Содержание

[1. Введение 4](#_Toc514636795)

[1.1 Актуальность 4](#_Toc514636796)

[1.2 Цель работы 5](#_Toc514636797)

[1.3 Методы исследования 5](#_Toc514636798)

[2. Постановка задачи о рюкзаке 6](#_Toc514636799)

[2.1 Описание 6](#_Toc514636800)

[2.2 Варианты постановки задачи о рюкзаке 6](#_Toc514636801)

[2.3 Постановка задачи о 0-1 рюкзаке (ЗОР) 7](#_Toc514636802)

[2.3.1 Словесная 7](#_Toc514636803)

[2.3.2 Математическая 7](#_Toc514636804)

[2.4 Постановка задачи о неограниченном рюкзаке (ЗОНР) 7](#_Toc514636805)

[2.4.1 Словесная 7](#_Toc514636806)

[2.4.2 Математическая 8](#_Toc514636807)

[2.5 Свойства задачи о неограниченном рюкзаке 8](#_Toc514636808)

[2.5.1 Доминирование 8](#_Toc514636809)

[2.5.2 Периодичность 9](#_Toc514636810)

[2.6 Области применения задачи о рюкзаке 10](#_Toc514636811)

[2.6.1 Изучение в математике 10](#_Toc514636812)

[2.6.2 Изучение в экономике 10](#_Toc514636813)

[2.6.3 Изучение в криптографии 11](#_Toc514636814)

[3. Рассматриваемые алгоритмы решения задачи о рюкзаке 14](#_Toc514636815)

[3.1 Точные алгоритмы решения ЗОР 14](#_Toc514636816)

[3.1.1 Метод динамического программирования 14](#_Toc514636817)

[3.1.2 Метод ветвей и границ 15](#_Toc514636818)

[3.2 Точные алгоритмы решения ЗОНР 17](#_Toc514636819)

[3.2.1 Классический метод динамического программирования 17](#_Toc514636820)

[3.2.2 Метод эффективного динамического программирования 19](#_Toc514636821)

[3.2.3 Метод ветвей и границ с использованием верхней границы по трем предметам с лучшим соотношением цены и веса (U3) 21](#_Toc514636822)

[3.3 Генетический алгоритм 22](#_Toc514636823)

[4. Генетический алгоритм 24](#_Toc514636824)

[4.1 Интерпретация задачи о ранце и операторов генетического алгоритма c помощью понятий популяционной генетики 24](#_Toc514636825)

[4.2 Представление генотипа 25](#_Toc514636826)

[4.3 Операторы создания начальной популяции 26](#_Toc514636827)

[4.3.1 Алгоритм Данцига 26](#_Toc514636828)

[4.3.2 Случайный алгоритм 26](#_Toc514636829)

[4.4 Операторы скрещивания 26](#_Toc514636830)

[4.4.1 Одноточечный кроссовер 27](#_Toc514636831)

[4.4.2 Двуточечный кроссовер 27](#_Toc514636832)

[4.4.3 Однородный кроссовер 27](#_Toc514636833)

[4.5 Операторы мутации 27](#_Toc514636834)

[4.5.1 Точечная мутация 27](#_Toc514636835)

[4.5.2 Инверсия 28](#_Toc514636836)

[4.5.3 Сальтация 28](#_Toc514636837)

[4.5.4 Транслакация 28](#_Toc514636838)

[4.6 Оператор обработки ограничений 28](#_Toc514636839)

[4.6.1 Метод штрафных функций 28](#_Toc514636840)

[4.7 Операторы селекции 32](#_Toc514636841)

[4.7.1 Линейно-ранговая схема селекции 32](#_Toc514636842)

[4.5.2турнир 33](#_Toc514636843)

[4.6 Оператор восстановления 34](#_Toc514636844)

[5. Классы тестовых задач 35](#_Toc514636845)

[5.1 Задачи без корреляции 35](#_Toc514636846)

[5.2 Задачи с корреляцией 35](#_Toc514636847)

[5.3 Задачи с сильной корреляцией 36](#_Toc514636848)

[6. Программный модуль 38](#_Toc514636849)

[7. Эксперименты 39](#_Toc514636850)

[7.1 Описание эксперимента по поиску эффективных комбинаций параметров ГА 39](#_Toc514636851)

[7.2 Описание эксперимента по исследованию производительности и качества алгоримтов 40](#_Toc514636852)

[7.3 Результаты 42](#_Toc514636853)

[7.2.1 Результаты эксперимента по поиску эффективных комбинаций параметров ГА 42](#_Toc514636854)

[8. Заключение 46](#_Toc514636855)

[9. Литература 48](#_Toc514636856)

[9. Приложение 49](#_Toc514636857)

# **1. Введение**

## 1.1 Актуальность

Классическая задача о ранце относится к числу широко известных задач дискретной оптимизации. Впервые она была сформулирована Д. Данцигом и с тех пор находится в активном исследовании. Основные сферы применения находятся в областях планирования и управления экономическими, производственными и транспортными системами. В частности, формулируемые в рамках классической задачи о ранце задачи объемного планирования для предприятий с единичным и мелкосерийным характером производства, сбора биржевого пакета и загрузки транспортных средств (складское планирование).

