

## **ГЕНЕРАЦИЯ ДАННЫХ ПАССАЖИРОПОТОКА МЕТОДОМ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ**

**Белоус Ф.В.<sup>1</sup>**

**Научный руководитель - кандидат физико-математических наук, доцент ВШ ЦК**

**Графеева Н.Г.<sup>1</sup>**

<sup>1</sup>НИУ ИМТО

fvdelows@mail.ru

### **Введение**

Разработка и тестирование современных аналитических и оптимизационных систем для пассажирского транспорта требуют объёмных и правдоподобных наборов данных. Реальные операционные данные часто недоступны из-за правовых ограничений, коммерческой тайны или их чувствительного характера, что тормозит исследования и разработку новых решений в области интеллектуальных систем SmartCity. При этом открытые данные расписаний перевозок описывают лишь идеализированную статичную картину, не отражая реальных стохастических отклонений и временных зависимостей.

Имитационное моделирование позволяет заполнить эту брешь, создавая синтетические данные, которые наследуют структурную корректность расписаний, но обогащаются статистическими свойствами реальных исторических записей. Зарубежный опыт в этой области развивается по трём направлениям: статистические модели, генеративно-состязательные сети (GAN) и имитационное моделирование [1, 2]. В России успешным примером применения имитационного подхода является проект «Демобазы 2.0» [3], генерирующий семантически согласованные данные авиаперевозок, что служит концептуальным ориентиром для настоящей работы.

Актуальность исследования обусловлена перспективностью имитационного моделирования как метода создания синтетических баз данных с гарантией соблюдения семантических ограничений системы: ссылочную целостность и временную согласованность данных, – которая заложена в самой архитектуре моделирующей системы (генератора) [2].

### **Основная часть**

Предлагаемый метод генерации данных основывается на дискретно-событийном имитационном моделировании. Архитектура генератора включает статическую модель предметной области (структура сущностей и связей) и стохастические распределения, описывающие вариативность поведения системы во времени. Основой системы являются конечный автомат, управляющий переходами рейсов между состояниями, и очередь событий с приоритетом, обеспечивающая корректную хронологическую последовательность генерируемых записей. Благодаря этому имитационный подход гарантирует соблюдение семантических ограничений, в отличие от машинного обучения [1].

Также важным преимуществом является то, что метод не требует больших объёмов исторических данных для обучения нейросети: достаточно выборочных статистических характеристик и правильной настройки, чтобы, – что как нельзя лучше подходит для областей с ограниченным доступом к реальным данным. Кроме того, имитационный подход, в отличие от GAN, не запоминает реальные записи, предоставляя естественные гарантии приватности.

## **Выводы**

Разработанная методика позволяет создавать реалистичные демонстрационные наборы данных для отладки аналитических и оптимизационных систем. Универсальность подхода открывает возможности для его применения в других предметных областях в качестве основы для создания цифровых двойников. Практические результаты работы, которые будут представлены в докладе, подтверждают состоятельность метода и его ценность. Дальнейшие исследования будут направлены на усложнение стохастических моделей и интеграцию с пуассоновскими потоками событий для моделирования случайного характера событий из реальных данных.

## **Литература**

1. Challagundla R. et al. Synthetic tabular data generation: A comparative survey for modern techniques //arXiv preprint arXiv:2507.11590. – 2025.
2. Lautrup A. D. et al. Systematic review of generative modelling tools and utility metrics for fully synthetic tabular data //ACM Computing Surveys. – 2024. – Т. 57. – №. 4. – С. 1-38.
3. Демобазы 2.0 для PostgreSQL: правдоподобные данные для разработки и тестирования // Habr.com: [сайт]. – URL: <https://habr.com/ru/companies/postgrespro/articles/956096/> (дата обращения: 09.02.2026).