

# Генетический алгоритм SPEA для решения многокритериальной задачи формирования инвестиционных портфелей

Т. И. Романенкова, email: romanenkova.tr@gmail.com<sup>1</sup>

В. В. Коротков, email: chasecrunk@gmail.com<sup>1</sup>

<sup>1</sup> Воронежский государственный университет

***Аннотация.** В данной работе рассматриваются различные способы решения задач многокритериальной оптимизации, в особенности, генетический алгоритм SPEA, а также производится расчет оптимальных инвестиционных портфелей на базе криптовалюты.*

***Ключевые слова:** многокритериальная оптимизация, SPEA, алгоритмы, множество Парето.*

## Введение

При решении различного рода задач, достаточно часто появляется необходимость найти решения многокритериальных задач. В базовых реализациях оптимизации рассматриваются следующие подходы [1, 2]:

- Метод уступок;
- Метод главного критерия;
- Метод свертывания критерия

Их основным недостатком является упрощение модели и нахождение единственного решения.

При рассмотрении многокритериальной оптимизации, результатом получается некое множество решений, в котором ни одно не является оптимальным в сравнении с другим по всем критериям. Такое множество называется множеством оптимальных решений по Парето. Описанные выше решения не являются оптимальными по Парето.

Существенней учитываются параметры в более сложных алгоритмах. Одними из таких являются популяционные. Самым изученным и активно используемым является генетический алгоритм [3].

## 1. Вариации генетических алгоритмов

Генетический алгоритм [4, 5] часто используется в своей классической форме, но существует ряд его модификаций, значительно

улучшающих его работу. Среди таких вариаций можно выделить несколько основных:

- VEGA (Vector Evaluated Genetic Algorithm);
- FFGA (Fonseca and Fleming's Multi-Objective Genetic Algorithm) или его также называют MOGA (Multi-Objective Genetic Algorithm);
- NPGA (Niche Pareto genetic algorithm);
- SPEA (Strength Pareto Evolutionary Algorithm).

Каждый из этих алгоритмов имеет свои уникальные преимущества и недостатки.

В VEGA для каждого из критериев селекция производится отдельно. После, все полученные таким образом подпопуляции объединяются в одну, и для нее происходит скрещивание и мутация.

FFGA представляет собой процедуру ранжирования индивидов, основанную на Парето-доминировании. Ранг в данном алгоритме определяется для каждого индивида с помощью количества доминирующих его индивидов. Величина обратная рангу, определяет пригодность.

NPGA имеет свои особенности и не похож на вышеописанные методы. Его особенностью является способность поддерживать разнообразие. Он соединяет в себе два подхода: турнирную селекцию и концепцию доминирования по Парето. В NPGA пригодность рассчитывается как модифицированная схема деления, с использованием ниш.

Алгоритмы VEGA и FFGA не способны обеспечить равномерное покрытие множества Парето. Алгоритмы NPGA и SREA, рассмотренные ниже, способны обеспечить хорошее покрытие, но, чтобы это реализовать, им необходимо использовать значительные затраты на вычисления, и механизмы, обеспечивающие поддержание разнообразия, иногда пересекают область Парето.

## **2. Алгоритм SPEA**

Алгоритм SPEA многие специалисты выбирают как более оптимальный и универсальный.

Преимущества данного алгоритма заключается в том, что [6]:

- Он реализует все подходы, описанные ранее, в одном алгоритме;
- Пригодность определяется для каждого индивида относительно множества недоминируемых особей;
- Лучшие индивиды, хранящиеся в отдельном множестве, все равно участвуют в отборе;

- Ниши, в которых деление общей пригодности, происходит на основе Парето доминирования, позволяют обойти преждевременную сходимость;
- Так как размер множества недоминированных особей влияет на работу данного алгоритма, предусмотрено уменьшение его размера, с сохранением его основных характеристик.

В основе реализации средств решения многокритериальной оптимизации лежат принципы Парето, благодаря которым можно оценить качество решений одновременно по всем критериям. Результатом решения поставленной задачи является теоретически бесконечное множество Парето-оптимальных решений [7].

На данные момент разделяются три поколения генетических алгоритмов:

- Поколение 0. Суть заключается в том, что принципы Парето не используются. Это сильно отражается на отсутствии поддержания разнообразия популяции.
- Поколение I. Используются принципы Парето для оценивания решений. Применяются концепции ниш, а также используется разделяющая функция для поддержания разнообразия популяции.
- Поколение II. Сохраняются все особенности реализации предыдущего поколения. Особенностью является развитая реализация стратегии элитизма с помощью использования Парето-архивов и управлением плотностью решений вдоль границы Парето.

Одним из алгоритмов поколения II является SPEA. В настоящее время именно это поколение является самым продуктивным и актуальным [6].

