

Хранилище симитированных типовых сигналов как основа разработки быстрых алгоритмов*

И.И. Дейкин, В.В. Сюзев, Е. В.Смирнова, А.В. Пролетарский

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», г. Москва, Россия

Аннотация. Статья посвящена структуре хранилища многомерных типовых сигналов и процессов в рамках задачи цифровой фильтрации, основанной на быстрых алгоритмах имитации и операторе универсального адаптивного матричного преобразования. Хранилище содержит актуальные уравнения имитации; симитированные сигналы в виде таблиц значений и изображений; оригинальный код на Python, C++ и других языках программирования; дополнительную лицензионную информацию. В статье рассмотрены различные подходы к созданию хранилища. Выбранный путь заключается в формировании набора данных, его публикации на ресурсе GitHub и построении онтологии на данном наборе. Принципы структурирования данных и организации такой онтологии представлены в статье. Представлен готовый опубликованный набор данных.

Ключевые слова: хранилище данных, хранилище сигналов, многомерные сигналы, имитация сигналов, цифровая обработка сигналов, ЦОС, онтология, гармонические методы.

DOI: 10.14357/20790279230105

Введение

Задача имитации сигналов, рассматриваемая в рамках цифровой обработки сигналов, становится особенно актуальной в условиях цифровой экономики. Имитация позволяет не только проводить экономически выгодное натурное моделирование, посылая реалистичные сигналы на проектируемое оборудование, но также моделировать состояния различных систем в задачах калибровки существующего оборудования, тренировки персонала и прогнозирования [1, 2]. Существуют различные методы имитации сигналов. Распространенные методы имитации сигналов предлагают выбирать между качеством имитации и экономией вычислительных ресурсов [3]. Одним из перспективных направлений исследований является исследование гармонических методов имитации сигналов, представляющих возможность совмещения качества и экономного расхода вычислительных ресурсов

для определенных типов сигналов. Гармонические методы имитации разнообразны, и поэтому становится целесообразным рассмотрение различных гармонических базисов и различных подходов к имитации в рамках гармонической имитации [4].

Коллективная разработка новых методов и инструментов цифровой обработки сигналов нуждается в организации хранилища результатов: математических выкладок, результатов их экспериментальной проверки и алгоритмов реализации методов имитации сигналов и процессов, часть из которых защищена авторскими свидетельствами на Программы для ЭВМ [5 – 8]. Вследствие необходимости их анализа предлагается разработать хранилище симитированных эталонных сигналов и сопутствующих материалов: теоретических выкладок, кодов алгоритмов, лицензионных документов.

Разработка современного хранилища опирается на совершенствование алгоритмов, структур данных и их взаимодействия. Тенденции вроде кэш-осознанности и энергетической осознанности, параллельные и облачные вычисления отражают требования, предъявляемые современным технологиям, и мотивируют поиск способов совершенствования базового канона алгоритмов

* Исследование выполнено за счет гранта Российского научного фонда №22-11-00049, <https://rscf.ru/project/22-11-00049/> («Разработка корреляционной теории моделирования многомерных сигналов и процессов в гибридных системах искусственного интеллекта реального времени»).

и структур данных общего назначения [9]. Особенно важна организация устойчивого развития в фундаментальной для современного общества сети Интернет, работа которой требовательна к вычислительным и, значит, к энергетическим ресурсам [10]. Хотя стандартные алгоритмы и простые структуры данных полезны при разработке новых систем и составляют основу для будущего развития, существует потребность в новых умных способах организации взаимодействия данных и алгоритмов.

Проект #22-11-00049 при поддержке Российского Научного Фонда, включающий сравнительный анализ различных методов имитации сигналов, требует организации взаимного доступа исследователей к результатам исследования этих методов. Разрабатываемое хранилище призвано позволить исследователям, работающим над разными методами, проверять себя и коллег путем отработки методов на одинаковых экспериментальных настройках и сравнения полученных результатов. Также предполагается доступ третьих лиц к результатам научной деятельности команды.

1. Возможные подходы

Существует множество различных подходов к организации открытого доступа к научным результатам [11 – 20]. Используемые в них алгоритмы и структуры данных соответствуют базовому канону алгоритмов и структур данных общего назначения, что эффективно для ресурсов малого масштаба, но может представлять проблему в их дальнейшем развитии. Однако такой простой подход создает базу для дальнейшего совершенствования. Анализ существующих решений позволил систематизировать их по методу доступа к данным и выделить следующие виды хранилищ:

- хранилище как отдельное программное приложение;
- хранилище как интерактивное Интернет-приложение;
- хранилище как опубликованный в сети Интернет набор данных.

