_F 稱為實現 ANN 的數學模型。所描述的方案通常用於建構函數的連續性，特別是在插值問題中。

「學習」一詞在數學上意味著什麼？「訓練」過程是調整激活函數 Θ_k 的權重 w_k 的過程，以減少所需 (目標) 和結果輸出向量之間的某些給定度量中的「距離」。最終，這是合成變換 R 問題的解決方案，其中在給定的一組「訓練」對 (x, y) 上滿足關係 $y=R(x)$ 。權重 w_k 和激活函數 Θ_k 是相對於給定功能 (度量) 的先驗選擇的最佳化過程來找到的。

函數 F 在參數集 $x=(x_1, x_2, \dots, x_n) \in D \subset \mathbb{R}_n$ 上的值的近似計算理論的典型表述包括從給定的函數集 $\tilde{R}(x, w_1, w_2, \dots, w_k) \approx R(x)$ 中選擇最佳多參數近似，其中 $(w_1, w_2, \dots, w_k)$ 是參數集，透過該參數集進行上述最佳化。例如，對於給定的參數值 $\tilde{x}^{(l)} \in D, l=1, 2, \dots, N$ ，假設

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

對處於給定的 $(\tilde{x}^{(l)}, \tilde{y}^{(l)}), 1 \leq l \leq N$ 關係中，則需要 $\tilde{R}$ 範數的最小偏差。

這種方法對於預測、迴歸和插值多項式的建構問題來說是典型的[2]。如果最小偏差的條件意味著 $\tilde{y}^{(l)} = R(\tilde{x}^{(l)}) = \tilde{R}(\tilde{x}^{(l)}, w_1, w_2, \dots, w_k), 1 \leq l \leq N$ ，的值一致，那麼這種構造近似的方法稱為內插（在這種情況下，最佳點處的目標函數 $F(w_1, w_2, \dots, w_k)$ 取最小值等於 0）。請注意，函數 F 的確定可以根據專家對一組 Y 值的度量（距離）選擇的評估以不同的方式進行，這些 Y 值確定要解決的問題的目標。在自然科學、經濟學中規劃實驗時，在建構模式辨識演算法時，在解決物體分類反問題時，在設計控制系統、導引、瞄準、導航等時，都會出現這種問題。 [2]。

基於多變量多項式構建 ANN 的問題本質上與通過相應泰勒級數的部分和來逼近足夠平滑的映射有關，這需要計算訓練集上未知函數的導數的估計。這導致需要求解具有近似給定右側的第一類卷積類型的 Volterra (Fredholm) 積分方程式。事實上，這種方法是流行名稱「卷積人工神經網路」的基礎。請注意，可求函數空間上的第一類 Fredholm 方程式是一個典型的病態問題，具有明顯的計算不穩定性，可以使用 A.N. Tikhonov [17] 的正則化方法來克服。

穩定性和收斂性的基本問題。 讓我們注意與所考慮的範例相關的更一般的人工神經網路的重要概念。

1. ANN 記憶體容量，由訓練集大小決定

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

2. 人工神經網路在逼近給定類別近似關係時的準確性 R . 設定目標的可實現性。

3. 基於給定演算法 F 的 ANN 學習速度。

4. 人工神經網路在重新配置其結構方面的「靈活性」（即動態改變人工神經元 Θ_i 的功能集、選擇權重係數 w_i 的規則、改變訓練集等）。

5. 人工神經網路的效能。

6. 當輸入資料變化時，所得輸出資料的穩定性。

7. 人工神經元計算程序的可行性及其可靠性。

8. 人工神經網路的精確度和速度對其尺寸的依賴性。

9. 在運作期間自動減少 ANN 尺寸以優化其特性的可能性。

10. ANN 的容錯性、功能冗餘性和抗雜訊性。

11.透過基於合併它們的圖和相互資料交換來擴展結構來與其他人工神經網路相容。

12. ANN 結構的拓撲不變量和不同 ANN 實現的同倫等價。

請注意，第 11 條的內容指的是由各種性質的 ANN 組成的複合體，可以將其描述為人工智慧 (AI) 系統的近似。因此，在未來，透過人工智慧技術，我們可以根據人類智力在該集合上先驗引入的順序關係來理解互連人工神經網路的有序層次結構。顯然，透過這種方法，人工智慧是自然（人類）智慧工作的衍生物。

