

**ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ ДЛЯ ВЫБОРА ПЕРЕМЕННЫХ В ЗАДАЧАХ
НЕЦЕЛЕВОГО КАЧЕСТВЕННОГО ХИМИЧЕСКОГО АНАЛИЗА**

Станиславчук-Абовский Д.Б. (СПБГУ)

**Научный руководитель – доктор химических наук, профессор Кирсанов Д.О.
(СПБГУ)**

Современные инструментальные методы анализа генерируют огромные массивы данных, усложняя построение классификационных моделей в рамках нецелевого качественного анализа (например, при прогнозировании заболеваний по спектрам биообразцов). Одним из ключевых этапов такого анализа выступает отбор наиболее информативных переменных среди множества сигналов (спектральных, хроматографических и т.д.). Эволюционные алгоритмы [1] эффективно решают эту задачу, поскольку способны исследовать обширные комбинаторные пространства и учитывать сложные взаимодействия признаков. Однако типичные реализации этих алгоритмов используют случайную инициализацию популяции и неуправляемые операторы мутации, что нередко приводит к избыточности набора переменных и замедленной сходимости.

В работе предлагается новый эволюционный алгоритм VIP-GA, позволяющий преодолеть эти недостатки. Его отличительной особенностью является использование VIP-оценок (Variable Importance in Projection) признаков в качестве вероятностей при инициализации решений. В процессе оптимизации VIP-оценки пересчитываются и комбинируются с частотой нахождения признаков в лучших решениях, что позволяет эффективно управлять мутацией. Также реализована схема адаптации для поддержания разнообразия популяции.

Эффективность предложенного алгоритма оценивалась путем сравнения с классическими методами выбора переменных, такими как генетический алгоритм, метод роя частиц и дифференциальная эволюция, а также их современными гибридными модификациями. Были изучены несколько наборов данных: хромато-масс-спектрометрическое исследование образцов плазмы крови для поиска маркеров неонатального энтероколита, потенциометрическое мультисенсорное исследование образцов мочи для оценки возможности скрининга рака простаты, данные ближней инфракрасной спектроскопии от образцов кожи для установления возможности распознавания тканей с онкологическими патологиями. Результаты показали, что алгоритм VIP-GA формирует более компактное подмножество переменных и обеспечивает лучшую точность классификации по сравнению с известными эволюционными методами выбора признаков.

Список литературы

- [1] Sorochan Armstrong M.D., De La Mata A.P., Harynuk J.J. // Review of Variable Selection Methods for Discriminant-Type Problems in Chemometrics // Front. Anal. Sci. 2022. Vol. 2. P. 867938.