Все три вида нуждаются в рассмотрении по-отдельности, в выделении позитивных и негативных черт с целью дальнейшего сравнительного анализа.

Отдельные приложения позволяют получать имитационные результаты лицам, имеющим доступ к этим приложениям. При этом люди нередко не доверяют приложениям неизвестных разработчиков и не готовы запускать такие приложения на своем оборудовании. Такая проблема не возника-

ет при обмене приложением в команде разработчиков-коллег, но может мешать распространению научных результатов за пределами команды разработчиков. Приложения не предоставляют открытого доступа к исходному коду методов имитации. Несмотря на то, что различные инструменты могут использоваться для извлечения этого кода, подобные риски малы, когда речь идет об узкоспециальном программном обеспечении.

Разработка приложения является отдельной проблемой. Исследование методов имитации сигналов в различных базисах сопровождается созданием их программной реализации. Так, участниками команды были разработаны программные системы имитации сигналов в базисе Фурье, в базисе Хартли и в других базисах [11, 12]. Разработка проводилась отдельными разработчиками под определенные задачи, что привело к гетерогенности результатов: использовались разные языки программирования, разные технологии, разные среды разработки. Объединение таких разрозненных продуктов в единую систему имитации является важной задачей и требует анализа имеющихся результатов, перевода части программного кода на другие языки программирования, проектирования обновленного интерфейса и структуры системы.

Размещение веб-приложения в сети Интернет позволяет обеспечить доступ к результатам научной деятельности третьим лицам. Среди положительных черт можно обозначить поддержку интерактивности приложения, доступ из любой точки мира и с любого устройства, отсутствие необходимости настройки системы на оборудовании пользователя. Интерактивность находит применение в исследовании спектров атомов различных веществ и спектров космических тел в ресурсе университета Вашингтона и ресурсе NIST [13, 14]. Особенный интерес предоставляет внедрение тетрадей Jupyter Notebook в веб-сайты с возможностью как их запуска с последующей интерактивной работой, так и просмотра их программного кода [15].

Организация веб-приложений, тем не менее, также имеет негативные черты. Разработка веб-приложений отличается от разработки обычных приложений и требует от команды разработчиков иных компетенций. Применяются другие языки, системы и парадигмы разработки, необходимо выбрать между поддержкой приложения на стороне клиента или на сервере, что потребует настройки и поддержки такого сервера. Широта доступа подразумевает самодокументированность веб-приложения и его предсказуемое поведение, обработка возможных ошибок ввода и информа-

ционная безопасность ресурса. За счет задержки, существующей при передаче данных в сети Интернет, страдает скорость работы приложения.

Простейшим способом распространения результатов имитации сигналов может считаться организация публично доступных наборов данных или же обычная публикация файлов в сети Интернет. В таком случае доступ к исходному коду возможен, только если разработчики сами разместят его, но и доступ к результатам неинтерактивен. Только опубликованные данные оказываются доступными посетителям набора данных. Возникает задача определения важности тех или иных данных и их структуризация. При этом обмен файлами является одним из традиционных применений сети Интернет, поэтому существует множество различных простых способов реализации этой функции.

Отдельные разработчики публикуют свои наборы сигналов на ресурсе GitHub, который обычно предназначен для контроля версий разрабатываемого программного обеспечения, но также успешно справляется с задачей публикации наборов данных. Например, набор коммуникационных сигналов для автоматической классификации модуляции, представленный на GitHub, включает одиннадцать типов сигналов модуляции одной несущей с различными шумами [16]. Он собран из сигналов с различными отношениями сигнал-шум на основе реальной географической среды. Набор данных содержит 22 000 выборок, и каждая выборка состоит из сигнала во временной области и метки. Изначально информация о наборе была опубликована в журнале *Sensors* с открытым доступом в 2018 г. в статье исследователей из Китая «Автоматическая классификация модуляции на основе глубокого обучения для беспилотных летательных аппаратов» [17].

Следующий пример – сжатый набор радарных сигналов *RadarCommDataset*, размещенный на Github. В статье “Multi-task Learning Approach for Automatic Modulation and Wireless Signal Classification”, в частности, говорится, что набор

данных о беспроводных сигналах выпущен для публичного использования в соответствии с лицензией некоммерческого использования [18, 19]. Отсутствие существующих многозадачных размеченных наборов данных для машинного обучения в области беспроводной связи и явилось основной причиной выпуска этого набора данных.

