

## АНОТАЦІЯ

Вдовиченко Р. О. Розріджено-розподілене подання структур даних у нейронних мережах. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 113 – Прикладна математика. – Інститут кібернетики ім. В. М. Глушкова Національної академії наук України, Київ. – 2022.

Дисертаційна робота присвячена дослідженню феноменологічних моделей пам'яті людини та тварин шляхом подання в нейронних мережах даних, що мають певну структуру (ієрархічну, семантичну тощо). Основним завданням дисертаційної роботи є побудова та аналіз гібридного семантичного сховища, яке би мало можливість зберігати цілісні дані (наприклад, структури взаємопов'язаних і послідовних пар ключ-значення) у нейронній мережі. Конструкції пам'яті для вирішення такої задачі пропонувались у 1990-их, проте не є практичними через недостатню масштабованість та низьку щільність зберігання. Запропонована модель CS-SDM за рахунок використання третьої теорії – стискаючих вимірювань - заповнює існуючий розрив між двома феноменологічними підходами до моделювання біологічної пам'яті. Усе вище перераховане зумовлює актуальність дисертаційної роботи.

Наукова новизна роботи полягає в розробці та дослідженні характеристик нової гібридної моделі розріджено-розподіленої пам'яті CS-SDM. Ця модель вперше поєднала два напрями феноменологічного моделювання пам'яті, забезпечивши умови ефективного використання розріджено-розподіленої пам'яті типу SDM. Також уперше було запропоноване застосування теорії стискаючих вимірювань (CS) для моделювання природної пам'яті. Ефективність CS-SDM доведена як формально, так і

експериментально. CS-SDM є першою штучною нейромережею, що у повній мірі та з практично придатною ємністю дозволяє зберегати структуровані данні, тобто придатна для збереження семантики.

Модель CS-SDM має ряд прикладів застосування. CS-SDM з практично придатною ємністю дозволяє зберігати струкуровані данні, що відкриває перспективи її використання у різноманітних задачах штучного інтелекту та як складової нейромережних моделей у машинному навчанні. Також CS-SDM може використовуватись у тих галузях людської діяльності, де застосовується штучний інтелект: робототехніці, семантичному пошуку, генерації контенту у соціальних мережах, медичної діагностики тощо. В ході дослідження була розроблена програмна бібліотека з відкритими кодами, що реалізує CS-SDM на графічних процесорах (на платформі NVIDIA CUDA) і також містить реалізації адаптованих до збереження розріджених векторів конструкцій SDM Канерви і Джекела. Бібліотека впроваджена у складі Базового програмного забезпечення суперкомп'ютерного комплексу СКІТ у Центрі колективного користування обладнанням суперкомп'ютерного комплексу "СКІТ" (ЦККО СКК "СКІТ") в Інституті кібернетики ім.В.М.Глушкова НАН України. Код бібліотеки є відкритим та доступним для інших дослідників на платформі GitHub.

**Зміст дисертації.** У **вступі** обґрунтовано актуальність теми, сформульовано мету та задачі досліджень, розкрито наукову новизну та практичну цінність роботи, представлено її загальну характеристику.

У **розділі 1** розглянуто основні етапи розвитку моделювання пам'яті, зокрема розріджено-розподілених конструкцій, а також їхнє застосування. Дано характеристику розвитку методів цифрової обробки даних, які були використані у дослідженні. Проаналізовано роботи вітчизняних та закордонних вчених: Глушкова В. М., Івахненка О.Г., Куссуль Е.М., Рачковського Д.А., Тульчинського В.Г., Albus J.S., Billings G., Brindley G.S., Brogliato M.S., Candes E., Clark W.A., Donoho D., Fodor

J.A., Forrest S., Frady E.P., Furber S.B., Gayler R.W., Hebb D.O., Hinton G.E., Holland J.H., Hopfield J., Ito M., Jaeckel L., Kanerva P., Karlsson R., Kohonen T., Kristoferson J., Laiho M., Linnainmaa S., Lustig M., Marr D., McCulloch W., Minsky M., Mnih A., Nyquist H., Papert S., Perelson A.S., Pylyshyn Z.W., Rogers D., Romberg. J., Rosenblatt F., Rumelhart D.E., Salakhutdinov R., Schelegel K., Shannon C.E., Sjodin G., Smith D.J., Smolensky P., Tao T., Turing A., Von Neumann J., Werbos P.J., Williams R.J., Yoshida M. У роботах вищенаведених авторів викладено загальновідомі підходи до моделювання пам'яті та класичні конструкції, аспекти їх якісних характеристик і обмежень.

