

**МОДЕЛЬ ВЕРСИОНИРОВАНИЯ НЕЙРОСЕТЕВЫХ МЕТОДОВ РЕШЕНИЯ
ЗАДАЧ, ПРЕДСТАВЛЕННЫХ В БАЗЕ ЗНАНИЙ**

Ковалев Михаил Владимирович¹, Зотов Никита Владимирович¹

¹Белорусский государственный университет информатики и радиоэлектроники, кафедра интеллектуальных информационных технологий

<https://doi.org/10.5281/zenodo.7856119>

Аннотация: В статье рассматривается процесс проектирования нейросетевых методов решения задач в базах знаний интеллектуальных систем. Предложена модель версионирования нейросетевых методов решения задач, описанных на специализированном языке представления нейросетевых методов решения задач.

Ключевые слова: интеллектуальная система, решатель задач, искусственная нейронная сеть, граф знаний, база знаний, Открытая семантическая технология для проектирования интеллектуальных систем (*Open Semantic Technology for Intelligent Systems, OSTIS*).

ВВЕДЕНИЕ

Современное развитие всех направлений *Искусственного интеллекта* направлено на построение интеллектуальных систем, автоматизирующие все более сложные виды человеческой деятельности. Текущее состояние в области разработки интеллектуальных систем нового поколения [1] показывает, что такие системы должны обеспечивать:

- унификацию представления и согласованность различных видов знаний и методов решения задач;
- интеграцию и конвергенцию различных методов решения задач в единой базе знаний с целью обеспечения согласованности семантики этого набора методов;
- представлять и интерпретировать как можно большее число классов методов решения задач (программ).

Интеграция различных методов решения задач в единой базе знаний гарантирует согласованность семантики этого набора методов. При решении задач с использованием подобных методов система не общается с внешней средой посредством передачи входных и выходных данных. Вместо этого единая база знаний позволяет системе отслеживать изменение входных знаний в режиме реального времени при помощи широкого спектра методов, что обеспечивает возможность интроспекции и пояснения принятых решений системой. Единая база знаний для методов решения задач и знаний, используемых для их решения, позволяет системе рефлексировать над процессом решения задач, объяснять причины своих решения, находить там ошибки.

Активно развивающимся классом методов решения задач являются искусственные нейронные сети (*и.н.с.*). Это объясняется, с одной стороны, бурным развитием теоретических основ искусственных нейронных сетей, а с другой - увеличением вычислительных возможностей машин, используемых для их обучения.

Были получены впечатляющие результаты в области решения задач с помощью искусственных нейронных сетей [2]. Среди положительных характеристик *и.н.с.* следует выделить их возможность эффективно решать задачи в условиях отсутствия известных

закономерностей, а также способность к решению проблем без обязательной разработки проблемно ориентированных *методов решения задач*.

Однако существуют серьезные проблемы *нейросетевых методов решения задач*:

- Эвристический характер процесса разработки *нейросетевых методов решения задач*. Процесс подбора архитектур *и.н.с.* и параметров их обучения предъявляет высокие требования к уровню знаний инженеров *и.н.с.*

- Отсутствии явного выделения семантических связей между знаниями в процессе решения задач. Они выделяются неявно, статистически, исходя из данных, которые использовались для обучения. Отсутствие явного выделения *смысла* приводит к проблеме “*черного ящика*” [3]. Появилось целое направление *Explainable AI*, в рамках которого исследователи предпринимают попытки к объяснению решений *и.н.с.* [4, 5].

Формализация *и.н.с.* в *базе знаний интеллектуальной системы* вместе с другими *методами решения задач* позволяет нивелировать перечисленные проблемы *и.н.с.*, так как в таких системах задача проектирования *и.н.с.* и задача объяснения решений этой *и.н.с.* представлены в единой для всей базы знаний форме и могут решаться с помощью любого из представленного в этой *базе знаний метода решения задач*.

Зачастую, *и.н.с.* активно изменяется в процессе проектирования и интерпретации (конфигурация связей, количество слоев, весовые коэффициенты синапсов, функции активации и т. д.). Для решения задачи проектирования *и.н.с.* система должна иметь возможность анализировать решения одной и той же задачи на различных версиях одного и того же *нейросетевого метода решения задач* с целью оценки успешности тех или иных решений в проектировании данного метода, к примеру, успешность подбора функций активаций, обучающей выборки, алгоритма обучения, конфигурации связей в слоях и т. п.