Его особенность заключается в том, что происходит сохранение так называемых элитных хромосом в основной популяции. Остальные хромосомы располагаются в Парето-архиве. Там же они и обновляются, а при необходимости интегрируются в основную популяцию.

Элитными хромосомами во втором поколении понимаются хромосомы, которые соответствуют недоминируемым решениям.

Принцип работы алгоритма SPEA относительно не сложен. Он заключается в том, что на каждой итерации работы этого алгоритма недоминируемые хромосомы копируются в Парето-архив. После чего для каждого элемента вычисляется Парето-сила. Ее значение пропорционально числу хромосом из основной популяции, которые доминирует выбранная хромосома из Парето-архива (1).

$$S(c_i) = \frac{|\{c_j \mid c_j \in P \wedge c_i \succ c_j\}|}{n_p + 1}, \quad (1)$$

где  $S(c_i) \in [0,1)$  – Парето-сила хромосомы  $c_i$ ,  $i = 1, \dots, n_A$ , из Парето-архива  $A$  размером  $n_A$ .

Так же SPEA отличается тем, что приспособленность хромосом популяции определяется только в зависимости от значения разделяющей функции соответствующих хромосом Парето-архива.

Все хромосомы популяции и Парето-архива участвуют в операторе отбора. И еще одной особенностью является то, что для сокращения размера Парето-архива, без потери качества границы Парето, используются методы кластеризации.

Чтобы поддерживать разнообразие популяции и контроля равномерного распределения решений вдоль границы Парето используется следующий способ.

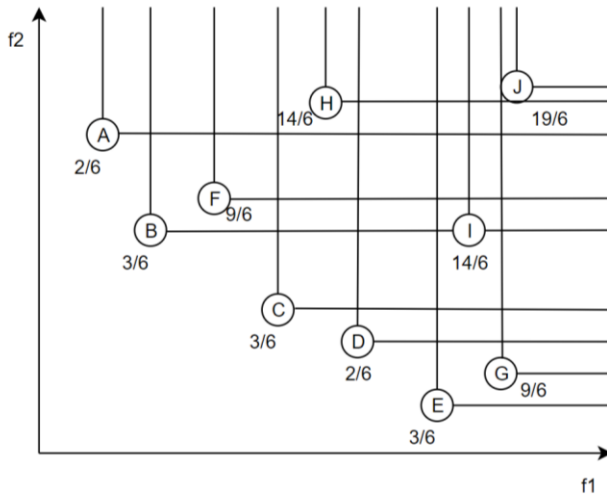


Рис. 1. Популяции.

На рисунке 1 размещены все возможные популяции. А, В, С, D и E – в своей совокупности образуют основную популяцию. Они же и лежат на границе Парето. F, G, H, I, J – в свою очередь образуют Парето-архив. Под каждым решением помечена приспособленность (Парето-сила) каждой соответствующей хромосомы. При этом координаты решений разделяют плоскость пространства критериев на прямоугольные

области. Разные решения могут принадлежать разному числу таких областей – чем их больше, тем больше решений доминируют данное. К примеру, решение  $J$  доминируется всеми остальными. Вместе с этим данные прямоугольной области могут рассматриваться как ниши. Принадлежность им определяется не с позиции некоторой меры близости, а на основе Парето-доминирования.

Если обобщать вышесказанное, то генетический алгоритм SPEA:

- Хранит недоминированные решения в основной популяции;
- Использует концепцию доминирования Парето для назначения скалярных значений пригодности для отдельных особей;
- Выполняет кластеризацию для уменьшения числа недоминированных решений, не разрушая Парето-фронт;
- Приспособленность особи определяется только от хранящихся в основной популяции;
- Все решения, находящиеся в основной популяции, участвуют в определении приспособленности;
- Поскольку метод основан на Парето, то для сохранения разнообразия в популяции не требуется какого-либо параметра расстояния.

### 3. Алгоритм работы SPEA

Порядок выполнения генетического алгоритма SPEA:

1. Сгенерировать начальную популяцию  $P_0$  и создать пустую основную популяцию  $P$  (недоминированное множество).
2. Скопировать все недоминированные особи  $P_0$  в  $P$ .
3. Удалить особи, которые повторяются в  $P$ .
4. Если количество особей в основной популяции  $P$  превышает заданный максимум  $N$ , то сократить  $P$ , используя кластеризацию.
5. Вычислить приспособленность каждой особи в  $P_0$  опираясь на гены, которые доминируют над генами особей  $P$ .
6. Выбрать особи из  $P_0 + P$ , пока не будет заполнен массив особей, участвующих в скрещивании. В данной теории рассматривается бинарный выбор скрещивания с заменой.
7. Применить скрещивание и мутации.
8. Если максимальное число генераций превышено, то остановиться, если нет, то вернуться к пункту 2.