Публикация на ресурсе GitHub снимает с разработчиков задачу обеспечения бесперебойной работы файлового сервера и предоставляет ряд дополнительных преимуществ. Одним из преимуществ является контроль версий, который позволяет сохранять различные итерации разработки набора данных. При этом, конечно, GitHub, как универсальное решение, ограничен с точки зрения гибкости организации хранилища – не позволяет реализовать собственные способы структуризации и поиска данных. Другой негативной чертой является ограниченный набор доступных лицензий, который требует согласования со стандартами, принятыми в отдельных странах – в Российской Федерации в данном случае. Традиционным способом публикации данных является публикация на собственном сайте, где разработчики имеют большую творческую свободу по организации хранения и поиска данных. Примером является раздел публикации разнообразных наборов данных общества обработки сигналов IEEE, который показывает, что в большинстве ситуаций такая гибкость оказывается невостребованной, так как все ограничивается простой публикацией файлов с простой древовидной файловой иерархией и без специфических инструментов поиска, применяемых, например, в базах знаний [20].

Описанные рассуждения записаны в таблице 1 в рамках сравнительного анализа рассмотренных способов организации доступа с целью выбора способа, который следует применить в проекте, выполняемом авторами статьи.

В табл. 1 рассмотрены следующие характеристики: интерактивность, закрытость исходного кода, гибкость, сложность разработки и доступа,

Табл. 1

Сравнительный анализ подходов к организации доступа к результатам имитации

Подход, пример	Преимущества	Недостатки
Отдельное приложение 1D Fourier basis signal simulation system [4]	Интерактивность; Закрытый код; Гибкость.	Сложность разработки; Сложность доступа.
Веб-приложение UW cosmic spectra [10]; NIST atom spectra [11].	Интерактивность; Закрытый код; Гибкость; Легкий доступ.	Сложность разработки; Сетевые проблемы; Поддержка сервера; Безопасность.
Набор данных RadarCommDataset [16]	Закрытый код; Простота разработки; Простота доступа.	Нет интерактивности; Мало гибкости.

безопасность и, в случае веб-приложений, специфика веб-разработки. Интерактивность подразумевает взаимодействие с пользователем, открытость пользовательскому вводу. Гибкость подразумевает возможность точного подбора результатов под конкретную задачу.

Реализация метода имитации сигналов в любом случае приведет к разработке отдельного приложения. Такое приложение отвечает критериям интерактивности и гибкости, защищает исходный код. В промышленных сферах применения имитации сигналов веб-приложения нецелесообразны по причине наличия задержки в сети Интернет. Даже при игнорировании задержки в исследовательских задачах сложности при разработке веб-приложения не оправдывают преимуществ по сравнению с отдельным приложением. Разработка отдельного приложения должна быть намечена на будущее, однако соответствующие трудозатраты могут быть чрезмерными на исследовательском этапе. Вариант набора данных позволяет обмениваться результатами имитации разными методами, элементами исходного кода, сравнивать эффективность разных подходов.

В результате сравнительного анализа было принято решение публиковать данные научных экспериментов в виде набора данных на ресурсе GitHub. Подготовка такого набора предполагает структурирование имеющихся данных, что также необходимо и при разработке объединяющего приложения в его отдельном или Интернет представлении.

2. Структура данных

Как указано выше, при построении набора данных важным становится структура набора – данные должны быть организованы в древовидную иерархию. Используются данные, необходимые для описания процесса гармонической имитации и получаемые в результате его выполнения:

- формулы, задающие функцию спектральной плотности энергии (ФСПЭ);
- исходные параметры: параметры дискретизации N_1 , N_2 ; периоды сигнала T_1 , T_2 ; граничные частоты ω_1 , ω_2 ;
- графики и таблицы значений полученных сигналов, ФСПЭ, теоретической, алгоритмической и экспериментальной автокорреляционных характеристик.

Цель публикации данных – сравнение результатов имитации на одинаковых и сходных параметрах исследователями, обзор результатов имитации третьими лицами, подготовка основы

для дальнейшего построения онтологии на наборе данных.

Структурирование данных проводится по экспериментам – папка эксперимента с конкретными начальными значениями содержит соответствующие данные. Параметры эксперимента решено вынести в имена папок для упрощения навигации. Фрагмент полученной иерархии показан на рис. 1.

