

АННОТАЦИЯ
научных исследований, проведенных в 1 полугодии 2018 года

Договор между Министерством образования и науки Российской Федерации, федеральным государственным автономным образовательным учреждением высшего образования "Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского",

(наименование российской образовательной организации высшего образования, научного учреждения, подведомственного Федеральному агентству научных организаций, государственного научного центра Российской Федерации - получателя гранта)

и Горбанем Александром Николаевичем _____

(фамилия, имя, отчество (при наличии) ведущего ученого)

о выделении гранта Правительства Российской Федерации для государственной поддержки научных исследований, проводимых под руководством ведущих ученых в российских образовательных организациях высшего образования, научных учреждениях, подведомственных Федеральному агентству научных организаций, и государственных научных центрах Российской Федерации

от 5 февраля 2018 г. № 14.У26.31.0022

Область наук Компьютерные и информационные науки

Направление научного исследования Масштабируемые сети систем искусственного интеллекта для анализа данных растущей размерности

Наименование лаборатории «лаборатория перспективных методов анализа многомерных данных»

1. Краткое описание выполненных работ научного исследования

В ходе реализации проекта были проведены следующие работы согласно плану работ:

- I. Обобщение теорем о стохастической отделимости;
- II. Построение теории разделения данных в многомерных пространствах малыми нейронными сетями;
- III. Построение теории аппроксимации, основанной на кусочно-квадратичных потенциалах субквадратичного роста (PQSQ potentials);
- IV. Разработка, программная реализация и тестирование универсальных алгоритмов аппроксимации данных, основанных на кусочно-квадратичных потенциалах ошибки с субквадратичным ростом;
- V. Разработка, программная реализация и тестирование алгоритмов, основанных на методе топологических грамматик, для отделения истинно многомерных проблем от редуцируемых проблем с малой внутренней размерностью;
- VI. Подготовка тестовых заданий для анализа больших и многомерных видеопотоков в реальном времени, выбор и подготовка баз сравнения;
- VII. Отбор коллекций и потоков данных для детального анализа;
- VIII. Мастер-классы для студентов, аспирантов и молодых ученых, проводимые А.Н. Горбанем;
- IX. Подготовка к публикации статей по результатам проведенной работы;
- X. Организация симпозиума по тематике проекта;
- XI. Участие ведущего ученого и членов научного коллектива в конференциях, научных семинарах, симпозиумах;
- XII. Оснащение лаборатории оборудованием, материалами и комплектующими для проведения исследований;
- XIII. Анализ крупномасштабных записей нейронной активности, разработка алгоритмов для анализа кальциевого имиджинга и электрофизиологических данных, разработка и анализ математических моделей.

2. Краткое описание полученных научных результатов

I, II. Было проведено обобщение серии теорем стохастической отделимости и разделения данных в многомерных пространствах малыми нейронными сетями. Подготовлены и представлены научные статьи.

A.N. Gorban, A. Golubkov, B. Grechuk, E.M. Mirkes, I.Y. Tyukin, "Correction of AI systems by linear discriminants: Probabilistic foundations", содержащая новые версии теорем о стохастической отделимости и разделения данных в многомерных пространствах малыми нейронными сетями представлена в журнал "Information Sciences"

Системы искусственного интеллекта (ИИ) иногда делают ошибки и время от времени будут делать ошибки в будущем. Эти ошибки обычно являются неожиданными и могут привести к серьезным последствиям. Интенсивное развитие ИИ и его практическое применение делает проблему ошибок более важной. Полная реорганизация систем может создавать новые ошибки и не всегда возможна из-за ресурсных ограничений. Важной задачей является разработка быстрых методов исправления ошибок без ущерба для существующих навыков. Мы сформулировали технические требования к «идеальным» корректорам. Такие корректоры включают бинарные классификаторы, которые отделяют ситуации с высоким риском ошибок от ситуаций, когда системы ИИ работают должным образом. В статье представлена вероятностная основа быстрой неразрушающей коррекции систем ИИ. Доказана серия новых теорем стохастического разделения. Эти теоремы предоставляют новые инструменты для быстрой неитеративной коррекции ошибок больших систем ИИ. Новые подходы становятся эффективными в области высоких измерений, для

коррекции высокоразмерных систем в многомерном мире (т. е. для обработки существенно высокоразмерных данных большими системами).

Мы доказали, что свойство сепарабельности справедливо для широкого класса распределений, включая лог-вогнутые распределения и распределения со специальным свойством «SMeared Absolute Continuity» (SmAC), определенным через отношения между объемом и вероятностью множеств исчезающего объема. Эти классы намного шире, чем гауссовские распределения. Требование независимости и идентичного распределения данных значительно ослаблено.