Целью данной работы является разработка подхода к *версионированию нейросетевых методов решения задач* в *базе знаний интеллектуальной системы*.

ПРЕДЛАГАЕМЫЙ ПОДХОД

Для решения перечисленных выше проблем предлагается *Технология OSTIS* [1]. Интеллектуальные системы, реализуемые по принципам *Технологии OSTIS*, называются *ostis-системами*.

Подходы к *интеграции и.н.с. с базами знаний* в *ostis-системах* рассматриваются в работе [6]. Подходы к интеграции “по входам и выходам” были апробированы и описаны в работах [7, 8]. Полная *интеграция и.н.с. с базами знаний*, т. е. использование нейросетевых методов решения задач путем их формализации в базе знаний *ostis-системы* описаны в работе [9], в рамках которой описана *Денотационная и Операционная семантика Языка представления нейросетевых методов решения задач (Нейро-SCP)*. Настоящая работа является развитием данного языка.

Язык представления нейросетевых методов решения задач позволяет представлять и интерпретировать в памяти *ostis-системы* нейросетевые методы. Данный язык является подязыком *Языка SCP* [1]. Любой метод, представленный на *Языке SCP* является фрагментом базы знаний, поэтому задача *версионирования нейросетевых методов решения задач* в *базе знаний интеллектуальной системы* сводится к задаче версионирования любых *фрагментов базы знаний*.

Под *фрагментом базы знаний* подразумевается формальная спецификация любого знака сущности или понятия, представленного в *базе знаний* заданной системы. То есть *фрагмент базы знаний* - это ничто иное как некоторая семантическая окрестность или структура, включающая знания об объекте *предметной области базы знаний* этой системы. Процесс *версионирования фрагмента базы знаний* подразумевает полное представление и описание его состояний, а также предоставление возможностей и средств для обработки и анализа состояний этого *фрагмента базы знаний*. Под *состоянием фрагмента базы знаний* подразумевается объединение результатов действий, выполненных над этим *фрагментом базы знаний*, начиная с момента его существования в *базе знаний* (то есть с начального состояния этого *фрагмента базы знаний*). *Версионирование фрагмента баз знаний* требует строгой идентификации всех состояний с момента начала существования этого *фрагмента базы знаний*, то есть требует построения *биективного соответствия* между *состоянием фрагмента базы знаний* и его уникальным идентификатором в рамках всей *базы знаний* системы [11]. В процессе *версионирования фрагмента база знаний* строится древовидно линейная структура состояний этого фрагмента базы знаний (*дерево состояний фрагментов базы знаний*) и древовидно линейная структура идентификаторов состояний этого фрагмента базы знаний (*дерево идентификаторов состояний фрагментов базы знаний*). Обе структуры представляются на *SC-коде*.

Состоянием фрагментом базы знаний являются пары из двух множеств. Элементами первого множества являются дуги принадлежности частных подструктур декомпозиции более общей структуры, которые в версии этого *состояния фрагмента базы знаний* не были изменены. Элементы второго множества - это дуги принадлежности частных подструктур декомпозиции более общей структуры, которые были изменены в этой версии. Оба множества могут представлены в виде ролевого отношения (атрибутивного множества), означающего то, какие и в каком виде находятся частные структуры общей структуры состояний. Эти два множества играют важную роль в операционной семантике, используемой при реализации агентов работы с историей изменений структуры.

Восстановление версии по *состоянию фрагмента базы знаний* происходит только за один обход поддерева, то есть за обход этого состояния. Структуры, располагающиеся в иерархии ниже чем та структура, в которую входит дуга принадлежности композиции более общей структуры, которая в свою очередь является элементом атрибутивного множества дуг принадлежности изменённых (неизменённых) структур, считаются изменёнными (неизменёнными) до тех пор, пока не встретится дуга из второго атрибутивного множества, а те, которые выше дуги, считаются неизменёнными (изменёнными) до тех пор, пока не встретится дуга из первого атрибутивного множества. Такой подход прост в реализации и не требует большого количества копий связей декомпозиции и дуг принадлежности изменённых структур к ним по сравнению с предыдущими случаями.

