

Теория высших категорий как метаязык сильного искусственного интеллекта

М.Г. Щербаков

Статья посвящена архитектуре нейронной сети сильного искусственного интеллекта. В статье обосновывается тезис о том, что математическая модель сильного искусственного интеллекта может быть описана с помощью математической теории высших категорий. Суть такого подхода состоит в описании не объектов, а их морфизмов, функторов и естественных преобразований. Автор отмечает, что в основе построения нейронной сети сильного искусственного интеллекта лежит две аксиомы теории высших категорий: тождественности и композиции. Автор приводит примеры использования методологии теории высших категорий для описания биологической нейронной сети головного мозга человека, в которой веса нейронов представляют собой не количественные значения, а связи и их преобразования.

В статье автор приводит аналогию между биологической нейронной сетью головного мозга человека и нейронной сетью сильного искусственного интеллекта, способного решать творческие задачи как человек, проводить ассоциации, учиться на ограниченных данных.

В проведенном исследовании используются системно-методологический подход, гипотетико-дедуктивный подход, методология теории высших категорий.

В заключение автор приходит к выводу о том, что теория высших категорий представляет собой метаязык для описания математической модели нейронной сети сильного искусственного интеллекта.

Ключевые слова: теория высших категория, сильный искусственный интеллект, морфизм, функтор, естественное преобразование, нейронная сеть, математическая модель, метаязык, ассоциации.

This article discusses the architecture of a strong artificial intelligence neural network. It argues that the mathematical model of strong artificial intelligence can be described using the mathematical theory of higher categories. The essence of this approach is to describe not objects, but their morphisms, functors, and natural transformations. The author notes that the construction of a strong artificial intelligence neural network is based on two axioms of the theory of higher categories: identity and composition. The author provides examples of using the methodology of higher category theory to describe the biological neural network of the human brain, in which neuron weights represent not quantitative values, but connections and their transformations.

In the article, the author draws an analogy between the biological neural network of the human brain and the neural network of a strong artificial intelligence capable of solving creative problems like a human, making associations, and learning from limited data.

This study utilizes a systems-methodological approach, a hypothetical-deductive approach, and the methodology of the theory of higher categories.

In conclusion, the author concludes that higher-category theory represents a metalanguage for describing the mathematical model of a strong artificial intelligence neural network.

Keywords: higher-category theory, strong artificial intelligence, morphism, functor, natural transformation, neural network, mathematical model, metalanguage, associations.

В цифровую эпоху обладание технологиями, основанными на сильном искусственном интеллекте, представляют его обладателю существенные конкурентные преимущества на мировом рынке.

В настоящее время активно применяются технологии, связанные с использованием узкого искусственного интеллекта. Между тем ведутся разработки по созданию сильного искусственного интеллекта, способного решать творческие задачи, оперировать неполными данными и быстро обучаться на

ограниченных данных. К сожалению, в настоящее время отсутствует теоретическая основа для создания математической модели нейронной сети сильного искусственного интеллекта, следовательно, практические разработки могут не привести к задуманному результату.

В настоящем исследовании предпринята попытка описать биологическую нейронную сеть головного мозга человека, следовательно, создать теоретическую основу для создания математической модели нейронной сети сильного искусственного интеллекта.

Искусственный интеллект представляет собой технологию, способную имитировать когнитивную деятельность человека (мыслить).

П.М. Морхат отмечает, что искусственный интеллект - полностью или частично автономная самоорганизующая (самоорганизующаяся) компьютерно-аппаратно-программная виртуальная или киберфизическая, в том числе биокибернетическая, система, наделенная/обладающая способностями и возможностями мыслить, самоорганизовываться, обучаться, самостоятельно принимать решения и т.д.[Морхат 2018].

В США закон S.2217, принятый Конгрессом в 2017 г., различает два типа искусственного интеллекта (ИИ): общий (широкий) и специфический (узкий).¹ Общий искусственный интеллект определяется как гипотетическая система будущего с уровнем интеллектуального поведения, сопоставимым с человеческим или превосходящим его, включая когнитивные, эмоциональные и социальные аспекты.