Был подготовлен препринт, включающий основные результаты статьи:

A.N. Gorban, B. Grechuk, I.Y. Tyukin, Augmented Artificial Intelligence: a Conceptual Framework, arXiv:1802.02172 [cs.AI]. <https://arxiv.org/pdf/1802.02172.pdf>.

I. - Рассмотрена следующая задача, возникающая при анализе эффективности алгоритмов линейного отдаления и алгоритмов, основанных на данном подходе. Множество точек в d -мерном пространстве называется 1-выпуклым, если все его точки являются вершинами выпуклой оболочки этого множества. Ранее в работе [Gorban A. N., Tyukin I. Y. Stochastic separation theorems. Neural Networks 94, 255–259 (2017)] была получена оценка мощности множества точек, являющихся 1-выпуклым с вероятностью, близкой к 1. На данном этапе получена новая более сильная оценка. Опубликован препринт:

Sidorov S. On the 1-convexity of random points in the d -dimensional spherical layer <https://arxiv.org/abs/1806.04732>

II, III, IV. - Было проведено: описание теории разделения данных в многомерных пространствах малыми нейронными сетями; описание теории аппроксимации, основанной на кусочно-квадратичных потенциалах субквадратичного роста (PQSQ potentials); описание универсальных алгоритмов и разработано программное обеспечение для аппроксимации данных, основанных на кусочно-квадратичных потенциалах ошибки с субквадратичным ростом; приведены результаты тестирования, описание алгоритмов и программного обеспечения, основанных на методе топологических грамматик, для отдаления истинно многомерных проблем от редуцируемых проблем с малой внутренней размерностью; приведены соответствующие результаты тестирования. Подготовлены и представлены научные статьи. Представлен доклад на Всемирном Конгрессе.

V. - Разработан новый метод обработки и аппроксимации больших данных сложной структуры, основанный на топологических грамматиках, применимый при больших уровнях шума. Метод может быть широко применим для анализа данных от геномики до астрономии. Представлена статья в PNAS, опубликован препринт:

Albergante L., Mirkes E.M., Chen H., Martin A., Faure L., Barillot E., Pinello L., Gorban A.N., Zinovyev A. Robust and scalable learning of data manifolds with complex topologies via EIPiGraph. arXiv preprint arXiv:1804.07580. 2018 Apr 20.

Большие наборы данных, представленные многомерными облаками данных, часто имеют нетривиальные распределения с ветвящимися траекториями и исключенными областями. Заметным примером здесь являются недавние одноцелевые транскриптомические исследования развития эмбриона. Сокращение сложности и создание компактных и интерпретируемых представлений таких данных остается сложной задачей. Большинство существующих вычислительных методов основаны на изучении отношений соседних точек данных, шага, который может плохо выполняться в случае многомерных и шумных данных. Здесь мы представляем EIPiGraph - масштабируемый и надежный метод аппроксимации наборов данных со сложными структурами, который не требует вычисления полной матрицы расстояний для данных или графа окрестности точки данных. Этот метод способен выдерживать высокие уровни шума и аппроксимировать сложные топологии через главные графовые ансамбли, которые могут быть объединены в консенсусный главный граф. EIPiGraph эффективно работает с большими и сложными наборами данных в различных областях биологии, где его можно использовать для

вывода динамики генов из одноклеточной РНК-Seq, и астрономии, где его можно использовать для изучения сложных структур в распределении галактик.

V. - Разработан новый метод для обработки биологических данных нового поколения: данных об индивидуальных клетках и траекториях их развития. Представлена статья в Nature Communication, опубликован препринт:

Chen H., Albergante L., Hsu J.Y., Lareau C.A., Bosco G.L., Guan J., Zhou S., Gorban A.N., Bauer D.E., Aryee M.J., Langenau D.M., Zinovyev A., Buenrostro J.D., Yuan G.-C., Pinello L. STREAM: Single-cell Trajectories Reconstruction, Exploration And Mapping of omics data. bioRxiv. 2018 Jan 1:302554.

Одноцепочечные транскриптомические анализы позволили провести наряду с характеристикой клеточной гетерогенности и состояний реконструкцию дерева траекторий дифференциации. Было разработано несколько методов для реконструкции траекторий развития из одноцелевых транскриптомических данных. Однако усилия по анализу одноэлементных эпигенетических данных и визуализация траектории остаются ограниченными. Нами был представлен интерактивный конвейер STREAM, способный распутывать и визуализировать сложные траектории ветвления как из одноцепочечных транскриптомических, так и эпигемных данных.