根據哥德爾定理，在任何一致的形式系統中，用該系統的語言，都有可能連同其否定一起表述出不可推導的命題，即形式系統是不完備的。為了解決不確定性，有必要將該陳述或其否定添加到所考慮的正式系統的公理中。對於圖靈機來說，這意味著它外部存在一個更智慧開發的系統，該系統每次都需要選擇將哪個新公理添加到正式系統中，並根據這個選擇來修正圖靈機的演算法。這樣的智慧系統可以形成一個日益複雜和全面的系統層次結構，而人們處於頂層。對這種層次結構的需求源自於萊文海姆-斯科勒姆定理，該定理指出，任何公理系統都允許比創建時假設的更多顯著不同的解釋。公理沒有對解釋或模型設定限制，因此校正者一人一的存在是必要的。萊文海姆-斯科勒姆定理適用的公理系統旨在指定一種非常具體的解釋，並且適用於完全不同的模型，因此它們不符合其目的。綜上所述，人工神經網路和深度學習方法的突破為反映人類外圍功能的問題提供了解決方案。人工神經網路尚未重現思考、計畫行動（「行為」控制）、邏輯「推理」和抽象思考的過程。這些過程將以某種方式透過將各種人工智慧方法（不僅是神經網路方法）與模擬建模相結合來重現[18]。

上述用於建構 ANN 和相應分層人工智慧系統的插值思想的自然發展是在邏輯變數空間中進行插值，以及在與哥德爾和 Löwenheim-Skolem 定理相關的邏輯推理有序結構上應用最佳化程序。

大數據以及基於它們訓練人工神經網路的有效性。至於訓練集 $\tilde{R}_N$ 的數量無限增加時 ANN 的準確性出現的問題，K.D.T.Runge [3] 和 S.N.Bernstein [4] 的例子針對插值多項式的數量無限增加插值節點在這裡是指示性的。

Runge、Bernstein 的例子和 Faber 的一般結果表明了一種深刻的誤解，即 ANN 中神經元的無限增長和構成訓練集 $\tilde{R}_N$ 基礎的大量數據 $N \gg 1$ （大數據）可以確保其在人工智能框架內的有效性。這同樣適用於許多關於透過不斷增長的 $\tilde{R}_N = \{(\tilde{x}^{(1)}, \tilde{y}^{(1)}), (\tilde{x}^{(2)}, \tilde{y}^{(2)}), \dots, (\tilde{x}^{(N)}, \tilde{y}^{(N)})\}$ 訓練資料集來建構所謂「數位雙胞胎」的說法。

這些評論涉及在增加訓練集數量時分析人工神經網路的計算穩定性的一般基本問題，以及因此建立在交互人工神經網路複合體上的人工智慧系統的可行性和安全適用性的限制。

我們強調經典的 Weierstrass 定理 [5] 關於緊集上連續函數空間中多項式集合的密度，允許人們斷言基於多項式構造的收斂 ANN 的存在，但 Faber 定理也是如此作為 K.D.T 的例子。Runge 和 S.N. Bernstein 指出，使用基於滿足條件 (3) 的多項式和訓練集 $\tilde{R}_N$ 和 $N \gg 1$ 建構的 ANN 時存在重大錯誤的風險。

值得注意的是，單純緊複形上多維連續映射的單純近似具有明確定義的 ANN 結構（我們稱之為單純 ANN），如圖 1 所示。Compacta 上單純複形[7]的重心細分過程與它們上映射的一致連續性相結合，使我們能夠聲明，由此構建的單純 ANN RF 均勻收斂於 Compacta 上的 R，其直徑為 $\delta_N > 0$ 個重心細分複數趨於零：