Рассмотрим пример версионирования конкретного нейросетевого метода, представленного на *Нейро-SCP* и который был описан в работе [9]. Данный метод решает классическую задачу “ИСКЛЮЧАЮЩЕГО ИЛИ” [13]. На рисунке 1 представлена схема *и.н.с.*, решающее эту задачу.

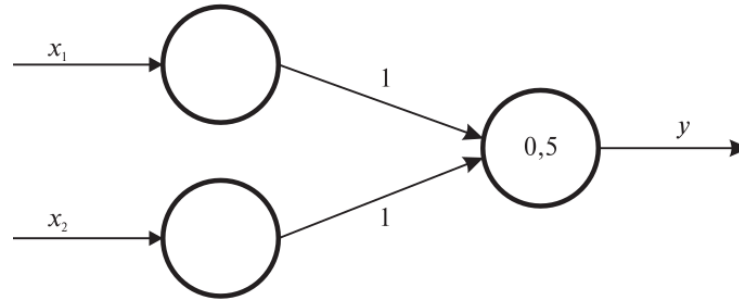


Рисунок 1. Схема и.н.с., решающей задачу ИСКЛЮЧАЮЩЕЕ ИЛИ [11]

На *Нейро-SCP* реализация этой и.н.с. сводится к одному оператору интерпретации слоя, для которого задана матрица весов синапсов, порог и функция активации. На рисунке 2 представлен этот оператор.

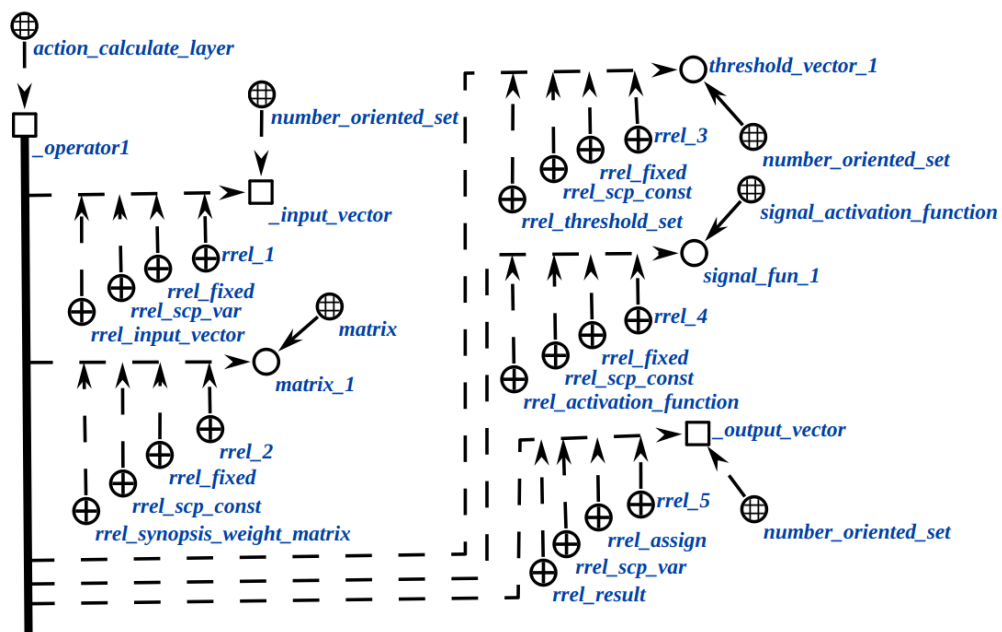


Рисунок 2. Оператор интерпретации слоя и.н.с. на Нейро-SCP

Любой из этих элементов может измениться в процессе обучения и реконфигурации и.н.с.. В рамках предлагаемого подхода такой и.н.с. будет соответствовать свой *дереву состояний*. Предположим, что данная и.н.с. была обучена и у нее изменилась матрица весов синапсов единственного слоя.

На рисунке 3 представлено дерево состояний, в котором хранятся две версии одной и той же и.н.с. - до обучения и после обучения. Корнем данного дерева является экземпляр класса истории (history). С корнем связаны два состояния, каждое из которых является ориентированной парой множеств.

Для каждого состояния задано время начала существования состояния. Все элементы, описывающие состояния, являются *темпоральными*, так как существуют временно и каждое состояние может смениться другим.