В работе рассматривается задача формирования инвестиционного портфеля, где в роли цены акции выступает вес, а выгодностью акции является цена в классической задачи о ранце.

Классическая задача о рюкзаке – NP - трудная задача с множеством применений.

Предлагаемый метод решения задачи о ранце – **генетический алгоритм**. Данный алгоритм выбран в качестве **объекта исследования**.

Генетические алгоритмы — широко развивающийся класс оптимизационных алгоритмов, использующих идеи искусственного интеллекта и предназначенный для ведения направленного поиска. С их помощью могут быть найдены решения многих задач в различных областях, таких как составление расписаний, построение конечных автоматов, построение и настройка искусственных нейронных сетей и т.д.

**Эффективность использования ГА, заключается в следующем:**

Если о пространстве поиска задачи есть некоторая дополнительная информация (как, например, пространство для известной задачи о ранце), метод поиска ГА, использующий эвристики, определяемые пространством, часто превосходит любой универсальный метод. При достаточно сложном рельефе функции приспособленности методы поиска с единственным решением в каждый момент времени, например, простой метод спуска, могли "затыкаться" в локальном решении, однако ГА, так как они работают с целой "популяцией" решений, имеют меньше шансов сойтись к локальному оптимуму и робастно функционируют на многоэкстремальном ландшафте.

Конечно, такие предположения не предсказывают строго, когда ГА будет эффективной процедурой поиска, конкурирующей с другими алгоритмами.

Эффективность ГА сильно зависит от таких деталей, как метод кодировки решений, операторы, числовые параметры (количество итераций, количество особей и т. д.), частный критерий успеха.

В моей работе в качестве **предмета исследования выбран подбор параметров ГА для наиболее эффективного поиска решения задачи формирования инвестиционного портфеля**.

**Актуальность** исследования предопределена широкой распространенностью и важностью различных прикладных проблем, решаемым генетическим алгоритмом.

## 1.2 Цель работы

Главной целью моей работы является:

* Реализация генетического алгоритма для классической задачи о ранце.
* Подбор параметров ГА для наиболее эффективного поиска решения задачи формирования инвестиционного портфеля (ЗФИП).

ЗФИП имеет особенности в виде корреляций между ценой и весом, поэтому исследование проводится для 4 классов тестовых задач:

* задачи без корреляции,
* коррелированные задачи,
* сильнокоррелированные задачи,
* задачи с подсуммами.

Оптимизация инвестиционного портфеля с помощью генетических алгоритмов повышает качество инвестирования финансовых средств в виде надеждого сбережения капитала или получения максимального дохода при допустимом риске.

## 1.3 Методы исследования

В качестве метода исследования выступает эксперимент, который описан в главе 6.

# **2. Постановка задачи о рюкзаке**

## 2.1 Описание

Задача о ранце (рюкзаке) — оптимизационная задача укладки максимального числа ценных предметов в рюкзак при условии, что общая его емкость ограничена. Она является актуальной и достаточно востребованной с точки зрения приложения в реальной жизни. Задача о загрузке и её модификации возникают в таких областях, как: прикладная математика, экономика, криптография, генетика и логистика (например, для нахождения оптимальной загрузки транспорта или склада).

*NP-*трудность задач

## 2.2 Варианты постановки задачи о рюкзаке[[1]](#footnote-1)

Существует множество разновидностей задачи о ранце, отличия заключаются в условиях, наложенных на рюкзак, предметы или их выбор.

1. **Рюкзак 0-1** ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) 0-1 Knapsack Problem): не более одного экземпляра каждого предмета.
2. **Ограниченный рюкзак** (англ. Bounded Knapsack Problem) - обобщение классической задачи, когда любой предмет может быть взят некоторое количество раз.
3. **Неограниченный рюкзак** (целочисленный рюкзак) ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) Unbounded Knapsack Problem (integer knapsack)): произвольное количество экземпляров каждого предмета.
4. **Непрерывный рюкзак** (англ. Continuous knapsack problem) - вариант задачи, в котором возможно брать любою дробную часть от предмета, при этом удельная стоимость сохраняется.
5. **Рюкзак с мультивыбором** ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) Multiple-choice Knapsack Problem): предметы разделены на группы, и из каждой группы требуется выбрать только один предмет.
6. **Мультипликативный рюкзак** ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) Multiple Knapsack Problem): есть несколько рюкзаков, каждый со своим максимальным весом. Каждый предмет можно положить в любой рюкзак или оставить.
7. **Многомерный рюкзак** ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) Multy-dimensional knapsack problem): вместо веса дано несколько разных ресурсов (например, вес, объём и время укладки). Каждый предмет тратит заданное количество каждого ресурса. Надо выбрать подмножество предметов так, чтобы общие затраты каждого ресурса не превышали максимума по этому ресурсу, и при этом общая ценность предметов была максимальна.