VI, VII. - Предложена методология использования конкурирующих нейронных сетей для генерации сигналов ЭКГ с той или иной патологией. Данный подход может быть использован для обучения классификаторов для автоматической диагностики заболеваний сердечно-сосудистой системы с небольшими обучающими выборками. Набрана экспериментальная база данных электрокардиограмм (1000 шт.). Проведена сегментация ЭКГ - разбиение ЭКГ на сердечные циклы с выделением всех характерных особенностей, RR-интервал, QRS комплексов и др. Данные сняты с помощью 12-канального электрокардиографа. Исследованы особенности собранного набора данных ЭКГ, и выявлены его проблемы (в частности - при большой размерности входных данных почти все классы являются редкими, что делает борьбу с переобучением нейросетей нетривиальной) и достоинства (заданная экспертами информативная иерархия диагнозов, большое их количество). Набор данных ЭКГ адаптирован к использованию в качестве входных данных для искусственной нейросети. Проведена подготовительная инженерная работа для удобства проведения численных экспериментов с разными архитектурами нейросетей. Написан вариант сверточной нейросети для дальнейшего улучшения

VI, VII. - Проведен отбор и коллекция экспериментальных данных ЭМГ-паттернов. Подготовлена и опубликована научная статья.

Lobov, S.; Krilova, N.; Kastalskiy, I.; Kazantsev, V.; Makarov, V.A. Latent Factors Limiting the Performance of sEMG-Interfaces. Sensors 2018, 18, 1122. doi: 10.3390/s18041122

Для дальнейшей работы по классификации данных за отчетный период набрана экспериментальная база данных по электромиографическим (ЭМГ) паттернам. Паттерны представляют собой совокупность «сырых» (RAW) сигналов, регистрируемых с помощью 8-канального миографа. Испытуемые выполняли следующую серию статичных жестов кистью руки: (1) кисть в расслабленном положении, (2) сжатие кисти в кулак, (3) сгибание кисти, (4) разгибание кисти, (5) приведение кисти в лучевую сторону, (6) приведение кисти в локтевую сторону, (7) вытягивание пальцев, (8) супинация кисти, (9) пронация кисти. Данную серию планируется в дальнейшем использовать для оптимизации задачи классификации ЭМГ-паттернов в нейромышечном интерфейсе. Часть испытуемых выполняла дополнительно четыре составных жеста, являющихся промежуточными между жестами (3) и (5), (4) и (5), (3) и (6), (4) и (6). Данную серию планируется использовать для оптимизации задачи ЭМГ-управления объектом с двумя степенями свободы. Для всех записей велось протоколирование выполняемого жеста (регистрации метки класса).

В настоящий момент база данных содержит:

а. ЭМГ паттерны, зарегистрированные с помощью браслета MYO Thalmic (частота дискретизации 300 Гц):

42 человека – выполнение шести основных жестов;

10 человек – выполнение девяти основных жестов;

77 человек – выполнение шести основных жестов и четырех дополнительных (составных).

б. ЭМГ паттерны, зарегистрированные с помощью миографической системы DELSYS Trigno (частота дискретизации 2000 Гц): 8 человек – выполнение девяти основных жестов.

В. ЭМГ паттерны, зарегистрированные с помощью браслета MYO Thalmic при выполнении игровой задачи на управление виртуальным объектом – 25 человек. В данном случае запись включала в себя миографические паттерны при выполнении большого количества движений кистью и метка класса не регистрировалась. Данную серию планируется в дальнейшем использовать для алгоритмов обучения без учителя.

VIII. - Под руководством Горбаня А.Н. и Осипова Г.В. была проведена серия установочных семинаров и мастер-классов для участников проекта, где были обсуждены траектории развития научных исследований участников проекта, определены результаты и сроки реализации (докладчики семинаров - Бирюков Р.С., Серeda Я.А., Смирнов Л.А., Колосов А.В., Яхно В.Г., Лобов С.А., Болотов М., Кузенков О., Стасенко С.В., Тельных А.А., Казаков А., Болотов М.). А.Н. Горбань провел публичные лекции для студентов (в ННГУ) и школьников (Лицей "Вторая Школа", Москва).