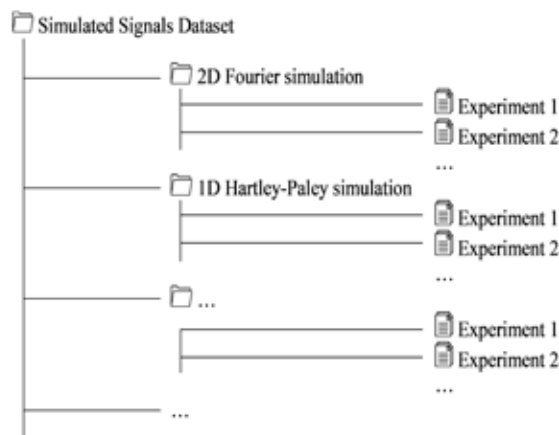


Рис. 1. Структура набора данных симитированных сигналов

Описанная на рис. 1 структура достаточна для формирования и публикации набора данных и отвечает поставленной цели публикации данных. В целях гомогенизации данных результаты экспериментов решено представить в виде таблиц, содержащих значения, формата «.csv» и рисунков различных форматов. При публикации рисунков приоритет предоставляется векторным рисункам формата «.svg», что мотивировано перспективой применения рисунков в научной деятельности и их публикации, так как векторные рисунки выигрывают по качеству. Рисунки и таблицы располагаются в папках, описывающих эксперименты. Исходные параметры каждого эксперимента указываются в именах соответствующих папок с целью организации автоматизированного поиска. Параметры ω_1 , ω_2 могут быть записаны как w_1 , w_2 для упрощения дальнейшей автоматизации, так как английский алфавит доступнее в средах программирования. Папки экспериментов в свою очередь организуются в разделы, посвященные конкретным методам имитации.

Для организации онтологии на наборе данных предложенной структуры предлагается добавить сущность «задача», описывающую конкретную задачу реального мира [21]. Отношение между сущностями «метод», «эксперимент» и «задача» показано на рис. 2.



Рис. 2. Взаимоотношение сущностей в онтологии

Сущность «задача» связана с экспериментами, так как эксперименты принимают на вход параметры, соответствующие конкретным задачам реального мира. Также сущность «задача» связана с методами, так как конкретные методы оказываются наиболее полезными при решении определенных задач. Включение сущности «задача» выводит хранилище за пределы обычного набора данных, но позволяет автоматизировать подбор методов имитации под конкретные задачи. Автоматизированный подбор состоит из следующих этапов:

- сопоставление экспериментов и методов путем сравнения значений автокорреляционных функций, определяющего точность эксперимента имитации сигнала;
- сопоставление экспериментов и задач путем вычисления вероятностей значений входных параметров имитации для каждой задачи;
- формирование правила вывода или ассоциации, связывающей задачу и метод, основываясь на предыдущих сравнениях, в целях поддержки

решения по применению отдельного метода для решения конкретной задачи.

Обработка больших объемов данных может приводить к необходимости применения методов работы с большими данными. Первые два шага направлены на создание экспертной подсистемы. Формирование правил вывода, характерных для онтологий, должно учитывать возможные неопределенности входных данных в случаях новых задач или ранее не проводимых экспериментов. Законченное хранилище должно представлять собой гибридную интеллектуальную систему, основанную на онтологии, объединяющей экспертные подсистемы.

3. Результаты

Формирование репозитория под названием Simulated-Signals-Dataset произведено на ресурсе GitHub [22]. Итоговый вид верхнего каталога репозитория показан на рис. 3.

Репозиторий содержит разделы, соответствующие трем методам имитации сигналов. Создана отдельная папка для хранения имеющихся в научной команде свидетельств государственной регистрации разработанных программных продуктов. Также имеется раздел, посвященный отдельным фрагментам математической составляющей метода имитации. На данный момент в репозитории наиболее полно представлены результаты экспериментов по имитации двумерных сигналов в базисе Фурье, расположенные в разделе “2D Fourier”. Содержимое раздела показано на рисунке 4. Название каждой папки эксперимента содержит значения всех входных параметров, описывающих этот эксперимент.