Викладено теоретичні основи векторно-символьних обчислень і розріджено-розподілених представлень, а також методу стискаючих вимірювань. На основі проведеного огляду зроблено висновки та поставлено задачі для дослідження.

У **розділі 2** запропоновано конструкцію гібридної розріджено-розподіленої пам'яті CS-SDM. CS-SDM являє собою інтеграцію класичних підходів до моделювання пам'яті сімейства SDM та сучасних досягнень теорії стискаючих вимірювань (Compressive Sensing). Доведено відповідність схеми CS-SDM вимогам Compressive Sensing. Наведено алгоритми базових операцій CS-SDM (ініціалізація, пошук активованих комірок, запис, зчитування, фіналізація). Отримано оцінки їх обчислювальної складності для випадків послідовної та паралельної реалізації. Представлено повну схему обробки семантичних та структурованих даних із застосуванням CS-SDM у якості очищуючої пам'яті. Побудовано загальну схему алгоритму постановки експериментів, використано кешування для прискорення відновлення бінарних розріджених векторів.

У **розділі 3** отримано ймовірнісні оцінки активації фізичних комірок та середньої кількості активованих комірок CS-SDM. Наведено результати експериментів для векторів із різною кількістю ознак  $S$  та різним ступенем стискання даних. Експерименти демонструють, що ємність CS-SDM як очищувальної пам'яті для

бінарних розріджено-розподілених представлень суттєво перевищує ємність більш ранніх конструкцій SDM (Kanerva, Jaeskel). Проведено експерименти із різними методами відновлення бінарних векторів (CoSaMP, LinProg). Показано, що вибір методу відновлення бінарного вектора суттєво впливає на якість відновлення при граничному стисканні. Досліджено застосування CS-SDM для більш складних задач негомогенної розрідженості.

У розділі 4 досліджено використання обчислювальної платформи CUDA для реалізації та проведення експериментів із CS-SDM: розглянуто програмну модель, ієрархію потоків, типи пам'яті, мову програмування CUDA C/C++. Наведено програмну реалізацію операцій CS-SDM, запропонованих в якості алгоритмів у розділі 2. Показано реалізацію відновлення бінарних розріджених векторів мовою програмування Python двома методами (CoSaMP і LinProg) із використанням кешування. Описано програмну бібліотеку для роботи із CS-SDM та двома класичними конструкціями SDM (Kanerva, Jaeskel), яку було розроблено в ході дисертаційної роботи; код бібліотеки є відкритим для інших дослідників.

**Ключові слова:** розріджено-розподілені подання, нейронні мережі, стискаючі вимірювання, векторно-символьні архітектури.

## ABSTRACT

Vdovychenko R.O. Sparse Distributed Representation of Structured Data in Neural Networks. – Qualifying scientific work as a manuscript.

Dissertation for a Doctor of Philosophy Degree by specialty 113 Applied mathematics.  
– V.M. Glushkov Institute of Cybernetics of the National Academy of Science of Ukraine.  
– Kyiv, 2020.

The dissertation is focused on studying phenomenological models of human and animal memory by presenting data with a specific structure (hierarchical, semantic, etc.) in neural networks. The main task of the dissertation is constructing and analysing a hybrid semantic store that would be able to store complete data (for example, structures of interconnected and consecutive key-value pairs) in a neural network. Memory designs to solve this problem were proposed in the 1990s but are not practical due to insufficient scalability and low storage density. The proposed CS-SDM model fills the existing gap between two phenomenological approaches to biological memory modelling by using the third theory – Compressive Sensing. All the statements above determine the relevance of the dissertation work.

The scientific novelty of the work consists of developing and researching the characteristics of a new hybrid model of sparse-distributed memory CS-SDM. For the first time, this model combined two directions of phenomenological modelling of memory, providing conditions for the effective use of sparse-distributed memory of the SDM type. Also, applying the theory of Compressive Sensing (CS) was proposed for the first time to model natural memory. The effectiveness of CS-SDM has been proven both formally and experimentally. CS-SDM is the first artificial neural network that entirely and practically suitable capacity allows you to store structured data, which is suitable for preserving semantics.