В данной работе мы будем рассматривать задачи 1 и 3.

## 2.3 Постановка задачи о 0-1 рюкзаке (ЗОР)

## 2.3.1 Словесная

Пусть дан набор предметов, каждый из которых имеет два параметра – вес и ценность. Дан рюкзак c заданным значением емкости. Задача заключается в том, чтобы заполнить рюкзак предметами, получить максимальную суммарную ценность, соблюдая при этом весовое ограничение.

## 2.3.2 Математическая

Пусть дано n предметов. Для каждого i-го предмета задан вес *> 0* и ценность  *> 0, i =1,.., n.* Задано ограничение на максимальный вес рюкзака . Каждый может принимать только одно из двух значений: *xi = 1*, если *i-*й предмет попадает в рюкзак, или*,* в противном случае. Требуется выбрать из заданного множества предметов набор с максимальной суммарной ценностью при одновременном соблюдении ограничения на суммарный вес найденного набора .

**(2.3.1)**

## 2.4 Постановка задачи о неограниченном рюкзаке (ЗОНР)

## 2.4.1 Словесная

Пусть дан набор предметов, каждый из которых имеет два параметра – вес и ценность. Каждый предмет можно класть неограниченное число раз. Дан рюкзак c заданным значением емкости. Задача заключается в том, чтобы заполнить рюкзак предметами, получить максимальную суммарную ценность, соблюдая при этом весовое ограничение.

## 2.4.2 Математическая

Пусть дано n предметов. Для каждого i-го предмета задан вес *> 0* и ценность  *> 0, i =1,.., n.* Задано ограничение на максимальный вес рюкзака . Каждый может принимать значение в диапазоне. Требуется выбрать из заданного множества предметов набор с максимальной суммарной ценностью при одновременном соблюдении ограничения на суммарный вес найденного набора .

**(2.4.1)**

## 2.5 Свойства задачи о неограниченном рюкзаке

## 2.5.1 Доминирование

Если i-й предмет менее выгодный и тяжелее чем j-й предмет, то он никогда не будет присутствовать в оптимальном решении, так как всегда лучше заменить любую копию i-ого предмета одной или более копиями j-ого предмета без уменьшения выгодности.

Такое отношение между предметами называется **простым**(**однократным) доминированием**. Его расширением является **множественное доминирование,** при котором i-й предмет множественно доминируем j-ым предметом, если *,* то есть всегда лучше заменить одну копию i-ого предмета копиями j-ого предмета.

Существует **пороговое доминирование**, при котором i-й предмет порогово доминируем набором предметов , если, для  *,* то есть всегда лучше заменить копий i-ого предмета некоторой линейной комбинацией предметов Для случая доминирование называется **коллективным**.



Рисунок 1. Типы доминирования

*Для каждой ЗОНР всегда существует оптимальное решение, которое не содержит любые просто-, множественно- или коллективно-доминируемые типы предметов.*

Это означает, что любые просто-, множественно- или коллективно-доминируемые типы предметов могут быть отброшены без изменения оптимального решения, значительно уменьшая пространство поиска.

## 2.5.2 Периодичность

Периодичность – это свойство ЗОНР, которое устанавливает, что емкость, большая чем некоторая емкость будет использовать только лучший предмет. И так, когда такая емкость достигнута (например, методом динамического программирования), оптимальное решение может быть вычислено по следующей формуле:

**(2.5.1)**

## 2.6 Области применения задачи о рюкзаке

Задача о ранце имеет применение в различных областях знаний: в математике, в экономике, информатике, криптографии. В вычислительной лингвистике в одной из работ предложена формулировка задачи [автоматического реферирования текстов](https://en.wikipedia.org/wiki/Automatic_summarization) (англ.)[русск.](https://ru.wikipedia.org/w/index.php?title=%D0%90%D0%B2%D1%82%D0%BE%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D1%80%D0%B5%D1%84%D0%B5%D1%80%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D1%82%D0%B5%D0%BA%D1%81%D1%82%D0%BE%D0%B2&action=edit&redlink=1), вырожденный (более простой) случай которой соответствует постановке задачи о ранце.