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

我們強調，對於任意連續映射 R，複數 K 上的單純 ANN 的收斂速度由連續性模 $\omega_R(\delta) = \sup_{\|u-v\| \leq \delta} \|R(u) - R(v)\|_{C(K)}, \delta > 0$ 決定。很容易驗證單純複數 K 上的近似 $\|R_F - R\|_C$ 的誤差，它是一個緊複數 $\mathbb{R}_n$ 中的值從上方以值 $\omega_R(\delta)$ 為界，其中 $\delta > 0$ 為複數 $K = D$ 中單純形直徑的精確上限，其上值為中的值的映射 R。事實上，對於單純複形 K 中包含的每個單純形 T，以下公式成立：

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

其中 μ_i 是由頂點 $a_i, 0 \leq i \leq n$ 跨越的單純形 T 中點 x 的重心座標，n 是單純形 T 的尺寸。因此，由於單純形 T 的直徑不超過 $\delta > 0$ ，則

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

因此，我們獲得了單純 ANN $N \leq (n+1) \text{vol}(K) \delta^{-1}$ 的訓練集 $\tilde{R}_N$ 的體積 N 的上限估計，其中 $\text{vol}(K)$ 是單純複形 K 的體積估計。請注意，結合 ANN 的給定精度 ε ，透過連續模表示，我們得到條件

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

特別地，對於平滑映射 R，複數 K 上的近似 RF 的誤差與 N-1 的值成比例。然而，在大多數實際問題中，沒有關於連續性模 ω_R 的合理先驗信息，這導

致在直觀地選擇訓練集 $\tilde{R}_N$ 的體積 N 時可能出現粗近似誤差。顯然，由於這些原因，存在對於訓練集 $\tilde{R}_N$ 的給定體積 N ，沒有普遍準確的單純 ANN。根據有關連續性模 $\omega_R(\delta)$ 的先驗信息，在連續映射類上提出單純 ANN 的準確性問題是合理的。

還要注意的是，由於 $\omega_R(\delta)$ 的單調遞增， $x \in K$ 的自變數空間的維數 n 對公式 (6) 中的逼近誤差的大小有顯著的影響，這對於大維 $n \gg 1$ 的輸入資料很重要，特別是，這種情況在處理視訊資訊和模式識別時很典型。

應該強調的是，所考慮的 ANN 的有效性是由於函數 $R(x)$ 的「預測」值是由直徑相當小的單純形 $T \in K$ 內的點 x 的「包圍」來確保的 δ ，在其頂點指定來自 R_N 的訓練值，由於線性內插，這些訓練值被分佈到點 x 。同時，它的重心座標 $\{\mu_i\}$ 正是圖 1 對應的單純人工神經網路中連接的權重係數。當然，根據人工神經網路的構造，由於給定的設計特點，它根本不能用於公式化從準確性的角度來看合理的預測問題，超出了複雜 K 的邊界（外推法）。為此，有必要考慮更廣泛的 $K' \supset K$ 複合體的訓練集，這無疑需要由人類進行專家評估。

P.L. Chebyshev 在考慮偏離零最少的多項式問題時，透過沿著 Chebyshev 多項式的根定位插值節點來解決最小化給定類函數的插值誤差的問題 [8]。顯然，這是優化選擇用於建立人工神經網路的訓練集的廣泛研究的起點。

ANN 實際應用中最重要的方面是訓練集大小 N 的顯著減少。在此路徑上，有效的解與使用基於 ZBC (P.L. 切比雪夫大數定律) 的近似值的順序平均過程相關，這顯著加速了三角傅立葉級數理論中部分和的收斂 [9]。類似的效果與 Mazur 定理 [1] 相關，該定理允許透過平均過程將弱收斂序列轉換為強收斂序列。類似的現象出現在用於加速粗網格上差分格式收斂的 Richardson 方法中 [11]。

注意，所謂的多層 ANN 是幾個簡單 ANN 的疊加，其中還可以實現反饋連接等。它們的分解、準確性和效率問題目前在很大程度上是工程計算實驗的領域。

未來研究人工神經網路數學本質的一個重要點是尋找相關的拓撲不變量（同倫群）以及尋找實現人工神經網路的各種數學模型的結構等價的條件。

與大多數關於基於助記圖的 ANN 開發的出版物一樣，所提出的構建上述簡單 ANN 的設計（圖 1）在思想上與 A.N. 的作品相關。科爾莫哥洛夫和 V.I. Arnold [13] 關於連續函數的表示。1957 年，他們證明了一個定理（與 D. 希爾伯特的第 13 個問題直接相關）：任何多個變數的連續函數都可以表示為加法運算和一個變數的有限函數集的疊加。