В соответствии с регламентом Европейского Союза об искусственном интеллекте от 12.07.2024 г. № 2024/1689 искусственный интеллект относится к системам, которые демонстрируют разумное поведение, анализируя окружающую среду и предпринимая действия - с определенной степенью автономности - для достижения конкретных целей.²

¹

S. 2217 (115th): FUTURE of Artificial Intelligence Act of 2017 // <https://www.govtrack.us/congress/bills/115/s2217/text>

² Regulation (EU) 2024/1689 // https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

Сильный искусственный интеллект (AGI, Strong AI) - это результат совершенствования узкого искусственного интеллекта [Серл 1990].

Artificial General Intelligence представляет собой автономную самоорганизующаяся систему, способную выполнять творческие задачи как человек [Frankish 2014].

Основу нейронной сети искусственного интеллекта составляет искусственный нейрон, который в математическом аспекте можно представить следующей схемой (рис. 1).

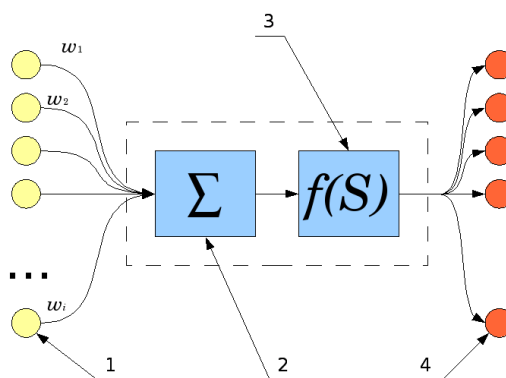


Рис. 1 Схема работы искусственного нейрона

В основе принцип работы нейронной сети искусственного интеллекта лежит обработка информации на основе обучения, а именно подбор синоптических весов в нейронных сетях, которые имеют числовые значения.

Веса в нейронных сетях - это числовые коэффициенты, определяющие силу связи между нейронами и степень влияния входных данных на результат. Они действуют как «важность» признаков, умножаясь на входные сигналы. В процессе обучения веса автоматически корректируются, чтобы минимизировать ошибку модели. Это «знания» нейросети, хранящиеся в виде числовых значений.

Нейронная сеть искусственного интеллекта представляет собой гиперсеть (рис. 2).

Важно отметить, что искусственная нейронная сеть представляет собой крайне упрощенную математическую абстракцию одного из аспектов работы

биологической нейронной сети, взятая из контекста невероятно сложной биологической системы.



Рис. 2 Гиперсеть нейронов искусственного интеллекта

Важно отметить, что биологическая нейронная сеть головного мозга человека использует не количественные значения синаптических весов, а их связи и преобразования на разных уровнях организации.

Например, биологическом, нейронном и когнитивном уровнях организации головного мозга человека (рис. 3).

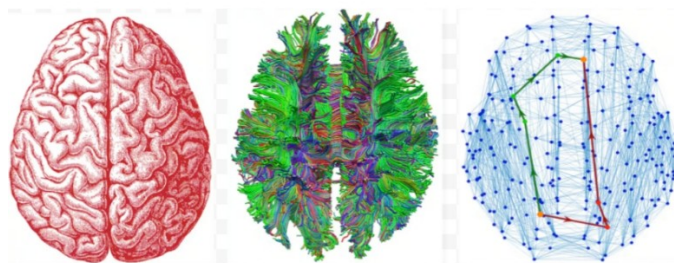


Рис. 3 Уровни организации головного мозга человека

Другими словами, обработка информации осуществляется на разных уровнях организации мозга и принципиально другим способом, основанным на подборе связей между нейронами и их преобразованиями.

Кроме того, связи и преобразования в биологической нейронной сети осуществляются на разных уровнях организации.

Конечно, такой способ обработки информации требует огромную архитектурную сложность биологической нейронной сети. Так, головной мозг человека представляет собой нейронную сеть, включающую в себя около 85

миллиардов нейронов. Количество связей между нейронами составляет 10^{15} [Дорогина 2019].

Когнитивистика рассматривает человеческий мозг как систему связей - коннектом (рис. 4) [McDowell 1996].