The CS-SDM model has numerous applications. CS-SDM, with a practically usable capacity, allows you to store structured data, which opens prospects for its use in various tasks of artificial intelligence and as a component of neural network models in machine learning. Also, CS-SDM can be used in those fields of human activity where artificial intelligence is used: robotics, semantic search, content generation in social networks, medical diagnostics, etc. In the course of the research, an open-source software library was developed that implements CS-SDM on graphics processors (on the NVIDIA CUDA platform) and also contains implementations of the Kanerva and Jackel SDM designs adapted to the preservation of sparse vectors. The library is implemented as part of the essential software of the SCIT supercomputer complex in the Center for collective use of the equipment of the SCIT supercomputer complex (CCKO SKK "SCIT") at the V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine. The library code is open and available to other researchers on the GitHub platform.

The content of the dissertation. The **introduction** substantiates the topic's relevance, formulates the research's purpose and tasks, reveals the work's scientific novelty and practical value, and presents its general characteristics.

**Chapter 1** discusses the main stages of the development of memory modelling, particularly sparse-distributed structures, and their application. A description of the development of digital data processing methods used in the study is given. The works of domestic and foreign scientists were analyzed: V.M. Hlushkova, O.G. Ivakhnenko, E.M. Kussul, D.A. Rachkovsky, V.G. Tulchinsky, Albus J.S., G. Billings, G.S. Brindley, M.S. Brogliato, Candes E., Clark W.A., Donoho D., Fodor J.A., Forrest S., Frady E.P., Furber S.B., Gayler R.W., Hebb D.O., Hinton G.E., Holland J.H., Hopfield J., Ito M., Jaeckel L., Kanerva P., Karlsson R., Kohonnen T., Kristoferson J., Laiho M., Linnainmaa S., Lustig M., Marr D., McCulloch W., Minsky M., Mnih A., Nyquist H., Papert S., Perelson AS, Pylyshyn ZW, Rogers D, Romberg. J., Rosenblatt F., Rumelhart D.E., Salakhutdinov R.,

Schlegel K., Shannon C.E., Sjodin G., Smith D.J., Smolensky P., Tao T., Turing A., Von Neumann J., Werbos P.J., Williams R.J., Yoshida M. Well-known approaches to memory modelling and classical designs, aspects of their qualitative characteristics and limitations are described in the works of the authors mentioned above.

The theoretical foundations of vector-symbol calculations, sparse-distributed representations, and the compressive measurements method are outlined. Based on the conducted review, conclusions were drawn, and research tasks were set.

In **Chapter 2**, the design of the CS-SDM hybrid sparse-distributed memory is proposed. CS-SDM integrates classic approaches to memory modelling of the SDM family and modern achievements of the Compressive Sensing theory. Compliance of the CS-SDM scheme with the requirements of Compressive Sensing is proven. Algorithms of basic CS-SDM operations (initialization, search for activated cells, writing, reading, finalization) are given. Estimates of their computational complexity were obtained for serial and parallel implementation cases. A complete scheme for processing semantic and structured data using CS-SDM as a cleaning memory is presented. The general scheme of the experiment setup algorithm was built, and caching was used to speed up the recovery of binary sparse vectors.

In **Chapter 3** probabilistic estimates of the activation of physical cells and the average number of activated CS-SDM cells are obtained. The results of experiments for vectors with a different number of S features and a different degree of data compression are given. Experiments demonstrate that the capacity of CS-SDM as a clearing memory for binary sparse-distributed representations significantly exceeds the capacity of earlier SDM designs (Kanerva, Jaeckel). Experiments with various recovery methods of binary vectors (CoSaMP, LinProg) were conducted. It is shown that the binary vector restoration method's choice significantly affects the restoration quality at the limit compression. The application of CS-SDM for more complex problems of inhomogeneous sparsity is studied.

In **Chapter 4** the use of the CUDA computing platform to implement and conduct experiments with CS-SDM is explored: the program model, thread hierarchy, memory types, and the CUDA C/C++ programming language are considered. The software implementation of CS-SDM operations, proposed as algorithms in Chapter 2, is presented. The implementation of recovery of binary sparse vectors in the Python programming language by two methods (CoSaMP and LinProg) with caching is shown. The software library for working with CS-SDM and two classic SDM designs (Kanerva, Jaeckel) is described, which was developed during the dissertation work; the library code is open to other researchers.

**Keywords:** Sparse Distributed Representations, Neural Networks, Compressive Sensing, Vector Symbolic Architectures.