從詳細研究和理論分析的角度來看，基於資料插值的人工神經網路是特殊且方便的，因為最小的最佳化功能導致精確可解決的問題，可以在這些問題上

研究人工神經網路的構造和操作的根本重要問題。

實際上，輸入資料空間的維度明顯大於輸出資料的維度。ANN 的典型用途是「壓縮」大量輸入資料的處理流，以用於分類、物件辨識、視訊和聲音資訊等。

壓縮大量資料的過程可以被視為希爾伯特空間 H 中緊緻連續算子 R 的作用，即 $R:H \rightarrow H$ 。因此，在算子 R 的作用下對輸入資訊進行「壓縮」的過程可以解釋為無限維希爾伯特空間 H 中有限的大數據輸入數據集被 ANN 變換為預輸出數據 $y \in H$ 的緊湊子集（幾乎有限集，即非常近似的有限 ε 網路）。

模式識別問題及其分類問題可以表述為尋找 $x \in H$ 逆問題的解

$$R(x)=y \quad (4)$$

基於所獲得的數據 $y \in H$ ，位於預緊集中。此類問題的本質是貝爾定理 [9]，即不可能將完整的度量空間表示為無處非稠密子集的可數並集。在希爾伯特空間 H 上使用緊緻連續算子 R 尋找反問題 (4) 的解的核心問題是缺少方程式 (4) 的解（即，並非對於所有值都有解）方程式(4)，或影像

$R(H) \subset H$ 上的逆算子 R^{-1} 缺乏連續性- 問題(4) 的解的基本計算不穩定性。

定理 1, [10]。令 $R:H_1 \rightarrow H_2$ 為希爾伯特空間 H_1, H_2 上的連續緊映射。令逆映射 $R^{-1}:R(H_1) \rightarrow H_1$ 定義在 $R(H_1) \subset H_2$ 的圖像上。則等式 (4) 無解的點集 $y \in H_2$ 在 H_2 中處處稠密，並且算子 $R^{-1}:R(H_1) \rightarrow H_1$ 在由在圖像 $R(H_1)$ 上的空間 H_2 引起的度量中的任何開球中都不連續。

注意，對於滿足該定理條件的線性算子 R ，逆算子 R^{-1} 在圖

$R(H_1) \subset H_2$ [11] 的每個點缺乏連續性。由於這種現象在整個密集的輸出資料集 $y \in R(H_1)$ 中都是成立的，因此基於使用緊湊算子 R 解決問題 (4) 來分類和識別輸入資料 x 的基於 ANN 的大資料流處理始終是一個不適定問題並獲得穩定的近似解 x 另外需要使用正規化方法[17]。特別地，這種現象對於基於訊號頻率濾波並抑制輸入訊號 $x \in H_1 = L_2[a, b]$ 的高頻成分的硬體數位資料處理是典型的。

因此，基於使用 ANN 作為逼近映射和使用緊湊算子構建反問題解決方案的方法的問題類別表現出共同的數學特性 - 處理大量數據時的計算不穩定。這些問題也類似於從「數位學生」中「恢復」影像的問題，因為它們具有相同的數學性質。

ANN 模型和不動點問題。 ANN 模型的建構可以被視為問題 (2)，即為資料集 $M \supset X \cup Y$ 上的某些映射 $f:M \rightarrow M$ 找到不動點 $\bar{x}(\bar{x}=f(\bar{x}))$ 。映射 f 的選擇將稱為 ANN 層。本質上，ANN 層的建構是多層 ANN 建構的一個階段，由 $f_n:M \rightarrow M$ 層的順序應用變換組成，使用者依序修改這些變換以實現一些「對使用者而言最佳的、理想的」結果。期望的結果是某個極