## 2.6.1 Изучение в математике[[2]](#footnote-2)

Одно из первых упоминаний о задаче о ранце можно найти в статье [Джорджа Балларда Мэтьюса](https://en.wikipedia.org/wiki/George_Ballard_Mathews) (англ.)[русск.](https://ru.wikipedia.org/w/index.php?title=%D0%9C%D1%8D%D1%82%D1%8C%D1%8E%D1%81,_%D0%94%D0%B6%D0%BE%D1%80%D0%B4%D0%B6_%D0%91%D0%B0%D0%BB%D0%BB%D0%B0%D1%80%D0%B4&action=edit&redlink=1), датированной 1897 годом. Интенсивное изучение данной проблемы началось после публикации [Д. Б Данцигом](https://ru.wikipedia.org/wiki/%D0%94%D0%B0%D0%BD%D1%86%D0%B8%D0%B3,_%D0%94%D0%B6%D0%BE%D1%80%D0%B4%D0%B6) в 1957 году книги «[англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) Discrete Variable Extremum Problem», особенно в 70-90-е годы 20-го века, как теоретиками, так и практиками. Во многом, данный интерес вызван достаточно простой формулировкой задачи, большим числом её разновидностей и свойств и в то же время сложностью их решения.

С практической точки зрения задача о рюкзаке может служить моделью для большого числа промышленных ситуаций:

* Размещение грузов в помещении минимального объёма.
* Раскройка ткани — для заданного куска материала получить максимальное число выкроек определенной формы.
* Расчет оптимальных капиталовложений.

С задачей о ранце сталкивается любой человек, собирающий рюкзак.

## 2.6.2 Изучение в экономике

Часто модификации задачи о ранце имеют практическое значение в экономике.

Предположим, что у нас есть 1000 миллионов долларов, которые нужно вложить в несколько возможных инвестиций. Каждое из этих вложений имеет различную стоимость и различный ожидаемый доход (выгодность). Мы должны решить, как потратить деньги, чтобы получить максимальную прибыль.  
 Подобные задачи называются задачами формирования инвестиционного портфеля. У нас есть несколько позиций (инвестиций), которые должны поместиться в портфель фиксированного размера (1000 миллионов долларов). Каждая позиция имеет свою прибыльность (выгодность). Необходимо найти набор позиций, которые помещаются в портфель и дают максимальную прибыль.

При составлении портфеля инвестор должен учитывать следующие факторы: степень риска - доходности, срок вложения, тип ценной бумаги. В зависимости от инвестиционной цели инвестор формирует портфель определенного типа. Сейчас принято выделять следующие типы портфелей.

* **Портфели роста**. Целью такого типа является рост капитала преимущественно не за счет получения дивидендов и процентов, а за счет роста курса ценных бумаг. По-другому такой портфель называют курсовым портфелем. Основные вложения делаются преимущественно в акции. В зависимости от соотношения ожидаемого роста капитала и риска можно выделить среди портфелей роста еще и виды портфелей:
  + **Портфели дохода.** Целью этого типа портфелей является получение дохода за набор дивидендов и процентов. Этот тип обеспечивает заранее спланированный уровень дохода при почти нулевом риске. Объектами инвестирования данного типа портфелей выступают высоконадежные ценные бумаги. По-другому портфель называют дивидендным.

Для портфелей роста свойственно быстрое изменение их структуры в зависимости от изменения курсов, входящих в портфель ценных бумаг. Портфели доходов имеют почти постоянные состав и структуру.

Перечисленные типы и виды представляют спектр возможных портфелей, но на практике инвесторы часто формируют портфели смешанного типа.

Наибольший вклад в создании теории инвестиционного портфеля был внесен американскими учеными Г. Марковицем и У. Шарпом, получившими за достижение в этой области Нобелевскую премию в 1990г.

Существует несколько моделей определения характеристик инвестиционного портфеля:

* Модель Марковица
* Модель Шарпа
* Модель «Квази - Шарп»