Для первого состояние первое множество было пустое, так как это первое состояние в дереве и оно не может ссылаться на предыдущие состояния. Второе множество по описанному выше принципу описывает все элементы, которые были добавлены в этой версии. Так как это первое состояние, то с помощью этого множества описываются все дуги принадлежности, участвующие в описании оператора.

Первое множество второго состояния, множество не изменившихся по сравнению с предыдущим состоянием элементов, описывается с помощью конструкций, участвующих в описании второго множества первого состояния. Второе же множество второго состояния описывает только изменения, а если быть точнее, описывает только те дуги принадлежности, которые были добавлены в новом состоянии. Таким образом, вы можем понять, что в новом состоянии изменилась матрица весовых коэффициентов. Можно предположить, что *и.н.с.* была переобучена.

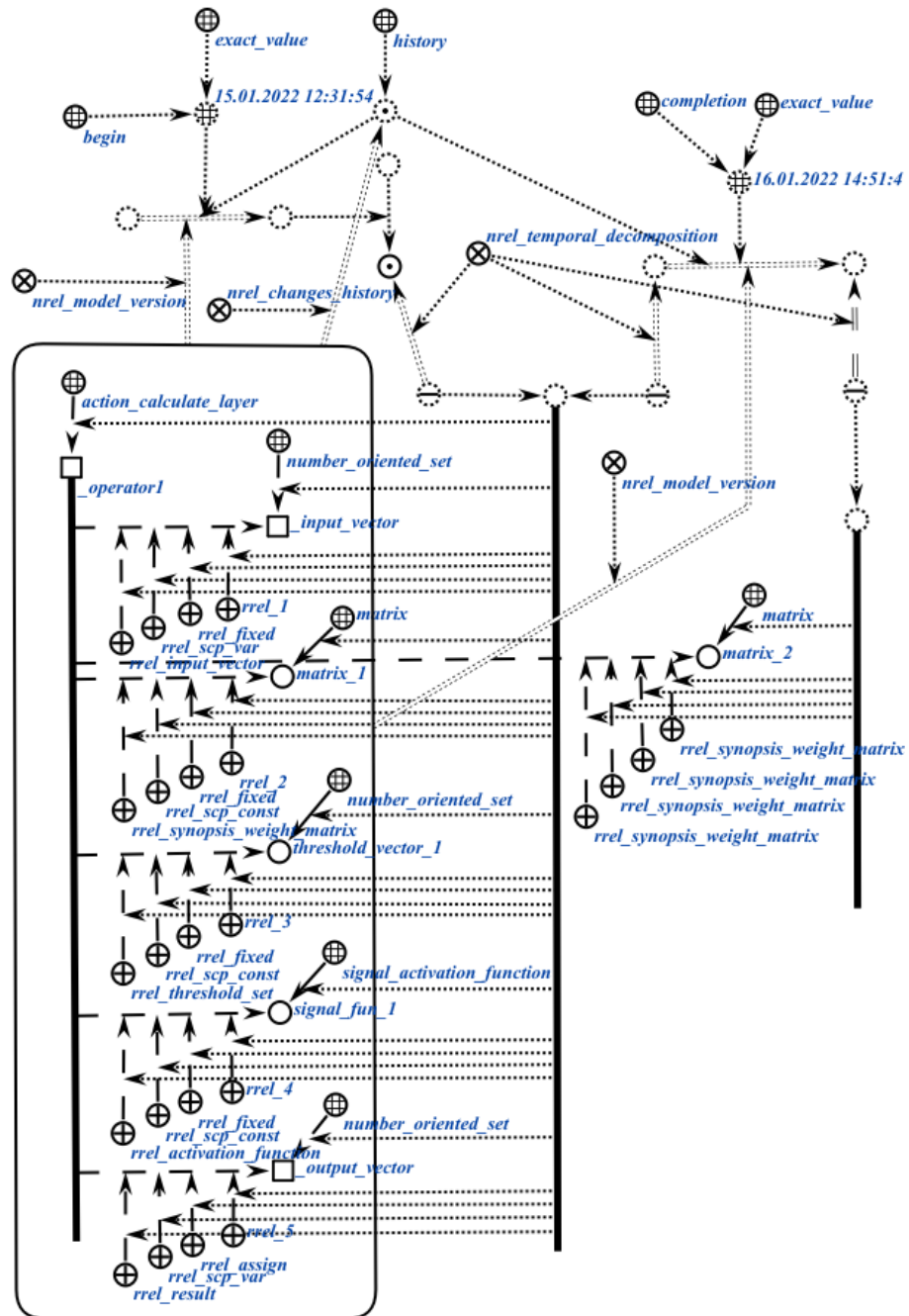


Рисунок 3. Пример дерева состояний и.н.с.