Рассмотрим подробнее процедуру вычисления приспособленности. Она представляет собой двухэтапный процесс. Во-первых, особи

основной популяции  $P$  ранжируются. Во-вторых, особи  $P_0$  оцениваются:

1. Каждой особи из основной популяции  $P$  присваивается значение в диапазоне  $[0, 1)$ , называемое приспособленностью. Оно пропорционально числу членов популяции  $P_0$ , для которых выбранная особь полностью доминирует генами, над соответствующими генами особей из  $P_0$ . Пусть  $n$  – это число особей в  $P_0$ , над которыми доминирует особь  $i$  из  $P$ .  $N$  – это размер популяции  $P_0$ . Тогда приспособленность  $s_i$  определяется как (2)

$$s_i = \frac{n}{N + 1}, \quad (2)$$

2. Приспособленность особи  $j$  из популяции  $P_0$  рассчитывается путем суммирования всех внешних  $i$  из  $P$ , над которыми доминирует  $j$ . Необходимо добавить 1 к общему количеству, что бы гарантировать, что особи  $P$  будут иметь лучшую приспособленность, чем члены  $P_0$ . Необходимо также обратить внимание, что приспособленность должна быть сведена к минимуму, то есть маленькие значения соответствуют высокой вероятности выживания) (3), где  $[0, 1)$ .

$$f_j = 1 + \sum_{i, j \succ i} s_i, \quad (3)$$

Теперь разберем особенности объединения путем кластеризации. В некоторых задачах Парето-оптимальных особей может быть чрезвычайно много. Однако с точки зрения лица, принимающего решение, предоставление всех найденных недоминированных решений бесполезно, когда их число превышает разумные пределы. Кроме того, размер основной популяции влияет на поведение всей модели SPEA.

С одной стороны, поскольку слишком много недоминированных решений может снизиться производительность поиска оптимального решения. С другой стороны, если точки в основной популяции  $P$  не распределены равномерно, то метод назначения пригодности, вероятней всего смещен в сторону определенных областей пространства поиска, что приводит к несбалансированному распределению в популяции.

Метод, который используется для решения этой проблемы – это кластерный анализ. В общем случае, он разбивает набор элементов в

группы относительно однородных элементов, где количество кластеров меньше количества особей.

Рассмотрим метод средней связи – это поход кластеризации, который хорошо проявляет свои положительные качества в данной задаче:

1. Инициализировать набор кластеров  $C$ , каждая вершина – это недоминируемая особь  $i$  популяции  $P$  и все они образуют кластеры (4).

$$C = U_i \{ \{ i \} \}, \quad (4)$$

2. Если  $|C| \leq N$ , то перейти на пункт 5, иначе 3. Вычислите расстояние всех возможных пар кластеров. Расстояние  $d$  между кластерами  $c_1$  и  $c_2$  определяется, как среднее расстояние между парами особей в двух кластерах (5), где метрика  $\| \cdot \|$  отражает расстояние между двумя особями  $i_1 - i_2$ .

$$d = \frac{1}{|c_1| * |c_2|} * \sum_{i_1 \in c_1, i_2 \in c_2} \| i_1 - i_2 \|, \quad (5)$$

3. Найти два кластера  $c_1$  и  $c_2$  с минимальным расстоянием  $d$ . Выбранные кластеры объединяются в более крупный кластер (6). Перейти к пункту 2

$$C = C \{ c_1, c_2 \} \cup \{ c_1, c_2 \}, \quad (6)$$

4. Необходимо в каждом кластере вычислить единственное недоминируемое значение. Для этого используются центроид – это точка с минимальным средним расстоянием до всех других точек в кластере.

#### 4. Формирование оптимального инвестиционного портфеля на базе криптовалюты

Для расчета оптимальных инвестиционных портфелей необходимы данные входящих в него активов. В данном рассмотрении – это криптовалюта. На сегодняшний день существует множество ресурсов, которые предоставляют всю необходимую информацию. Для ниже приведенного расчета портфелей была взята информация с сайта [coingecko.com](http://coingecko.com).

Расчет оптимальных инвестиционных портфелей можно провести, опираясь на теорию Марковица [8]. В своих учениях он сформировал модель, в которой рассчитывается один портфель. Но это является упрощением модели, поэтому для дальнейших расчетов теория была переделана в многокритериальную задачу, для нахождения решения

которой и будет использоваться алгоритм SPEA. Для получения инвестиционных портфелей, необходимо рассчитать ожидаемую доходность и риск входящих в него криптовалют. Риск выражается дисперсией – разброс цен на актив. Другими словами, чем выше разброс котировок, то есть волатильность актива – тем больше риск инвестора. Будущая доходность – это величина случайная. Мы можем лишь предполагать какие значения она может принять. Для возможности ее использования наиболее вероятное ее значение можно выразить математическим ожиданием. Для его расчета можно использовать величины на основе средних значений доходности активов за определенный период. В таблице 1 приведены данные доходности и риска выбранных криптовалют.