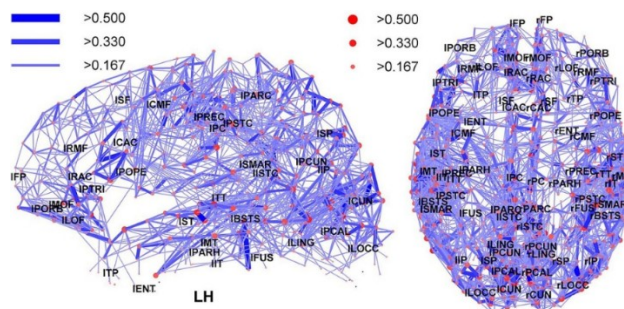


Рис.4 Коннектом головного мозга человека

Нейронная сеть головного мозга человека обладает динамикой и функционирует на разных уровнях организации, от отдельного нейрона, до всего мозга [Хайкин 2006].

Кроме того, нейронная сеть головного мозга человека представляет собой самоорганизующуюся систему, которая способна в процессе развития порождать новые уровни организации. Причем каждый такой новый уровень оказывает обратное воздействие на ранее сложившиеся, перестраивает их, в результате чего система обретает новую целостность [Степин 2016].

К сожалению, в настоящее время отсутствуют вычислительные мощности, способные описать полную архитектуру связей в головном мозгу человека.

Между тем, по нашему мнению, можно описать принципы построения математической модели биологической нейронной сети мозга человека, следовательно, нейронной сети сильного искусственного интеллекта.

Никола Бурбаки отмечает, что математика - это наука об отношениях между объектами, о которых ничего не известно, кроме описывающих их некоторых

свойств, - именно тех, которые в качестве аксиом положены в основание той или иной математической теории [Колмогоров 1929].

По нашему мнению, математическую модель биологической нейронной сети мозга человека и сильного искусственного интеллекта можно описать с помощью методологии теории высших категорий.

Д. Спивак рассматривает теорию категорий как универсальный язык для описания структур и отношений в различных системах [Spivak 2018].

Теория категорий - это математическая теория, основная идея которой заключается в том, что при изучении какого-либо класса математических объектов следует рассматривать этот класс как новый единый конгломерат, в котором учитываются все связи между этими объектами, но не учитывается их внутренняя структура [Landry 2009].

Другими словами, теория категорий, формируя новые категории, переносит акцент с элементов на их связи, сохраняя структуру (рис. 5).

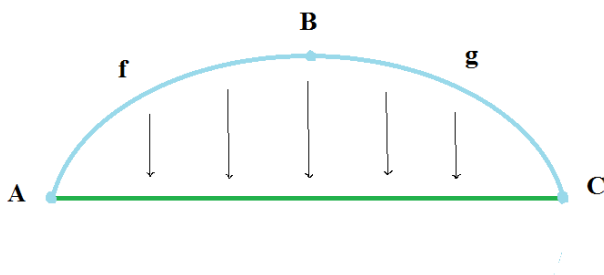


Рис. 5 Перенос акцента от объекта к связям

Категория - это не просто коллекция математических объектов, а объекты вместе с морфизмами (стрелками) между ними. Например, *set*: категория множеств и функций; *grp*: категория групп и гомоморфизмов; *top*: категория топологических пространств и непрерывных отображений; *vect*: категория векторных пространств и линейных отображений.

Стрелки (морфизмами) обозначают преобразование объектов, при котором сохраняется их структура. Например, категория (K) имеет объекты (AB) и морфизм (f) :

$$f: A \rightarrow B \quad \text{или} \quad A \xrightarrow{f} B$$

Важно отметить, что любые категории должны соответствовать следующим аксиомам: тождественности ($\text{id}_a : A \rightarrow A$) и композиции ($q = g \circ f$) (рис. 6).

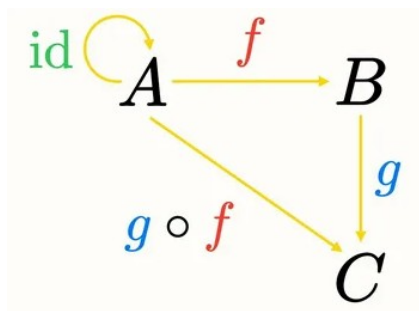


Рис. 6 Диаграмма категории

Так, для каждого объекта A задан тождественный морфизм $\text{id}_a \in \text{Hom}(A, A)$

Другими словами, кодирование объекта в нейронной сети должно быть тождественно самому себе. Например, на различные изображения Билла Клинтона (рис. 7) реагируют одинаковая совокупность нейронов (аксиома тождественности) [Анохин 2016].