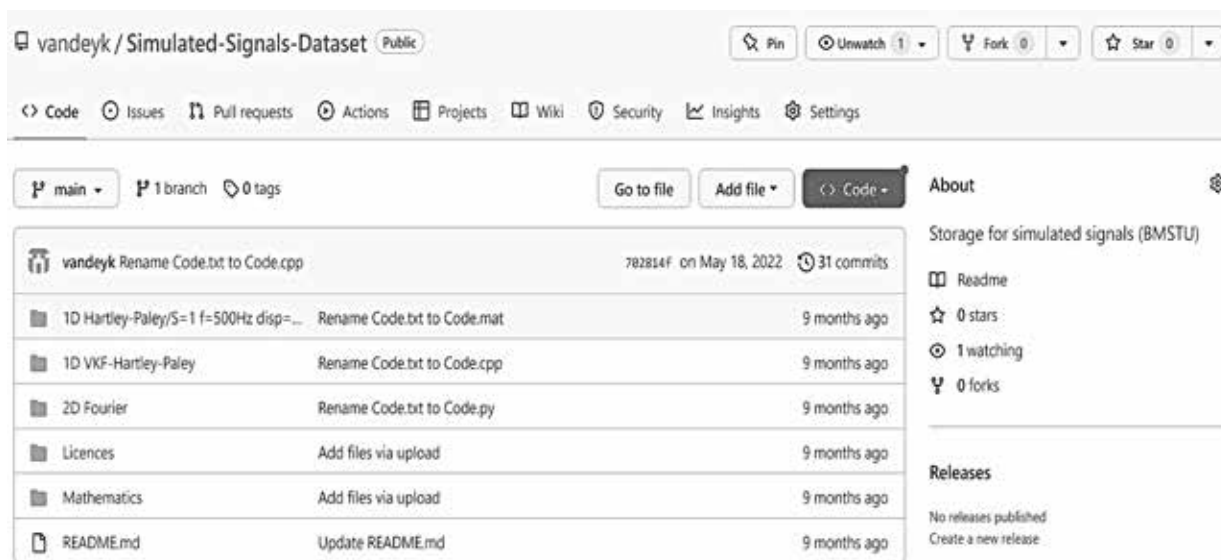


Рис. 3. Набор симитированных сигналов на GitHub

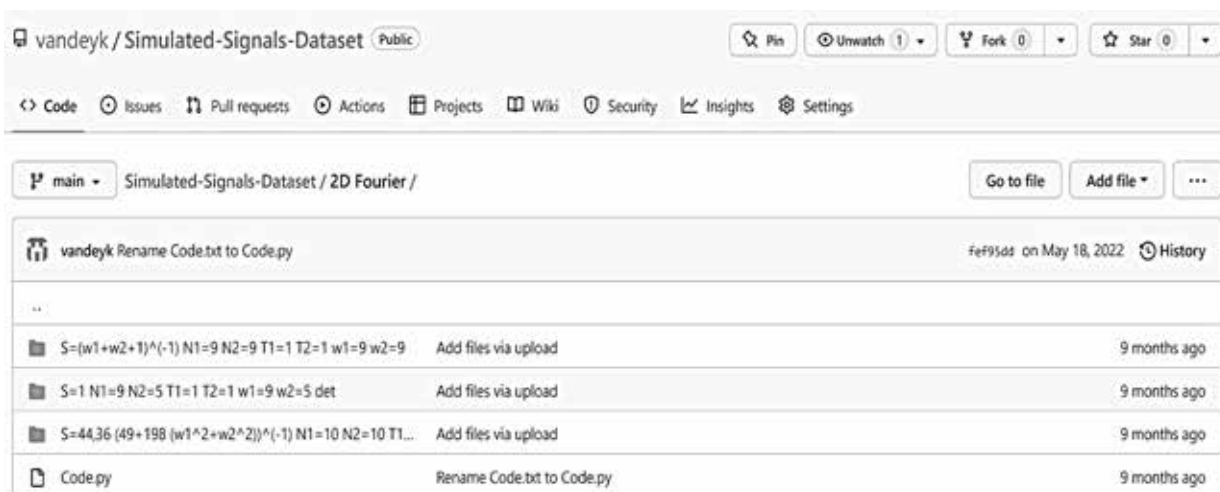


Рис. 4. Эксперименты по имитации в двумерном базисе Фурье в наборе данных

В разделах, посвященных отдельным методам, хранятся кроме экспериментов фрагменты программного кода, реализующего эти методы. Планируется расширение имеющегося раздела репозитория, а также добавление других методов имитации и их результатов.

Заключение

В рамках проекта с помощью разработанного ПО были получены результаты имитации сигналов, сформированные наборы сигналов были опубликованы на GitHub [22]. В результате работы научная команда и третьи лица могут получить доступ к научным результатам. Статья описывает процесс выбора подхода к организации доступа, процесс построения структуры хранилища и сам набор данных. Приведены принципы построения онтологии на основе созданного набора данных. В будущем планируется расширить и усовершенствовать набор данных, дополнить его новыми результатами, разработать отдельное приложение для работы с различными методами гармонической имитации сигналов. Будет проводиться разработка онтологии на основе существующего набора данных.