IX. - Опубликованные статьи в журналах из коллекции Web of Science:

1. AN Gorban, Model reduction in chemical dynamics: slow invariant manifolds, singular perturbations, thermodynamic estimates, and analysis of reaction graph, *Current Opinion in Chemical Engineering* 21, 48-59. – (**Current Opinion in Chemical Engineering, IF 2016 3.403, Q1 in Engineering, Chemical, GIF percentile 85.5%**)
2. Lages, J., Shepelyansky, D. L., & Zinovyev, A. (2018). Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks. *PLoS ONE*, 13(1). <https://doi.org/10.1371/journal.pone.0190812> - (**PLoS One, IF 2016 2.806, Q1 in Multidisciplinary Sciences, GIF percentile 77.3%**)
3. Lobov, S.; Krilova, N.; Kastalskiy, I.; Kazantsev, V.; Makarov, V.A. Latent Factors Limiting the Performance of sEMG-Interfaces. *Sensors* 2018, 18, 1122. doi: 10.3390/s18041122 – (**Sensors, IF 2016 2.677, Q1 in Instruments and Instrumentation, GIF percentile 83.6%**)
4. Naldi A., Hernandez C., Levy N., Stoll G., Monteiro P.T., Chaouiya C., Helikar T., Zinovyev A., Calzone L., Cohen-Boulakia S., Thieffry D., Paulevé L. The CoLoMoTo Interactive Notebook: Accessible and Reproducible Computational Analyses for Qualitative Biological Networks, *Front. Physiol.*, 19 June 2018, <https://doi.org/10.3389/fphys.2018.00680> - (**Frontiers in Physiology, IF 2016 4.134, Q1 in Physiology, GIF percentile 82.7%**)
5. Tyukin, I., Gorban, A. N., Calvo, C., Makarova, J., & Makarov, V. A. (2018, March 19). High-Dimensional Brain: A Tool for Encoding and Rapid Learning of Memories by Single Neurons. *Bulletin of Mathematical Biology*, pp. 1–33. Springer New York – (**IF 2016 1.26 (2017 1.48) Q3 in Biology, Q3 in Mathematical and Computational Biology, JIF percentile 40.5% (39% in 2017) in Biology, JIF percentile 39.4% (33% in 2017) in Mathematical and Computational Biology**)
6. Gorban, A. N., & Çabukoğlu, N. (2018). Basic model of purposeful kinesis. *Ecological Complexity*, 33, 75–83. <https://doi.org/10.1016/j.ecocom.2018.01.002> - (**IF 2016 1.784 (2017 1.634) Q3 in Ecology JIF percentile 46.7% (37.025 % in 2017)**)

- Подготовленные и представленные статьи:

- A.N. Gorban, A. Golubkov, B. Grechuk, E.M. Mirkes, I.Y. Tyukin, Correction of AI systems by linear discriminants: Probabilistic foundations, *Information Sciences*.

- O. Kuzenkov, A. Morozov. Towards constructing a mathematically rigorous framework for modelling evolutionary fitness// Bulletin of Mathematical Biology.
- Albergante L., Mirkes E.M., Chen H., Martin A., Faure L., Barillot E., Pinello L., Gorban A.N., Zinovyev A. Robust and scalable learning of data manifolds with complex topologies via EIPiGraph. PNAS US (arXiv preprint arXiv:1804.07580. 2018 Apr 20)
- S. Sidorov, On the 1-convexity of random points in the d-dimensional spherical layer, On the 1-convexity of random points in the d-dimensional spherical layer, <https://arxiv.org/abs/1806.04732>
- A.N. Gorban, B. Grechuk, I.Y. Tyukin, Augmented Artificial Intelligence: a Conceptual Framework, arXiv:1802.02172
- I.A. Lazarevich, S.S. Stasenko, M.A. Rozhnova, E.V. Pankratova, E.A. Ditatev, V.B. Kazantsev, Dynamics of the brain extracellular matrix governed by interactions with neural cells, Physical Review E.
- Chen H., Albergante L., Hsu J.Y., Lareau C.A., Bosco G.L., Guan J., Zhou S., Gorban A.N., Bauer D.E., Aryee M.J., Langenau D.M., Zinovyev A., Buenrostro J.D., Yuan G.-C., Pinello L. STREAM: Single-cell Trajectories Reconstruction, Exploration And Mapping of omics data. Nature Communication (bioRxiv. 2018 Jan 1:302554).

X. По тематике проекта идет организация симпозиума «NEURODYNAMICS AND ARTIFICIAL INTELLIGENCE» в рамках международной конференции «Volga Neuroscience meeting 2018», которая состоится с 22 по 27 июля 2018 года <http://conf.neuro.unn.ru/vnm-2018/>.