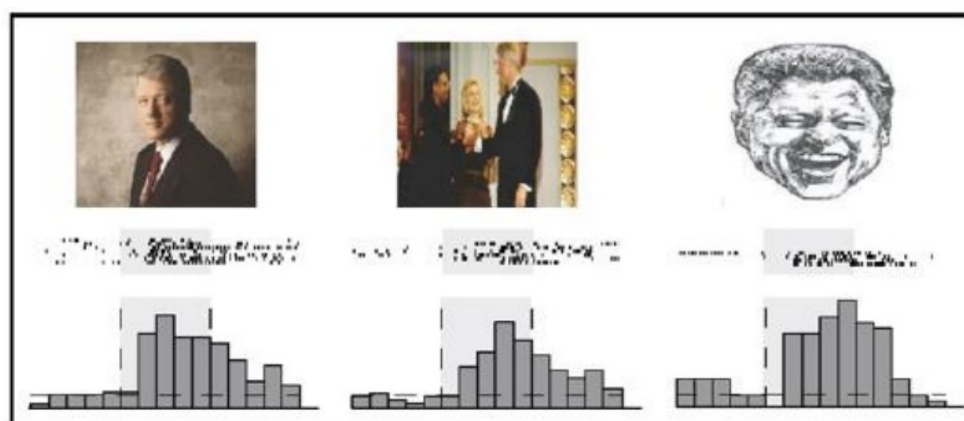


Рис. 7 Схема срабатывания нейрона

Теории категорий позволяет создавать ассоциативную композицию морфизмов. Например, на диаграмме (рис. 8) представлены объекты (X, Y, Z) и

морфизмы: $f: X \rightarrow Y$ и $g: Y \rightarrow Z$, а также ассоциативная композиция морфизмов: $h = f \circ g: X \rightarrow Z$.

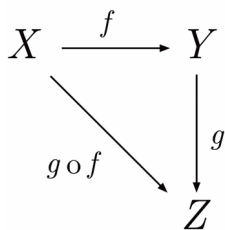


Рис. 8 Диаграмма композиции морфизмов

Ассоциативная композиция морфизмов лежит в основе функционирования ассоциативных нейронных сетей (рис. 9).

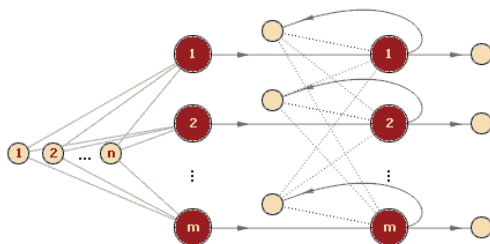


Рис. 9 Ассоциативная нейронная сеть Хемминга

По нашему мнению, в основе математической модели нейронной сети сильного искусственного интеллекта должны лежать две аксиомы теории высших категорий: тождественности и композиции.

Важно отметить, что принципиально иной механизм обработки информации, основанный на использовании не весов, а связей и их преобразований, может быть описан теорией высших категорий.

Теория высших категорий расширяет традиционное понимание категорий, оперируя не только объектами и отношениями между ними (морфизмами), но и отношениями между отношениями (функтор и естественное преобразование).

В аспекте теории высших категорий синоптический вес нейрона сильного искусственного интеллекта может быть представлен не в виде значения, а в виде структуры морфизмов, функторов и естественных преобразований. Например, на рис. 10 изображена схема отображения морфизмов и функторов, при которой сохраняется структура объектов в категории.