## 2.6.3 Изучение в криптографии[[3]](#footnote-3)

Проблема рюкзака лежит в основе первого алгоритма [асимметричного шифрования](https://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%B8%D0%BF%D1%82%D0%BE%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D1%81_%D0%BE%D1%82%D0%BA%D1%80%D1%8B%D1%82%D1%8B%D0%BC_%D0%BA%D0%BB%D1%8E%D1%87%D0%BE%D0%BC) (шифрования с открытым ключом). Идея [криптографии](https://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%B8%D0%BF%D1%82%D0%BE%D0%B3%D1%80%D0%B0%D1%84%D0%B8%D1%8F) с открытыми ключами была выдвинута [Уитфилдом Диффи](https://ru.wikipedia.org/wiki/%D0%A3%D0%B8%D1%82%D1%84%D0%B8%D0%BB%D0%B4_%D0%94%D0%B8%D1%84%D1%84%D0%B8), [Мартином Хеллманом](https://ru.wikipedia.org/wiki/%D0%A5%D0%B5%D0%BB%D0%BB%D0%BC%D0%B0%D0%BD,_%D0%9C%D0%B0%D1%80%D1%82%D0%B8%D0%BD) и независимо — [Ральфом Мерклом](https://ru.wikipedia.org/w/index.php?title=Ralph_Merkle&action=edit&redlink=1) ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) Ralph Merkle). Впервые она была представлена Диффи и Хеллманом на Национальной компьютерной конференции ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) National Computer Conference). Новизна по отношению к [симметричным криптосистемам](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BC%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D1%87%D0%BD%D1%8B%D0%B5_%D0%BA%D1%80%D0%B8%D0%BF%D1%82%D0%BE%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D1%8B) заключалась в использовании парных ключей — секретного ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) private key, secret key, SK) и открытого ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) public key, PK), создаваемых пользователем. Из названия понятно, что секретный ключ пользователь должен скрывать, а открытый может быть общедоступным. Открытый ключ нужен для шифрования, а секретный для расшифровки. Часто из секретного ключа получают открытый ключ.

**Криптосистема Меркла-Хеллмана** — первый, основанный на задаче о ранце, алгоритм для обобщённого шифрования с [открытым ключом](https://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%B8%D0%BF%D1%82%D0%BE%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D1%81_%D0%BE%D1%82%D0%BA%D1%80%D1%8B%D1%82%D1%8B%D0%BC_%D0%BA%D0%BB%D1%8E%D1%87%D0%BE%D0%BC). Разработан Ральфом Мерклом и Мартином Хеллманом в 1978 году. Был опубликован одностадийный ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) singly-iterated) и мультистадийный варианты ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) multiply-iterated). Алгоритм мог быть использован только для шифрования, но [Ади Шамир](https://ru.wikipedia.org/wiki/%D0%A8%D0%B0%D0%BC%D0%B8%D1%80,_%D0%90%D0%B4%D0%B8) адаптировал его для использования в [цифровых подписях](https://ru.wikipedia.org/wiki/%D0%AD%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%86%D0%B8%D1%84%D1%80%D0%BE%D0%B2%D0%B0%D1%8F_%D0%BF%D0%BE%D0%B4%D0%BF%D0%B8%D1%81%D1%8C).

В дальнейшем было предложено как множество модификаций криптосистемы Меркла-Хеллмана, так и совершенно новых криптосистем на основе задачи о ранце. Среди них:

1. Рюкзак Грэм-Шамира
2. Рюкзак Гудмана-Макколи
3. Рюкзак Накаше-Штерна
4. Рюкзак Шора-Ривеста

**Шифрование с помощью задачи о рюкзаке**

Сообщение шифруется как решение набора задач о ранце.

Рюкзачным вектором назовём упорядоченный набор из n предметов.

Для шифрования [открытого текста](https://ru.wikipedia.org/wiki/%D0%9E%D1%82%D0%BA%D1%80%D1%8B%D1%82%D1%8B%D0%B9_%D1%82%D0%B5%D0%BA%D1%81%D1%82) в двоичном представлении его разбивают на блоки длины  (например,  (11100) соответствует 5-ти предметам в рюкзаке). Считается, что единица указывает на наличие предмета в рюкзаке, а ноль на его отсутствие.

**Пример** шифротекста, полученного по данному алгоритму.

Пусть задан рюкзачный вектор Α = (3 4 6 7 10 11) с длинной n = 6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| открытый текст | 1 1 1 1 1 0 | 0 0 1 1 0 0 | 0 0 0 0 0 0 | 0 0 0 0 0 1 |
| вещи в рюкзаке | 3 4 6 7 10 11 | 3 4 6 7 10 11 | 3 4 6 7 10 11 | 3 4 6 7 10 11 |
| шифротекст | 3 + 4 + 6 + 7 + 10 = 30 | 6 + 7 = 13 | 0 | 11 |

Таблица 1. Пример шифрования

Для заданного Α - все криптосистемы - есть числа, не превышающие 41, суммарный вес всех предметов в рюкзачном векторе. Для каждого исходного текста существует единственный криптотекст. В указанном примере можно получить одинаковый шифротекст для векторов 100010 и 001100. Ошибка?