Литература

1. Smirnova, E., Syuzev, V., Gurenko, V., Alekhin, V. Software system's usage for multidimensional signal's simulation as an engineering staff training tool. INTED2020, pp. 6270-6279. 2020.
2. Суятинов, С. И. Синергетическая модель ситуационной осведомленности человека-оператора в эргатических системах управления подвижными объектами / С. И. Суятинов, Т. И. Булдакова, Ю. А. Вишневская // Мехатроника, автоматизация, управление. – 2022. – Т. 23. – № 6. – С. 302-308. – DOI 10.17587/mau.23.302-308.
3. Liu, Y., Li, J., Sun, S., Yu, B. Advances in Gaussian random field generation: A review. Computational Geosciences. 2019.
4. Сюев В.В. Алгоритмы многомерного имитационного моделирования случайных процессов / В. В. Сюев, Е. В. Смирнова, А. В. Пролетарский // Компьютерная оптика. – 2021. – Т. 45. – № 4. – С. 627-637. – DOI 10.18287/2412-6179-CO-770.
5. База Федерального Института Промышленной Собственности [сайт]: свидетельство о регистрации программы для ЭВМ RU 2020610822. URL: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2020610822&TypeFile=html (дата обращения: 01.06.2022).
6. База Федерального Института Промышленной Собственности [сайт]: свидетельство о регистрации программы для ЭВМ RU 2019610689. URL: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2019610689&TypeFile=html (дата обращения: 01.06.2022).
7. База Федерального Института Промышленной Собственности [сайт]: свидетельство о регистрации программы для ЭВМ RU 2017619635. URL: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2017619635&TypeFile=html (дата обращения: 01.06.2022).
8. База Федерального Института Промышленной Собственности [сайт]: свидетельство

- о регистрации программы для ЭВМ RU 2017619554. URL: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2017619554&TypeFile=html (дата обращения: 01.06.2022).
9. Black, P., Flater, D., Bojanova, I. Algorithms and Data Structures for New Models of Computation. IT Professional. 23. 9-15. 2021.
 10. Alan, I., Arslan, E., Kosar, T. Energy-aware data transfer algorithms. The International Conference for High Performance Computing, Networking, Storage and Analysis. 1-12. 2015.
 11. Smirnova, E.V., Syuzev, V.V., Gurenko, V.V., Bychkov, B. Spectral signal simulation as a scientific and practical task in the training of engineers. INTED2019, pp. 4511-4516. 2019.
 12. Smirnova, E., Syuzev, V., Samarev, R., Deykin, I., Proletarsky, A. High-Dimensional Simulation Processes in New Energy Theory: Experimental Research (Extended Abstract). Data Analytics and Management in Data Intensive Domains : Extended Abstracts of the XXII International Conference DAMDID, pp. 160-163. Voronezh State University, Voronezh. 2020.
 13. University of Washington [сайт]: spectral databases and tools. URL: <http://depts.washington.edu/naivpl/content/spectral-databases-and-tools> (дата обращения 31.05.2022).
 14. National Institute of Standards and Technology [сайт]: atomic spectra database. URL: <https://www.nist.gov/pml/atomic-spectra-database> (дата обращения 31.05.2022).
 15. Beg, M., Belin, J., Kluver, T., Konovalov, A., Ragan-Kelley, M., Thiéry, N., Fangohr, H. Using Jupyter for Reproducible Scientific Workflows. Computing in Science & Engineering, 1-11. 2021.
 16. Github [сайт]: набор данных «Communication Signal Dataset». URL: <https://github.com/bczhangbczhang/Communication-Signal-Dataset> (дата обращения 31.05.2022).
 17. Zhang, D., Ding, W., Zhang, B., Xie, C., Li, H., Liu, C., Han, J. Automatic Modulation Classification Based on Deep Learning for Unmanned Aerial Vehicles. Sensors. 18, 924. 2018.
 18. Jagannath, A., Jagannath, J. Multi-task Learning Approach for Modulation and Wireless Signal Classification. ICC2021, pp. 1-7. IEEE, New Jersey. 2021.
 19. Github [сайт]: набор данных «Radar and communication signal dataset and modulation classification dataset». URL: <https://github.com/ANDROComputationalSolutions/RadarCommDataset> (дата обращения 31.05.2022).
 20. Signal processing society [сайт]: dataset resources. URL: <https://signalprocessingsociety.org/publications-resources/dataset-resources> (дата обращения 31.05.2022).
 21. Skvortsov, N., Stupnikov, S. Managing Data-Intensive Research Problem-Solving Lifecycle. In: Sychev, A., Makhortov, S., Thalheim, B. (eds) Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2020. Communications in Computer and Information Science, vol 1427, pp. 29 – 33. Springer, Cham. 2021.
 22. Github [сайт]: BMSTU simulated signals dataset. URL: <https://github.com/vandeyk/Simulated-Signals-Dataset> (дата обращения 31.05.2022).
 23. Российский научный фонд [сайт]: карточка проекта № 22-11-00049 «Разработка корреляционной теории моделирования многомерных сигналов и процессов в гибридных системах искусственного интеллекта реального времени». URL: <https://rscf.ru/project/22-11-00049/> (дата обращения 5.09.2022).

