

## **ГИБРИДНЫЙ ПОДХОД К АНАЛИЗУ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ**

Статья посвящена исследованию применения методов эволюционных вычислений, в частности генетических алгоритмов и эволюционных стратегий, для решения ключевых задач анализа временных рядов. Рассматриваются теоретические основы ГА, генетического программирования и ЭС, включая их основные операторы и стратегии отбора. Основное внимание уделено практическому применению этих методов для автоматического подбора параметров прогнозных моделей оптимизации кластеризации временных данных и решения задачи обнаружения аномалий.

**КЛЮЧЕВЫЕ СЛОВА:** Временные ряды, генетический алгоритмы, генетическое программирование, прогнозирование, обнаружение аномалий, эволюционные стратегии

**Введение.** В современной науке и практике анализа данных особое внимание уделяется методам, способным решать сложные задачи оптимизации с высокой размерностью и множеством взаимосвязанных параметров. Одним из таких подходов являются генетические алгоритмы (ГА) — класс методов, вдохновлённых природными механизмами эволюции, включая естественный отбор, наследственность и генетические изменения. Генетические алгоритмы широко применяются для решения задач оптимизации, анализа и прогнозирования, где классические подходы могут столкнуться с проблемами, связанными с размерностью задачи или сложной структурой пространства решений.

Актуальность изучения и применения генетических алгоритмов высока в области анализа временных рядов. Временные ряды присутствуют во множестве областей человеческой деятельности: экономике, финансах, экологии, инженерных науках и информационных технологиях. Часто эти ряды имеют сложную структуру, характеризуются сезонностью, трендами и непредсказуемыми изменениями, что затрудняет применение традиционных

методов анализа. Генетические алгоритмы позволяют решать такие задачи благодаря своей способности адаптироваться, автоматически подбирать оптимальные параметры и модели, а также находить скрытые закономерности и зависимости в данных.

**Основы генетических алгоритмов.** Первые идеи, лежащие в основе генетических алгоритмов, были предложены Джоном Холландом в 1960-х годах. Холланд стремился создать методы, которые могли бы решать задачи оптимизации, имитируя процессы естественного отбора, наследования и изменчивости. Его работы заложили фундамент направления, получившего впоследствии широкое распространение в различных областях знаний.

Генетические алгоритмы основаны на трёх основных операторах: селекции (отбора), скрещивания (кроссовера) и мутации. Рабочий цикл классического ГА выглядит следующим образом:

- Инициализация: генерируется начальная случайная популяция возможных решений;
- Оценка приспособленности: каждому решению ставится в соответствие значение функции приспособленности, характеризующее его качество;
- Селекция: лучшие решения выбираются для создания следующего поколения;
- Кроссовер (скрещивание): создаются новые решения путём обмена генетической информацией между парами решений;
- Мутация: небольшие случайные изменения, обеспечивающие разнообразие решений и поиск новых областей пространства решений.

Генетическое программирование (ГП) – это разновидность генетических алгоритмов, направленная на автоматическую генерацию компьютерных программ и математических выражений. В отличие от классического ГА, где решения представлены фиксированными строками или последовательностями, в генетическом программировании решения представлены в виде древовидных структур, где внутренние узлы являются

функциями или операторами, а листья – терминальными элементами (переменные, константы).

Каждое дерево в генетическом программировании является отдельной программой, которую можно выполнить и оценить по заданной метрике качества. Применяя операторы эволюции (скрещивание и мутация) к древовидным структурам, ГП позволяет постепенно улучшать качество решений.