Таблица 1

*Параметры криптовалют*

<b>Аббревиатура криптовалюты</b>	<b>Доходность</b>	<b>Риск</b>
AVAX	3.85	9.60
CRO	2.30	3.57
DOGE	-0.43	0.11
MATIC	0.09	0.07
SHIB	2.69	5.27
SOL	2.40	3.96
XLM	-0.22	0.05
USDC	0.08	0.02
WBTC	0.39	0.13

В таблице 2 полученные 9 портфелей с помощью применения алгоритма SPEA. В ней указаны полученные доли криптовалют, а также риск и доходность сформированного портфеля.

Доходность портфеля – это сумма произведения доли каждого актива на его доходность, а риск – это сумма произведения долей попарно взятых активов на ковариацию между ними.

Количество рассчитанных портфелей можно изменять размером популяции недоминированных особей.

Таблица 2

*Рассчитанные портфели*

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
AVAX	0.36	0.30	0.39	0.10	0.24	0.04	0.08	0.24	0.14



CRO	0.04	0.26	0.04	0.42	0.05	0.06	0.31	0.21	0.28
DOGE	0.05	0.04	0.06	0.07	0.23	0.09	0.05	0.10	0.31
MATIC	0.01	0.016	0.21	0.01	0.01	0.01	0.20	0.08	0.01
SHIB	0.08	0.14	0.18	0.23	0.23	0.18	0.05	0.07	0.15
SOL	0.23	0.00	0.00	0.00	0.16	0.40	0.23	0.00	0.00
XLM	0.03	0.03	0.02	0.04	0.02	0.05	0.03	0.24	0.03
USDC	0.08	0.03	0.03	0.04	0.02	0.05	0.01	0.02	0.03
WBTC	0.11	0.05	0.07	0.08	0.04	0.10	0.05	0.04	0.05
Риск	8.63	7.66	7.33	6.65	6.53	5.27	4.99	3.98	3.88
Доходность	2.27	2.13	2.08	1.98	1.95	1.75	1.71	1.52	1.49

### **Заключение**

Существует много способов для нахождения решений многокритериальных задач. Многие из них сводятся к тому, что параметры упрощаются и выискивается только 1 решение. Для нахождения большего количества решений часто прибегают к генетическому алгоритму. И на данный момент одной из самых эффективной его реализаций является SPEA.

Преимущества SPEA в том, что сохраняются все преимущества классических реализаций генетических алгоритмов. Также пригодность индивида рассчитывается относительно всего множества недоминируемых особей. Обираются лучшие индивиды в отдельную популяцию, которая так же продолжает участвовать в отборе. Он в свою очередь проводится на основе Парето доминирования, что позволяет избежать преждевременную сходимость.

Классическая модель Марковица для расчета инвестиционного портфеля легко преобразуется в многокритериальную задачу, с которой хорошо справляется рассмотренный алгоритм SPEA. С помощью него было получено 9 оптимальных инвестиционных портфелей криптовалют.

### **Список литературы**

1. Ногин, В. Д. Множество и принцип Парето: Учебное пособие / В. Д. Ногин. — СПб.: Издательско-полиграфическая ассоциация высших учебных заведений, 2020. — 100 с.
2. Zitzler, E. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach / E. Zitzler, L. Thiele // IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION. – VOL. 3, NO. 4, NOVEMBER 1999. – P. 257-271.

3. Казаков, П. В. Генетические алгоритмы многокритериальной оптимизации. Обзор / П. В. Казаков // Информационные технологии. – 2011. – 10(182). – С. 2-8.
4. Holland, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence* / J. H. Holland. — The MIT Press, Cambridge, 1992. – pp. 228.
5. Панченко, Т. В. Генетические алгоритмы: учебно-методическое пособие / Т. В. Панченко /под ред. Ю. Ю. Тарасевича. — Астрахань: Издательский дом «Астраханский университет», 2007. — 87 с.
6. Черноруцкий, И.Г. Методы оптимизации и принятия решений / И.Г. Черноруцкий. – СПб.: Лань, 2001. – 384с.
7. Пантелеев, А.В. Методы оптимизации. Практический курс / А.В. Пантелеев, Т.А. Летова. – М.: Логос, 2011. – 110с.
8. Markowitz, H. M. Portfolio Selection / H. M. Markowitz // *Journal of Finance*. – 1952. – 7. № 1 – pp. 71-91.