

УДК 519.226.3

**Поиск аномальных значений в базах промышленных данных
методами вероятностного моделирования**

Рубин И.М. (ИТМО)

**Научный руководитель – кандидат технических наук, доцент Калюжная А.В.
(ИТМО)**

Введение. В современной промышленности и научных исследованиях, в том числе и в области метеорологических наблюдений, обнаружение аномалий в данных становится все более значимым аспектом. Это связано с тем, что данные играют ключевую роль в обеспечении безопасности, эффективности и надежности различных процессов [1]. Причины аномалий в данных могут быть разнообразными [2]. Естественные факторы, такие как погодные условия, изменения климата и сезонные вариации, могут вызвать аномалии в метеорологических данных. Технические проблемы, такие как сбои в оборудовании или ошибки в измерительных приборах, могут привести к аномалиям в производственных данных. Человеческий фактор также играет важную роль. Ошибки ввода данных, недостаточное обучение персонала или неправильные действия операторов могут привести к аномалиям в данных. Независимо от отрасли промышленности, мониторинг и анализ данных становятся неотъемлемой частью ежедневной работы для выявления нештатных ситуаций и принятия своевременных мер.

Основная часть. В рамках исследования используются методы вероятностного моделирования для решения задачи детекции аномалий в данных. Решаются следующие два типа задач:

- 1) Повышения значения метрик качества современных методов обнаружения аномалий в данных. Были рассмотрены современные алгоритмы детекции аномалий в табличных данных и их применение на метеорологических данных. Повысить метрику качества удастся с помощью добавления новых признаков в данные, сгенерированных с помощью Байесовских сетей, содержащие вероятности аномалии каждого признака исходных данных.
- 2) Детекция аномалий в данных на основе сравнения структур Байесовских сетей. Сравнение структур производится с помощью метрики SHD. Для сравнения берутся структуры, обученные на данных без аномалий, и структуры, обученные на проверяемых данных. Если различия существенны, можно предположить, что данные содержат аномалии.

Выводы. Разработан алгоритм детекции аномалий в данных с помощью вероятностного моделирования

Список использованных источников:

1. Methodology for Data Validation 1.0 / Marco Di Zio, Nadežda Fursova, Tjalling Gelsema, Sarah Gießing, Ugo Guarnera, Jūratė Petrauskienė, Lucas Quenselvon Kalben, Mauro Scanu, K.O. ten Bosch, Mark van der Loo, Kat-rin Walsdorfer; Essnet Validat Foundation, 2016. – С. 7-18.
2. Gao, J., Xie, C., Tao, C. (2016). Big Data Validation and Quality Assurance – Issues, Challenges, and Needs. В: 10th IEEE International Symposium on Service-Oriented System Engineering, Оксфорд, Великобритания.