Дейкин Иван Игоревич. Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)». Аспирант. Количество печатных работ 10. Область научных интересов: цифровая обработка сигналов. E-mail: deykinii@student.bmstu.ru (ответственный за переписку).

Сюзев Владимир Васильевич. Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)». Доктор технических наук, профессор кафедры ИУ6. Количество печатных работ: более 120 (в т.ч. 5 монографий). Область научных интересов: цифровая обработка сигналов, многомерные структуры, разработка новой энергетической теории математического представления и преобразования моделей, сигналов и процессов в системах управления динамическими объектами на основе имитационного моделирования сигналов. E-mail: k_iu6@bmstu.ru

Смирнова Елена Валентиновна. Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)». Доктор технических наук, профессор кафедры ИУ6. Количество печатных работ: более 80 (в т.ч. 3 монографии). Область научных интересов: цифровая обработка биомедицинских сигналов, интернет-программирование, качество инженерного образования. E-mail: evsmirnova@bmstu.ru

Пролетарский Андрей Викторович. Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)». Доктор технических наук, профессор кафедры ИУ6, заведующий кафедрой ИУ6, декан факультета ИУ. Количество печатных работ: более 200 (в т.ч. 5 монографий). Область научных интересов: интеллектуальные системы и искусственный интеллект, мониторинг и управление динамическими системами, большие данные. E-mail: pav@bmstu.ru

Simulated reference signals storage: basis for fast algorithms creation

I.I. Deykin, V.V. Syuzev, E.V. Smirnova, A.V. Proletarsky

Bauman Moscow State Technical University, Moscow, Russia

Abstract. The paper is devoted to organizing a storage structure for multidimensional reference signals and processes within the framework of digital filtering based on fast simulation algorithms and a universal adaptive matrix transformation operator. The storage contains relevant simulation equations, simulated signals as value tables and images, and original code in Python, C++ and other languages, as well as a supplementary license data. Different approaches to storage creation are analyzed in the paper. The chosen route consists of structuring all the data into a dataset, publishing the dataset on GitHub and building an ontology based upon the dataset. The principles for structuring the data and organizing such an ontology are given. The resulting published dataset is presented. The work is supported by the Russian Scientific Fund (Project #22-11-00049).

Keywords: *data Storage, Signal Storage, Multidimensional Signals, Signal Simulation, Digital Signal Processing, DSP, Ontology, Harmonic Methods.*

DOI: 10.14357/20790279230105

References

1. Smirnova, E., Syuzev, V., Gurenko, V., Alekhin, V. Software system's usage for multidimensional signal's simulation as an engineering staff training tool. INTED2020, pp. 6270-6279. 2020.
2. Suyatinov, S., Buldakova, T., Vishnevskaya, Y. Building a model of situational awareness of a human operator based on the principles of synergetics. Mathematical Methods in Technologies and Technics 9, 92-96. 2021.
3. Liu, Y., Li, J., Sun, S., Yu, B. Advances in Gaussian random field generation: A review. Computational Geosciences. 2019.
4. Syuzev, V.V., Smirnova, E.V., Proletarsky, A.V. Algorithms of multidimensional random process simulation. Computer Optics 45(4), 627–637. 2021.
5. Computer program registration certificate RU 2020610822 at the Federal Institute of Industrial Property database. Available at: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2020610822&TypeFile=html (accessed June 1, 2022)
6. Computer program registration certificate RU 2019610689 at the Federal Institute of Industrial Property database. Available at: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2019610689&TypeFile=html (accessed June 1, 2022)
7. Computer program registration certificate RU 2017619635 at the Federal Institute of Industrial Property database. Available at: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2017619635&TypeFile=html (accessed June 1, 2022)
8. Computer program registration certificate RU 2017619554 at the Federal Institute of Industrial Property database. Available at: https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2017619554&TypeFile=html (accessed June 1, 2022)
9. Black, P., Flater, D., Bojanova, I. Algorithms and Data Structures for New Models of Computation. IT Professional. 23. 9-15. 2021.