XI. Коллектив исполнителей проекта принял участие в следующих конференциях и семинарах:

- 6-12 июля 11-й нейронаучный форум FENS 2018, г. Берлин, Германия;
- 22-27 июля Volga Neuroscience Meeting 2018, Н.Новгород – Самара – Н.Новгород,
- 22 - 30 июня 2018 года, международный семинар «Complex Heterogeneous Multiphase Systems» Эдинбург, Великобритания
- 30 -7 июля 2018 года, международный семинар «International Workshop on Nonequilibrium Thermodynamics IWNET 2018» Синт-Михильсгестель, Нидерланды
- 7 - 15 июля 2018 года, объединенная конференция по нейронным сетям (The International Joint Conference on Neural Networks (IJCNN)) в рамках Всемирного Конгресса по Вычислительному Интеллекту (World Congress in Computational Intelligence, WCNN) в Рио-де-Жанейро, Бразилия. Организована специальная секция Всемирного конгресса «Нейронный интеллект послезавтра» (Neural Intelligence After Tomorrow) для представления результатов проекта. Представлено три доклада исполнителей проекта.

XII. Подготовлен проект оснащения лаборатории, включающий список оборудования, оргтехники и мебели для офисных помещений лаборатории.

XIII. В рамках пункта работ «Анализ крупномасштабных записей нейронной активности, разработка алгоритмов для анализа кальциевого имиджинга и электрофизиологических данных, разработка и анализ математических моделей» были выполнены следующие работы:

а). На основе анализа экспериментальных данных разработана математическая модель взаимодействия популяции нейронов с молекулами внеклеточного матрикса мозга (ВКМ). По результатам подготовлена научная статья.

Ivan A. Lazarevich, Sergey S. Stasenko, Maia A. Rozhnova, Evgeniya V. Pankratova, Victor B. Kazantsev, Dynamics of the brain extracellular matrix governed by interactions with neural cells, Physical Review E, 2018

б). Проведено исследование полученной регрессионной модели задачи предсказания грозовых разрядов в заданной области пространства на основе данных с доплеровских радаров wsd-88 и lightning detection network и показано, что данный метод пригоден для формирования прогноза грозовых разрядов в заданной области пространства на несколько часов вперед. Подготовлен вариант статьи по данному результату.

в). Проводилось изучение структур из различных дисперсных фаз в водных растворах. Показано, что неотъемлемым свойством воды и водных растворов является их структуризация на микроуровне - ранее не описанное явление. Подготовлен вариант статьи по данному результату.

г). Для управления мобильным агентом в меняющейся среде разработана схема, при которой управление представляется в виде двух взаимодействующих подсистем: локальной, отвечающей за непосредственное взаимодействие агента с окружением и обработку данных с датчиков, и глобальной, отвечающей за взаимодействие с окружением в рамках принятых абстракций и планирование пути в поставленной задаче. Реализована система для проведения компьютерных экспериментов, проведены эксперименты.

д). Проведены исследования динамики многомасштабных (то есть с несколькими характерными масштабами скоростей распространения сигналов) систем «реакция — диффузия». Показано, что в общем случае наличие второго быстрого масштаба существенно меняет набор характерных динамических режимов.

е). Проводится цикл работ, в которых исследование сложных многокомпонентных систем, описывающие в том числе динамику нейросетей, сводится к анализу маломерных моделей для параметров порядка - параметров, характеризующих пространственно-временные структуры сети. Исследованы процессы формирования сложных регулярных, хаотических, химерных структур в среде нелокально связанных неидентичных фазовых осцилляторов.

ж). Проведены исследования вращательной динамики системы трёх связанных маятников с топологией кольца. Они обобщены на случай произвольного числа N глобально связанных идентичных маятников. Разработана асимптотическая теория, справедливая в пределе малой диссипации, описывающая все найденные периодические колебания.

з). Проведена работа по подбору тестовых заданий для прогнозирования качественных характеристик поведения многокомпонентной водной экосистемы по наблюдаемым визуальным данным, собраны образцы сравнения по визуальным данным движения

Направлена статья совместно с А. Морозовым профессором университета Лестера (Великобритания)

O. Kuzenkov, A. Morozov. Towards constructing a mathematically rigorous framework for modelling evolutionary fitness// Bulletin of Mathematical Biology.

Создано и апробировано программное обеспечение для реализации предложенного алгоритма.

и). Разработан алгоритм проверки наличия или отсутствия изучаемых характеристик в заданной системе на основе анализа имеющейся базы образцов сравнения. Алгоритм использует новый подход, опирающийся на степенную аппроксимацию функции эволюционного фитнеса.

к). Разрабатывались нейроноподобные алгоритмы для распознавания символической информации. Рассматривается задача распознавания номерных знаков железнодорожных вагонов: захват объекта из видеопотока и распознавание объектов. Подготовлен вариант статьи по данному результату.