限變換 $f = \lim_{n \rightarrow \infty} f_n$ 的不動點。在許多情況下，尋找 $\bar{x}$ 的過程是基於考慮迭代序列

$$x_{n+1} = f_n(x_n) \quad (5)$$

通常，在多層人工神經網路的具體實現中，選擇的層是相同的，即 迭代 $f_n \equiv f$ 。一般來說，這種創建多層 ANN 的方法可能不會收斂。A.N. Sharkovsky 發現了一維迭代的動態混沌 $x_{n+1} = f_n(x_n)$ [19]。

因此，人工神經網路層鏈的建構必須進行修改，以確保迭代的收斂。下面討論具有這些屬性的演算法。

定理 2, [22]。設 (M, ρ) 為等壓算符序列作用於其上的完備度量空間

$$f_n: M \rightarrow M, n \in \mathbb{N} \quad (6)$$

即， f_n 的壓縮常數不超過值 $0 \leq q < 1$ 。讓一連串固定點 $\bar{x}_n, n \in \mathbb{N}$ ，映射 f_n 收斂到某點 $\bar{x} \in M$ 。然後可以找到點 $\bar{x}$ 作為迭代的極限

$$x_{n+1} = f_{n+1}(x_n), n = 0, 1, \dots, \quad (7)$$

其中起點 $x_0 \in M$ 是任意選擇的。

評論。點 $\bar{x}_n$ 收斂的充分條件是映射 f_n 到 M 的逐點收斂。

上述定理與具有唯一解的系統的構造有關。分類和模式識別的實際問題與建立傳入資料與許多物件之一的對應關係相關，每個物件都與多層人工神經網路的一個固定點相關聯。也就是說，對於此類映射，不動點不是唯一的，因此創建穩定的多層 ANN 最重要的方面之一是構建穩定迭代極限行為的演算法(7)，同時放棄與以下相關的定理 2 的要求均勻壓縮的性質。特別是，序列(5)、(6)的混沌行為是由於存在映射(6)相當有彈性的子集。一類相當廣泛的映射具有相似的屬性，其中自然包括透過實驗創建的人工神經網路。這種行為對於雙曲動態系統來說是典型的[21]。

與這些現象相關的是映射的傅立葉級數展開的高頻諧波行為的不穩定性的一般特性，即大量基底函數對應於大拉伸係數，這會產生迭代的混沌行為(5)。

定理 3, [22]。讓 f 到 $\mathbb{R}_n$ 的連續映射使有界閉球 M 不變。讓迭代序列(7)由下列公式定義

$$x_n = f\left(\frac{1}{n} \sum_{i=0}^{n-1} x_i\right) \quad (8)$$

假設幾乎處處都存在算術平均值 $\bar{x} = \frac{1}{n} \sum_{i=0}^{n-1} x_i$ 的極限，則對於幾乎所有初始點 $x_0 \in M$ ，迭代序列(8)都收斂到映射 $f: M \rightarrow M$ 的不動點 $\bar{x}$ 。

因此，演算法（8）可以用作創建穩定收斂 ANN 類別的基礎。

算術平均值收斂的基本問題 $\frac{1}{n} \sum_{i=0}^{n-1} x_i$ 是開放的，它已經在保測映射

$f: M \rightarrow M$ 的 Birkhoff-Khinchin 遍歷定理 [23] 的框架內得到了有意義的解決。作者的計算實驗證明了演算法（8）對於球到自身的一系列連續映射的可靠性和計算穩定性。

有序空間上映射的不動點（迭代的單調收斂）。建構和證明構成人工智慧結構的 ANN 層次結構的演算法 (2) 收斂性的一個有希望的方向是使用與 A.N. Tikhonov, Yu 的作品相關的局部凸和有序空間上的不動點定理。紹德和 A. 塔斯基[1]。這些定理允許在基於保序連續映射的 ANN 的許多實際問題中構造收斂到最小（最大）不動點的構造型演算法。特別是，這個陳述在程式語言的形式語意學中有著深入的應用（在抽象解釋中—電腦語言語意的近似理論，與 P. 和 R. Couso 的名字相關[25]）。

解決連續介質力學問題的浸入法和人工神經網路的應用。需要強調的是，從某種意義上來說，人工神經網路技術最發達、最普遍的理論是線性問題的計算方法理論。因此，從這個角度來看，透過改變描述的初始維度和擴展函數空間來將非線性問題「沉浸」到一般類別的線性問題中的過程是很重要的 [26]。