10. *Alan, I., Arslan, E., Kosar, T.* Energy-aware data transfer algorithms. The International Conference for High Performance Computing, Networking, Storage and Analysis. 1-12. 2015.
11. *Smirnova, E.V., Syuzev, V.V., Gurenko, V.V., Bychkov, B.* Spectral signal simulation as a scientific and practical task in the training of engineers. INTED2019, pp. 4511-4516. 2019.
12. *Smirnova, E., Syuzev, V., Samarev, R., Deykin, I., Proletarsky, A.* High-Dimensional Simulation Processes in New Energy Theory: Experimental Research (Extended Abstract). Data Analytics and Management in Data Intensive Domains : Extended Abstracts of the XXII International Conference DAMDID, pp. 160-163. Voronezh State University, Voronezh. 2020.
13. University of Washington spectral databases and tools. Available at: <http://depts.washington.edu/naivpl/content/spectral-databases-and-tools> (accessed May 31, 2022).
14. National Institute of Standards and Technology atomic spectra database. Available at: <https://www.nist.gov/pml/atomic-spectra-database> (accessed May 31, 2022).
15. *Beg, M., Belin, J., Klyuyver, T., Konovalov, A., Ragan-Kelley, M., Thiéry, N., Fangohr, H.* Using Jupyter for Reproducible Scientific Workflows. Computing in Science & Engineering, 1-11 2021.
16. Communication Signal Dataset on Github. Available at: <https://github.com/bczhangbczhang/Communication-Signal-Dataset> (accessed May 31, 2022).
17. *Zhang, D., Ding, W., Zhang, B., Xie, C., Li, H., Liu, C., Han, J.* Automatic Modulation Classification Based on Deep Learning for Unmanned Aerial Vehicles. Sensors. 18, 924. 2018.
18. *Jagannath, A., Jagannath, J.* Multi-task Learning Approach for Modulation and Wireless Signal Classification. ICC2021, pp. 1–7. IEEE, New Jersey. 2021.
19. Radar and communication signal dataset and modulation classification dataset on GitHub. Available at: <https://github.com/ANDROComputationalSolutions/RadarCommDataset> (accessed May 31, 2022).
20. Signal processing society dataset resources. Available at: <https://signalprocessingsociety.org/publications-resources/dataset-resources> (accessed May 31, 2022).
21. *Skvortsov, N., Stupnikov, S.* Managing Data-Intensive Research Problem-Solving Lifecycle. In: Sychev, A., Makhortov, S., Thalheim, B. (eds) Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2020. Communications in Computer and Information Science, vol 1427, pp. 29 – 33. Springer, Cham. 2021.
22. BMSTU simulated signals dataset on GitHub. Available at: <https://github.com/vandeyk/Simulated-Signals-Dataset> (accessed May 31, 2022).
23. Information about the project “Correlation theory creation for multidimensional signals and processes simulation in real-time hybrid artificial intelligence system” on Russian Science Foundation web-site. Available at: <https://rscf.ru/project/22-11-00049/> (accessed September 5, 2022).

Deykin I.I. Post-graduate student, Bauman Moscow State Technical University, 2-nd Baumanskaya, 5, Moscow, 105005, Russia, e-mail: deykinii@mail.ru.

Syuzev V.V. Doctor of technical sciences, full professor, Bauman Moscow State Technical University, 2 - n d Baumanskaya, 5, Moscow, 105005, Russia, e-mail: k_iu6@bmstu.ru

Smirnova E.V. Doctor of technical sciences, full professor, Bauman Moscow State Technical University, 2-nd Baumanskaya, 5, Moscow, 105005, Russia, e-mail: evsmirnova@bmstu.ru

Proletarsky A.V. Head of the Scientific and Educational Complex “Informatics and Control Systems”, head of Computer Systems and Networks department, doctor of technical sciences, full professor, Bauman Moscow State Technical University, 2-nd Baumanskaya, 5, Moscow, 105005, Russia, e-mail: pav@bmstu.ru