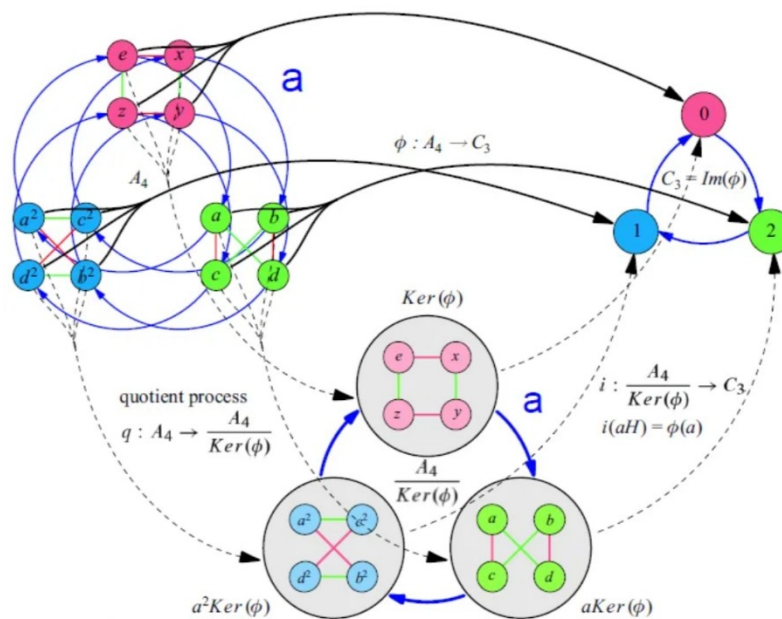


Рис. 10 Схема морфизмов и функторов

Важно отметить, что существенную роль в процессе преобразования играют функторы и естественные преобразования.

Функтор - это отображение между категориями, сохраняющее структуру (рис. 11).

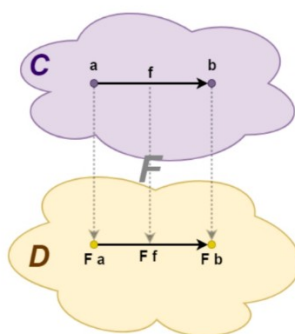


Рис. 11 Схема функтора

Например, нейроны, активирующиеся вместе, соединяются между собой (рис. 12).

Другими словами, изменяется характер их взаимосвязи (функтор).



Рис. 12 Соединение нейронов

Например, нейроны, активирующиеся вместе, образуют устойчивые связи, которые в последствии могут порождать ассоциации, следовательно, представлять механизм ассоциативного мышления.

Например, нейронные связи могут вызывать ассоциации у других нейронных связях, следовательно, представляют собой способность человека связывать идеи, понятия или образы на основе их схожести или связи, даже если они изначально не имеют прямого отношения друг к другу (рис. 13) [Хайкин 2006].

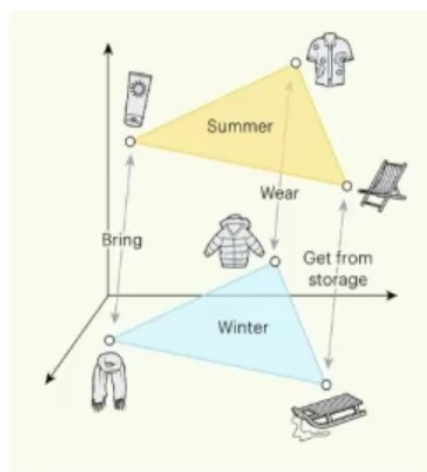


Рис. 13 Ассоциации между образами

Естественные преобразования - это отображения между функторами. Они позволяют говорить, когда две разные конструкции (функторы) «естественным образом» связаны друг с другом. Например, естественное преобразование между

функторами $F: C \rightarrow D$ и $G: C \rightarrow D$ это семейство морфизмов $\eta_A: F(A) \rightarrow G(A)$ для всех объектов A в категории C такова, что для всех морфизмов [Букур 1972].

$f: A \rightarrow B$ в категории C перемещает следующая диаграмма:

$$\begin{array}{ccc} F(A) & \longrightarrow & G(A) \\ \downarrow & & \downarrow \\ F(B) & \longrightarrow & G(B) \end{array}$$

В теории высших категорий объект категории полностью определяется его отношениями с другими объектами. Например, связи определяют свойства активации нейронов на определенный объект (рис. 14) [Анохин 2021].