Верно, ошибка. Для шифрования необходим сверхвозрастающий рюкзак, то есть элементы упорядоченного рюкзачного вектора должны являться [сверхвозрастающей последовательностью.](https://ru.wikipedia.org/wiki/%D0%A1%D0%B2%D0%B5%D1%80%D1%85%D0%B2%D0%BE%D0%B7%D1%80%D0%B0%D1%81%D1%82%D0%B0%D1%8E%D1%89%D0%B0%D1%8F_%D0%BF%D0%BE%D1%81%D0%BB%D0%B5%D0%B4%D0%BE%D0%B2%D0%B0%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%BE%D1%81%D1%82%D1%8C)

**2.6.4 Изучение в информатике**

При решении задачи о ранце всегда нужно выбирать между точными алгоритмами, которые не применимы для «больших» рюкзаков, и приближенными, которые работают быстро, но не обеспечивают точного решения задачи. Естественно, создание быстрого и достаточно точного алгоритма представляет большой интерес с точки зрения информатики.

# **3. Рассматриваемые алгоритмы решения задачи о рюкзаке**

В данной работе исследуются несколько подходов к решению задачи о 0-1 рюкзаке и неограниченном рюкзаке.

Подходы к решению ЗОР:

1. Метод динамического программирования
2. Метод ветвей и границ

Подходы к решению ЗОНР:

1. Классический метод динамического программирования
2. Метод эффективного динамического программирования
3. Метод ветвей и границ с использованием верхней границы по трем предметам с лучшим соотношением цены и веса (U3)

Так же рассмотрен генетический алгоритм с различной комбинацией параметров, применимый к обеим типам задач.

## 3.1 Точные алгоритмы решения ЗОР

### 3.1.1 Метод динамического программирования

В основе метода динамического программирования (ДП) лежит принцип оптимальности Беллмана:” Каково *бы ни было состояние системы перед очередным шагом, надо выбирать управление на этом шаге так, чтобы выигрыш на этом шаге плюс оптимальный выигрыш на всех последующих шагах был оптимальным*”. Проще говоря, оптимум на i-ом шаге находится исходя из найденных ранее на предшествующих шагах. Из этого следует, что для того чтобы найти оптимальное решение на последнем шаге надо сначала найти оптимальное решения для первого, затем для второго и так далее пока не пройдем все шаги до последнего.

оптимальное решение для текущей задачи:

**(3.1.1)**

находится по следующей формуле:

**(3.1.2)**

вычисляется по рекуррентной формуле*:*

**(3.1.3)**

В работе рассмотрены два подхода к реализации ДП:

* *Табличный* (прямой). По формуле (3.1.3) последовательно вычисляются ячейки таблицы .
* *Рекуррентный (обратный).* По формуле (3.1.3) рекурсивно вычисляется *.* В обратном методе вычисляются только необходимые ячейки таблицы.

Для хранения таблицы стоимости и запоминания того, брался каждый предмет или нет, требуется порядка памяти, временная сложность равна

**3.1.2 Метод ветвей и границ**

[Метод ветвей и границ](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%B2%D0%B5%D1%82%D0%B2%D0%B5%D0%B9_%D0%B8_%D0%B3%D1%80%D0%B0%D0%BD%D0%B8%D1%86) (МВИГ) является вариацией метода полного перебора с исключением заведомо неоптимальных решений.

Пусть есть оптимальное решение  . Попытаемся его улучшить, рассмотрев решение на другой ветви. Если на рассматриваемой в данной момент ветви решение становится хуже (с какого-то шага), чем , то прекращаем его исследование и выбираем другую ветвь дерева.

Применяя метод ветвей и границ, можно значительно сократить количество вариантов для перебора. Однако МВИГ работает эффективно не для всех наборов данных. Существуют примеры, в которых время выполнения будет таким же, как и для простого перебора.

**Пример применения МВИГ**

При использовании метода ветвей и границ строится сеть. По оси  откладываем количество предметов, по оси  — их вес. На первом шаге из начала координат строятся две линии: горизонтальная, соответствующая тому, что первый предмет не был взят, и наклонная, соответствующая взятому первому предмету. Их проекции на ось  равны весу предмета. На втором шаге опять строим 2 линии, горизонтальная (второй предмет не был взят) или наклонная (второй предмет взят). Положим длину горизонтальных дуг равной нулю, а наклонных — ценности предмета.

Таким образом, любому решению задачи соответствует некоторый путь в сети. Наша задача свелась к нахождению пути максимальной длины.

Пусть вместимость рюкзака .

|  |  |  |
| --- | --- | --- |
| i | Ценность | Вес |
| 1 | 3 | 5 |
| 2 | 5 | 10 |
| 3 | 4 | 6 |
| 4 | 2 | 5 |

Таблица 2. Пример задачи о ранце

На рисунке (Рисунок 4), в квадратных скобках [] стоит суммарная ценность на каждом шаге алгоритма. Видно, что для конкретного примера она равна 17.