上述非線性系統在線性動力學中的浸入可以在基於 Sobolev 廣義解的楊嵌入到配備吉洪諾夫拓撲的線性泛函空間的擬線性系統泛函解概念的框架內普遍實現 [27, 28]。

需要注意的是，弱穩定性條件[27, 28]本質上比非線性問題差分格式理論中的穩定性概念簡單得多；它反映了基本守恆定律的「正確」再現。因此，所提出的將非線性守恆定律嵌入到配備 A.N. Tikhonov 拓撲的函數解空間上的線性問題中的方法具有通用計算性質，在與圖 1 中的助記圖相對應的局部凸 ANN 框架內實現。1. 這是因為函數解的定義自然地擴展到包含柯西問題積分形式的給定初始資料的函數解的最小凸包。

Литература

References

參考書目

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, Bd. 46, S. 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. Берёзин 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. Бетелин В.Б., Галкин В.А. Математические задачи, связанные с искусственным интеллектом и искусственными нейронными сетями, Успехи кибернетики, 2021, №4, С. 6-14.
11. Николаев Е.С., Самарский А.А. Выбор итерационных параметров в методе Рундсона, ЖВМ и МФ, Т. 12, 1972, с. 960 - 973.
12. 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, Т. 495, с. 13–16.
13. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения, ДАН СССР, 1957, Т.114, №5, с. 953 – 956.
14. Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem, In Proceeding of the International Conference on Neural Networks, (In Eng.) (III, 11-14). 1987, NewYork: IEEE Press.
15. Funahashi K. On the Approximate Realization of Continuous Mappings by Neural Networks, Neural Networks Vol. 2, Issue 3, 1989, P. 183-192.
16. Cybenko G. Approximation by superposition of a sigmoidal function // Math. Control, Signals and Systems. 1989, V. 2, P. 303 - 314.
17. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач, М., Наука, 1986.
18. Бетелин В.Б., Галкин В.А., Дубовик А.О. Некоторые математические аспекты проблемы построения искусственных нейронных сетей, Успехи кибернетики. 2021;2(3): С. 19-22.
19. Крянев А.В. и др. Метрический анализ и обработка данных.
20. Ly T.Y., Yorke J.A. Period three implies chaos// Amer. Math. Monthly, 1975, V. 82, P. 982–985.
21. Аносов Д. В. Геодезические потоки на компактных римановых многообразиях, Труды Математического института им. В.А. Стеклова, вып. 90, № 1, 1967, с. 1– 235.
22. Александров П.С., Пасынков Б.А. Введение в теорию размерности. М., Наука, 1973.

23. Каток А.Б., Хасселблат Б. Введение в теорию динамических систем с обзором последних достижений, Пер. с англ. Под ред. А.С.Городецкого. М., МЦНМО, 2005. ISBN 5-94057-063-1.

24. Бетелин В. Б., Галкин В. А. О неподвижных точках непрерывных преобразований, связанных с построением искусственных нейронных сетей, Доклады Российской академии наук. Математика, информатика, процессы управления, 2022, т. 507, С. 22–25.

25. Кузо П., Кузо Р. Абстрактная интерпретация: унифицированная решетчатая модель для статического анализа путём построения или аппроксимации неподвижных точек, Протокол конференции 4-го симпозиума АСМ по принципам языков программирования. PDF. Лос-Анджелес, Калифорния, США. 1977. ACM Press. P. 238 – 252.

26. Бетелин В. Б., Галкин В. А. Математические и вычислительные проблемы, связанные с образованием структур в сложных системах, Компьютерные исследования и моделирование, 2022, т. 14, вып. 4, С. 805 - 815.

27. Галкин В. А. Теория функциональных решений квазилинейных систем законов сохранения, Труды семинара имени И.Г.Петровского, 2000, вып. 20, С. 81–120.

28. Galkin V.A. Background of mathematical models, based on conservation laws systems, Industrial Mathematics, Narosa Publishing House, 2006, New Delhi, India. P. 159–178.