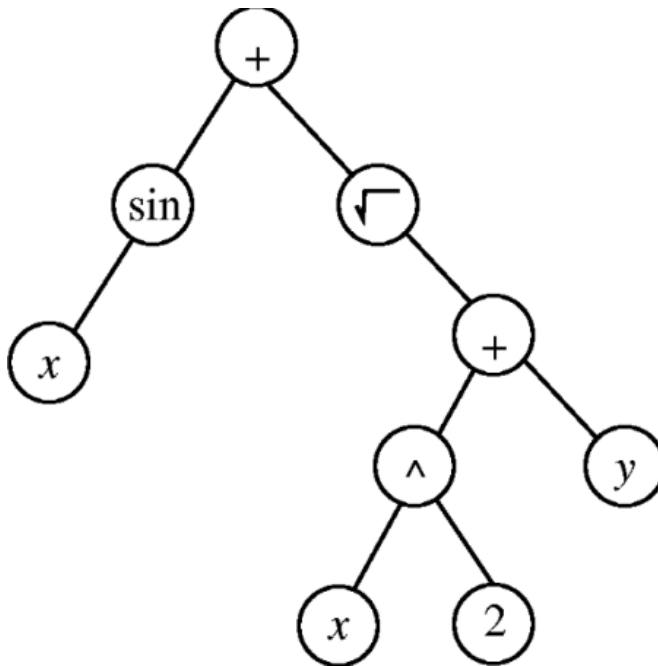


Рис. 1 – Пример древа ГП в задаче символьной регрессии

Эволюционные стратегии (ЭС) представляют собой отдельный класс методов эволюционных вычислений, предназначенных преимущественно для решения задач оптимизации с непрерывными параметрами. Изначально предложенные в 1960-е годы Ингеборгом Рехенбергом и Хансом-Полем Швэфелем, эти методы основаны на механизмах естественного отбора, аналогично другим эволюционным алгоритмам, но отличаются специфическими подходами к генерации новых решений и выбору лучших из них.

Центральная идея эволюционных стратегий заключается в том, что решения представлены вещественными векторами, а улучшение решений достигается за счёт применения оператора мутации, реализующего небольшие изменения текущих решений. Отличительной особенностью ЭС является механизм самоадаптации параметров мутации, благодаря чему алгоритм самостоятельно изменяет масштаб поиска в зависимости от текущей стадии оптимизации.

Существуют две основные стратегии отбора в ЭС:

- $(\mu, \lambda)$ -стратегия: из  $\lambda$  потомков нового поколения отбираются  $\mu$  лучших решений, родительские решения полностью отбрасываются. Данный подход способствует высокой скорости поиска новых решений и быстрому реагированию на изменения в данных, однако может привести к потере полезных решений, найденных ранее;
- $(\mu + \lambda)$ -стратегия: в данном подходе родительские решения объединяются с потомками в единый набор, после чего из этого объединённого множества выбираются  $\mu$  лучших решений. Подобная стратегия обеспечивает большую стабильность и сохраняет найденные хорошие решения, однако сходимость алгоритма может быть медленнее;

**Применение в анализе временных рядов.** Эволюционные стратегии являются готовым, мощным инструментом, применимым для работы со временными рядами. Временные ряды представляют собой последовательности данных, измеренных последовательно во времени, такие как показатели экономической активности, температура воздуха, спрос на продукцию и др. Анализ и прогнозирование таких данных имеют большое значение в таких областях деятельности человека как:

- Экономика и финансы – для оптимизации параметров моделей предсказания значений временного ряда;
- В кибербезопасности – для автоматической настройки моделей поиска аномалий

- Электронная торговля – для настройки параметров моделей рекомендации.

Задачи анализа временных рядов включают выявление трендов, сезонных колебаний, циклов и случайных факторов, а также построение моделей для прогнозирования будущих значений.

**Оптимизация моделей прогнозирования.** Одной из ключевых задач анализа временных рядов является выбор подходящей модели и подбор её оптимальных параметров. Генетические алгоритмы успешно используются для решения таких задач, автоматически подбирая параметры моделей прогнозирования, таких как:

- Модели ARIMA (авторегрессионные интегрированные модели скользящего среднего);
- Нейронные сети (LSTM, GRU);
- Модели экспоненциального сглаживания.

ГА позволяют автоматически и эффективно исследовать большое пространство возможных параметров, минимизируя ошибки прогнозирования.

Модель ARIMA зависит от трех параметров:

- Порядок авторегрессии
- Степень дифференцирования
- Порядок скользящего среднего

В виду конечности общего числа возможных решений, полная работа генетического алгоритма является перебором из  $n^3$  вариантов, где  $n$  – закрепленное максимальное значение параметров модели ARIMA.

Модели рекуррентных нейронных сетей применяются для прогнозирования значений временных рядов, исходя из доступных данных. Модели нейронных сетей, таких как LSTM, GRU обладают куда большим количеством переменных. С помощью моделей эволюционных стратегий возможно подобрать как топологию модели, так и ее параметры, такие как:

- Степень регуляризации;

- Функция активации;
- Скорость обучения;
- Количество эпох

Важно понимать, что исследование всего пространства решений в контексте подбора параметров нейронной сети – крайне сложная вычислительная задача, требующая колоссальных ресурсов. Для ее решения с применением генетических алгоритмов, на каждой итерации будет обучаться нейронная сеть, параметры которой соответствуют текущему поколению, предложенным эволюционной стратегией.

Генетические алгоритмы также широко применяются для оптимизации более простых моделей машинного обучения. Они могут быть применены для автоматического отбора признаков и определения оптимальных временных лагов, которые наиболее информативны для прогнозирования. Этот процесс включает следующие этапы:

- Формирование множества потенциальных признаков и лагов;
- Использование ГА для поиска оптимального подмножества признаков, минимизирующего прогнозную ошибку;
- Оценка качества полученных моделей и отбор лучших вариантов.