****

Рисунок 2. График нахождения оптимума методом ветвей и границ

**В работе рассмотрен следующий подход к реализации МВИГ:**

1. Создаем узел 1. {(индекс предмета)}. Добавляем узел 1 в контейнер. Тип контейнера зависит от выбранного пользователем обхода: в глубину – стек, в ширину – очередь. Задаем нижнюю оценку .
2. *Если* контейнер пуст*,*  – искомое оптимальное решение. *Иначе*, вынимаем узел из контейнера. Создаем узел :

**(3.1.4)**

1. Если , не добавляем этот узел в контейнер. Переходим к шагу 6.
2. Если то устанавливаем .
3. Считаем нижнюю оценку для текущего узла. Если нижняя оценка больше, чем, добавляем узел в контейнер.
4. Создаем узел . Узел, который не содержит предмет в решении.

**(3.1.5)**

Считаем нижнюю оценку для текущего узла. Если нижняя оценка больше, чем, добавляем узел в контейнер. Возвращаемся к шагу 2.

**Примечания:**

В качестве алгоритма нахождения нижней оценки выбрана **жадная нижняя оценка**. Предметы, нерассмотренные в текущем решении, сортируются по невозрастанию удельной стоимости и заполняют рюкзак, если не превышают его объем. Если не все предметы поместились в ранец, то берем последний непоместившийся i-ый предмет и увеличиваем нижнюю оценку по следующей формуле:

**(3.1.6)**

Перед запуском метода ветвей и границ предметы сортируются по невозрастанию удельной стоимости.

Сложность метода ветвей и границ .

## 3.2 Точные алгоритмы решения ЗОНР

## 3.2.1 Классический метод динамического программирования

Динамическое программирование (ДП) — это метод, использующий для решения больших задач решения меньших задач. Большие задачи, решаемые ДП, имеют перекрывающиеся подзадачи, которые используются для построения их оптимального решения. Решение каждой подзадачи обычно хранится в таблице, поэтому подзадачи никогда не рассчитываются более одного раза. ЗОНР может быть решена путем рассмотрения рюкзаков меньших мощностей в качестве подзадач, сохраняющих максимальную прибыль для каждой емкости. Оптимальное решение для рюкзака емкости можно найти, используя решение для рюкзаков с емкостью, меньшими .

Предположим, что рюкзак емкости 11 должен быть заполнен двумя предметами: один из них с весом 5 и прибылью 10, а другой с весом 2 и прибылью 3. Предположим, что решение (наибольшая сумма прибыли) для рюкзаков емкостей до 10 известны. Если мы используем первый элемент, мы по-прежнему будем иметь 6 единиц веса, поэтому, если оптимальное решение для ранца веса 6 равно 10, использование первого элемента приведет к решению с прибылью 10 + 10 = 20. Если мы, однако, используем второй элемент, мы по-прежнему будем иметь 9 единиц веса, если оптимальное решение для рюкзака с емкостью 9 равно 16, то при использовании второго элемента получается решение с прибылью 16 + 3 = 19. Таким образом, оптимальное решение для ранца с весом 11 равно 20. И это значение может быть использовано для решения рюкзаков емкостью более 11.

Способ, проиллюстрированный приведенным выше примером, суммируется в следующей функции:

**(3.2.1)**

– текущая емкость

Для вычисления функции 3.2.1 используется вектор . Этот вектор содержит наилучшую прибыль для каждой емкости . Классический алгоритм динамического программирования для ЗОНР показан на рисунке 3. Сначала в строке 1 устанавливается в ноль, так как все предметы имеют вес больше нуля. Затем для каждой емкости используется тип элемента, который вписывается в текущий рюкзак и который дает наибольшую прибыль. Обратите внимание, что предыдущие решения меньших задач никогда не пересчитываются, они извлекаются из вектора . После вычисления для каждого оптимальное решение можно получить из .



Рисунок 3. Псевдокод классического метода динамического программирования

Временная сложность данного метода — , строка 6 выполняется раз (количество предметов) для каждой емкости, меньшей . Сложность пространства — , так как прибыль для каждой емкости, меньшей , сохраняется.

## 3.2.2 Метод эффективного динамического программирования

Метод эффективного динамического программирования (ЭДП) состоит из двух основных фаз: **фаза сокращения** и **стандартная фаза**.

**Предварительная обработка**

ЭДП не требует предварительной обработки, так как любой тип предмета обрабатывается после того, как будет посчитан каждый объем, меньший чем вес текущего предмета, то есть будет уже посчитан для каждого предмета j с . Преимущество данного подхода в том, что просто-, множественно- и коллективно-пороговые доминирования могут быть определены простым тестом, выполняющимся за для каждого типа предмета: если для i-ого предмета выполняется неравенство , то i-ый предмет никогда не будет использоваться в оптимальном решении, поэтому он не попадает в набор недоминируемых предметов.