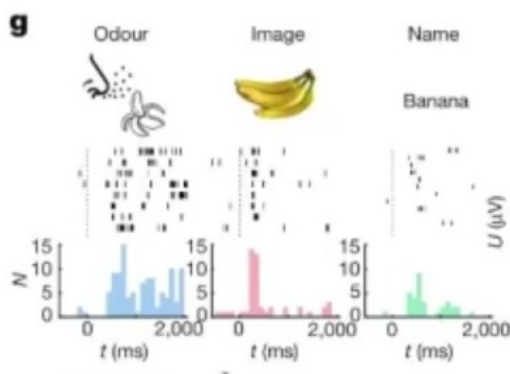


Рис. 14 Кодирование объекта в нейронной сети

Теория категорий позволяет проводить операции (инъекции и сюръекции) с морфизмами, в результате которой появляются мономорфизмы и эпиморфизмы.

Мономорфизм представляет собой категориальное обобщение инъективных функций:

$$f: X \rightarrow Y \text{ из } f \circ g_1 = f \circ g_2 = g_1 = g_2$$

Эпиморфизм является дуальным понятием по отношению к мономорфизму:

$$f: X \rightarrow Y, \text{ если для любых } h_1, h_2: Y \rightarrow Z \text{ из } h_1 \circ f = h_2 \circ f = h_1 = h_2$$

Инъекции и сюръекции морфизмов позволяют описать языком математики образование новых связей между нейронами, при которых сохраняется структура.

Например, в теории гиперсетей, предложенной К.В. Анохиным, коги - это когнитивные группы из рассеянных по мозгу, но связанных совместной деятельностью нервных клеток, а линкеры (лиги) - это связанные нейроны, одновременно входящую и в одну, и в другую когнитивную группу (рис. 15) [Анохин 2021].

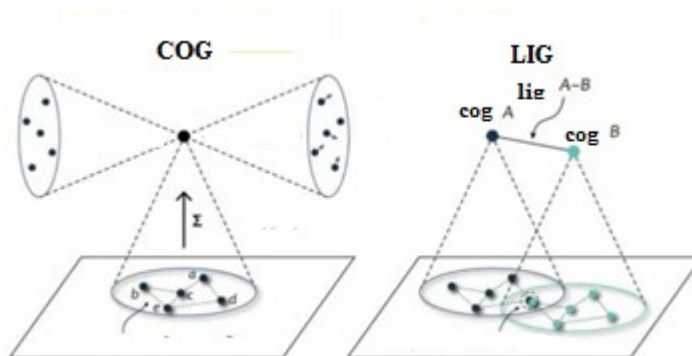


Рис. 15 Коги и лиги в нейронной сети мозга человека

Важно отметить, что морфизмы, функторы и естественные преобразования существуют на разных уровнях организации биологической нейронной сети.

Более того, разные уровни организации биологической нейронной сети непрерывно отображаются друг в друге.

Так, гомотопическая теория типов отмечает, что путь $(\gamma_3 \circ \gamma_2) \circ \gamma_1$ не равен, а гомотопен пути $\gamma_3 \circ (\gamma_2 \circ \gamma_1)$ [Aczel 2002].

Например, когнитом по отношению к коннектому представляет собой новый уровень организации нейронной сети (нейронную гиперсеть) (рис. 16) [Анохин 2021].

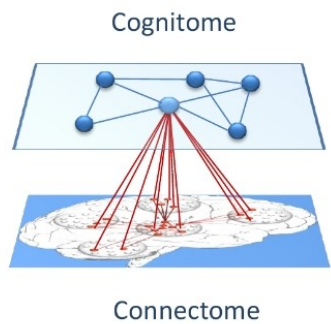


Рис. 16 Уровни организации нейронной сети в головном мозге

Важно отметить, что теория высших категорий имеет дело с непрерывным отображением из топологического пространства в подпространство, которое сохраняет положение всех точек в этом подпространстве (ретракция) (рис. 17).