ГА эффективно применяются в задачах сегментации и кластеризации временных рядов. Это особенно актуально при работе с данными, имеющими различные режимы поведения:

- ГА могут оптимизировать параметры алгоритмов кластеризации, определяя оптимальное количество кластеров и их состав;
- Они помогают выявить периоды смены тренда и сезонных колебаний, улучшая интерпретацию и прогнозные свойства моделей.

**Использование в задаче поиска аномалий.** Помимо решения задачи предсказания временных рядов, эволюционные стратегии применимы и в обширной задаче обнаружения аномалий. К основным группам методов детектирования аномалий во временных рядах можно отнести следующие:

- статистические тесты и пороговые правила;

- декомпозиция на сезонность, тренд и остаток;
- методы кластеризации;
- модели изолирующего леса.

Генетические алгоритмы оказываются полезны в работе с каждой группой методов. Например, с их помощью можно модифицировать классический метод трех сигм. В таком случае, в качестве порогового значения берется не значение, равное трем стандартным отклонениям, а более оптимальное, найденное после исследования пространства решений алгоритмом.

Методы кластеризации и декомпозиции временного ряда включают в себя большое множество различных подходов. Генетический алгоритм помогает найти оптимальные параметры для окон сезонности, тренда и порогового значения остатка, эффективно исследуя дискретное пространство решений. Для кластеризации точек временного ряда активно применяются в том числе и модели машинного обучения. В этом случае область применения генетических алгоритмов смежна с аналогичной задачей применения методов машинного обучения для прогнозирования значений временного ряда. Отличие наблюдается лишь в оптимизируемых параметрах, зависящих от выбранной модели.

Модель изолирующего леса является крайне популярной модификацией деревьев решений для решения задачи поиска аномалий, что позволяет вынести ее в отдельную группу методов. Данный метод строит большое количество случайных деревьев, изолирующих точки. Чем короче путь к листу – тем аномальнее точка. В этом случае генетические алгоритмы используются для подбора гиперпараметров.

**Формализация задачи поиска аномалии.** Пусть  $X$  — вектор гиперпараметров модели детекции. Требуется найти  $\operatorname{argmax} F(X)$ , где  $F$  — заданная фитнес-функция (например, F1-score). ГА реализует стохастическую

оптимизацию этой задачи без расчёта градиента. Каждая хромосома, с которой работает генетический алгоритм, кодирует набор параметров:

- порог;
- окно сезонности для декомпозиции;
- параметры для изолирующего леса;
- весовые коэффициенты в комбинированной метрике.

Генетический алгоритм для решения задачи поиска аномалии может быть представлен в следующем виде:

1. Инициализация случайной популяции размером.
2. Вычисление пригодности для каждого индивида.
3. Селекция (турнир, рулетка).
4. Кроссовер (одноточечный, арифметический).
5. Мутация с вероятностью.
6. Замена старого поколения новым.
7. Проверка критерия останова (уменьшение ошибки или лимит поколений).

Параметры алгоритма подбираются экспериментально.

**Обзор преимуществ и недостатков.** К явным преимуществам применения генетических алгоритмов можно отнести:

- возможность работы с разнородными данными;
- способность исследовать сложные многомерные пространства решений;
- гибкость в комбинировании с другими методами.

Не менее явные недостатки включают в себя:

- высокая вычислительная сложность при больших популяциях;
- необходимость балансировать между скоростью и точностью;

**Заключение.** Таким образом, генетические алгоритмы и эволюционные стратегии представляют собой мощный и гибкий инструмент для решения широкого спектра задач анализа временных рядов. От

прогнозирования и кластеризации до обнаружения аномалий — эти методы позволяют эффективно исследовать сложные пространства параметров, автоматически находя оптимальные решения там, где традиционные подходы часто оказываются неэффективными.

#### **ЛИТЕРАТУРА:**

1. Холланд, Дж. Г. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
2. Голдберг, Д. Э. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.
3. Коза, Дж. Р. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press.
4. Zitzler, E., & Thiele, L. (1999). *Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach*. *IEEE Transactions on Evolutionary Computation*, 3(4), 257–271.
5. Бэк, Т. (1996). *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press.
6. Гамильтон, Дж.Д. *Time Series Analysis*. Princeton University Press, 1994. – 825 с.
7. Нильсен, Э. *Практический анализ временных рядов: Прогнозирование с использованием статистики и машинного обучения*. 2019. – 540 с.
8. Хиндман, Р.Дж., Афанасопулос, Г. *Прогнозирование: Принципы и практика (2-е изд.)*. 2018. – 380 с.