ЗАКЛЮЧЕНИЕ

Описанная модель версионности позволяет с минимальными затратами хранить и восстанавливать любое состояние не только нейросетевого метода, описанного с помощью *Нейро-SCP* в базе знаний, но и любой фрагмент базы знаний в принципе.

С помощью модели можно описывать не только то, что было изменено, но и то, в какое время и под какими воздействиями. При необходимости, каждое состояние может быть описано с различных сторон - эффективности, достоверности, историей эксплуатации и т.д.

В деле конвергенции и интеграции и.н.с. с базами знаний интеллектуальных систем такая модель позволит рассматривать *и.н.с.* не только как *метод решения задач*, но и как объект проектирования. Автоматизация такого проектирования становится задачей, которую можно поставить той же *интеллектуальной системой*, в которой описана и интерпретируется *нейронная сеть*.

Наличие унификации представления знаний, описывающих задачу, и методов решения этих задач позволяет использовать для проектирования *и.н.с.* весь арсенал методов решения задач *интеллектуальной системы*. Такие методы могут добиться хороших результатов в деле автоматизации проектирования за счет доступа к знаниям, описывающим задачу, которую должна решать проектируемая *и.н.с.*, контексту решаемой задачи, истории решения интеллектуальной системой схожих задач и т. д.

Дальнейшим развитием данной работы является проектирование и реализация интеллектуальной среды проектирования *и.н.с.*, которая позволит автоматизировать различные виды деятельности проектировщиков *и.н.с.*

REFERENCES

1. Голенков В. В., Гулякина Н. А., Шункевич Д. В. Открытая технология онтологического проектирования, производства и эксплуатации семантически совместимых гибридных интеллектуальных компьютерных систем. – 2021.
2. Alzubaidi L. et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions //Journal of big Data. – 2021. – Т. 8. – С. 1-74.
3. Castelvechi D. Can we open the black box of AI? //Nature News. – 2016. – Т. 538. – №. 7623. – С. 20.
4. Ribeiro M. T., Singh S., Guestrin C. " Why should i trust you?" Explaining the predictions of any classifier //Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. – 2016. – С. 1135-1144.
5. Saul L. K., Weiss Y., Bottou L. (ed.). Advances in neural information processing systems 17: proceedings of the 2004 conference. – MIT Press, 2005. – Т. 17.
6. Golovko V. A. et al. Integration of artificial neural networks and knowledge bases //Открытые семантические технологии проектирования интеллектуальных систем. – 2018. – №. 8. – С. 133-147.
7. Golovko V. et al. Neuro-symbolic artificial intelligence: application for control the quality of product labeling //Open Semantic Technologies for Intelligent System: 10th International Conference, OSTIS 2020, Minsk, Belarus, February 19–22, 2020, Revised Selected Papers. – Cham : Springer International Publishing, 2020. – С. 81-101.
8. Kroshchanka A. et al. A Neural-Symbolic Approach to Computer Vision //Open Semantic Technologies for Intelligent Systems: 11th International Conference, OSTIS 2021, Minsk,

Belarus, September 16–18, 2021, Revised Selected Papers. – Cham : Springer International Publishing, 2022. – С. 282-309.

9. Kovalev M., Kroshchanka A., Golovko V. Convergence and integration of artificial neural networks with knowledge bases in next-generation intelligent computer systems. – 2022.
10. Ивашенко В. П. Семантическое протоколирование процессов обработки знаний. – 2017.
11. Ивашенко В. П. Идентификация бинарно порождаемых событий процессов обработки знаний для семантического протоколирования. – 2018.
12. Shunkevich D. V. Agent-oriented models, methods and tools of compatible problem solvers development for intelligent systems. – 2018.
13. Головко В. А., Краснопрошин В. В. Нейросетевые технологии обработки данных. – 2017.