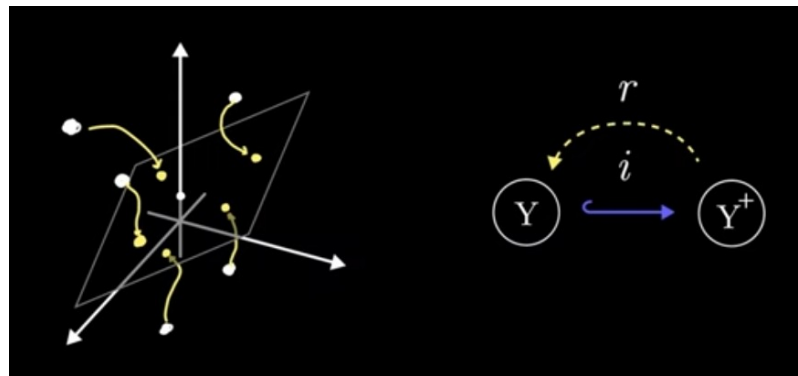


Рис. 17 Ретракция топологического пространства

Ретракция представляет собой левую инверсию некоторого морфизма: Если $f : X \rightarrow Y$ и $g : Y \rightarrow X$ - морфизмы, состав которых $f \circ g : Y \rightarrow Y$ является морфизмом идентичности в Y , то g - ретракция f , а f - ретракт g .

Теория высших категорий позволяет создавать математическую модель нейронной сети любой сложности.

Например, топологические пространства (объекты) имеют непрерывные преобразования (морфизмы), между морфизмами имеются непрерывные преобразования (гомотопии), что приводит к ∞ -Категории (бесконечномерной категории) [Родин 2010].

Таким образом, в основе математической модели биологической нейронной сети связи и их преобразования (морфизмы, функторы и естественные преобразования).

По нашему мнению, в основе математической модели нейронной сети сильного искусственного интеллекта лежат связи и их преобразования, которые могут быть описаны теорией высших категорий.

Другими словами, теория высших категорий представляет собой метаязык сильного искусственного интеллекта.

Заключение

В заключение можно сделать вывод о том, что теория высших категорий, основанная на аксиомах тождественности и композиции, описывающая морфизмы, функторы и естественные преобразования, представляет собой универсальный метаязык для описания математической модели сильного искусственного интеллекта.

Литература

Анохин К.В. Коннектом и когнитом: заполнение разрыва между мозгом и разумом // Седьмая международная конференция по когнитивной науке. М., 2016. С. 18-19.

Анохин К.В. Когнитом: в поисках фундаментальной нейронаучной теории сознания // Журнал высшей нервной деятельности им. И.П. Павлова. 2021. Т. 71, № 1. С. 39-71.

Букур И., Деляну А. Введение в теорию категорий и функторов. М.: Мир, 1972. 260 с.

Нейрофизиология: учеб. пособие / О. И. Дорогина; М-во науки и высш. образования Рос. Федерации, Урал. федер. ун-т. - Екатеринбург: Изд-во Урал. ун-та, 2019. - 100 с.

Колмогоров А. Н. Современные споры о природе математики // Научное слово. 1929. Вып. 6. С. 41-54.

Морхат П.М. К вопросу о правосубъектности «электронного лица» // Юридические исследования. 2018. Т. 4. - С. 6.

Родин А. В. Теория категорий и поиски новых математических оснований физики // Вопросы философии. 2010. № 7. С. 67–81. EDN: MSXDKX.

Серл Д. Разум мозга - компьютерная программа? // В мире науки. - 1990. - № 3 [Электронный ресурс]. - URL: <http://www.raai.org/library/books/sirl/ai.htm> (дата обращения: 16.01.2026).

Степин В.С. / Сложность. Разум. Постнеклассика. - 2013 - № 4 - С. 45-59.

Хайкин С. Нейронные сети: полный курс. М.: Вильямс, 2006. С. 1104.

Category Theory (Stanford Encyclopedia of Philosophy). [Электронный ресурс]. URL: <https://plato.stanford.edu/entries/category-theory/> (дата обращения: 16.01.2026).

Fong B., Spivak D. I. Seven sketches in compositionality: An invitation to applied category theory // arXiv. Preprint arXiv:1803.05316. 2018. (дата обращения: 16.01.2026).

Frankish K., Ramsey W.M. The Cambridge handbook of artificial intelligence. Cambridge: Cambridge University Press, 2014.

Homotopy Type Theory: Univalent Foundations of Mathematics. Princeton, NJ: The Univalent Foundations Program: Institute for Advanced Study, 2013. 480 p.

McDowell J.H. Mind and World: with a New Introduction. Cambridge, 1996. 191p.

Об авторе

Щербаков Михаил Геннадьевич, соискатель Института законодательства и сравнительного правоведения, г. Москва, pravovednalog777@mail.ru.