**Фаза сокращения**

На этом этапе предметы обрабатываются в срезах по t, отсортированные по увеличению веса. Держится список недоминируемых элементов F, упорядоченных по невозрастанию удельной стоимости (соотношению цены и веса). В начале F пусто, . Для каждого элемента вычисляется по формуле 3.2.1.

Если прибыль i-ого предмета больше чем текущая , то i-й предмет попадает в F. В противном случае, он отбрасывается, как порогово-доминируемый.

**Определение порогового доминирования**

В конце каждого среза предметов проверяется пороговое доминирование для каждого предмета. Предмет является порогово-доминируемым для емкостей, больших чем , если он неэффективный для какого-нибудь ранца с емкостью .

Чтобы эффективно определить это, вектор L сохраняет последнюю посчитанную емкость для каждого i-ого предмета***,*** где i-йбыл наиболее эффективным, используемым в оптимальном решении. Для поддержания данного вектора действительным, F сортируется по невозрастанию удельной стоимости, устанавливается равным , если . Итак, если полностью посчитана, любой предмет, для которого *,* может быть удален.



Рисунок 4. Псевдокод определения порогового доминирования

**Стандартная фаза**

На этом шаге высчитывается в срезах по q емкостям. В конце каждого среза проверяется пороговое доминирование для каждого предмета таким же образом, как описано выше. Если мощность F достигает единицы, алгоритм останавливается и использует формулу **периодичности**(2.5.1) для вычисления оптимального решения. Если этого не происходит, алгоритм останавливается, когда емкость достигает .

**Параметры**

Алгоритм требует два параметра:

* t– размер срезов предметов в фазе сокращения;
* q – размер срезов емкости в стандартной фазе.

В конце каждого среза проверяется пороговое доминирование для каждого предмета из F. Если значения параметров будут слишком малы, проверка будет занимать значительный промежуток времени. Если же они будут слишком велики, предметы, которые требуется удалить, будут занимать много времени в процедуре удаления. Подбор параметров t и q не исследуется в данной работе, но планируется в дальнейшем.

**Сложность**

Количество примитивных операций, выполняющихся в данном алгоритме, зависит от размера рюкзака и от количества предметов. Данный алгоритм можно рассматривать как вычисление таблицы. Хотя в большинстве реальных случаев периодичность достигается до вычисления , худший случай все еще может произойти. Заметим, что поскольку сложность алгоритма ограничена не размером задачи, а числовым значением ввода, его временная сложность называется псевдо-полиномиальной.



Рисунок 5. Псевдокод ЭДП

## 3.2.3 Метод ветвей и границ с использованием верхней границы по трем предметам с лучшим соотношением цены и веса (U3)

Метод ветвей и границ состоит в перечислении каждой комбинации типов предметов, сохраняя нижнюю и верхнюю границы на оптимальном решении. Нижняя граница вычисляется для каждой текущей ветки набора предметов, например, используя жадный алгоритм соотношения цены и веса. Верхняя граница вычисляется для всей задачи.

Для последней мы использовали алгоритм, называемый , который использует три наиболее эффективных типа предметов. Предположим, что *.* Верхняя граница определяется как максимум среди и . Пусть – это емкость, оставшаяся после использования максимального количества элементов типа 1, – это емкость, оставшаяся после использования максимального количества элементов типа 2 в рюкзаке емкостью и – это прибыль при решении, использовавшем максимальное количество предметов типа 1 и оставшуюся емкость с предметами типа 2. Таким образом, получаем две оценки:

**(3.2.2)**

**(3.2.3)**

Граница – это прибыль, полученная использованием оставшейся емкости с предметом 3-его типа, – это значение удаления некоторых предметов 1-ого типа из решения полученного из и замены их предметами 2-ого типа.

**(3.2.4)**

**Алгоритм МВИГ с верхней оценкой U3**:

1. Создаем узел 1. {(индекс предмета)}. Добавляем узел 1 в контейнер. Тип контейнера зависит от выбранного пользователем обхода: в глубину – стек, в ширину – очередь. Задаем нижнюю оценку . Верхнюю оценку высчитываем по формуле 3.2.4.
2. *Если* контейнер пуст*,*  – искомое оптимальное решение. *Иначе*, вынимаем узел из контейнера.

Создаем узлы для , изменяющегося от максильмально допустимого количества предмета типа дo 0:

**(3.2.5)**

* 1. Создаем узел:
  2. *Если* , не добавляем этот узел в контейнер. Переходим к шагу 2.6.
  3. *Если* то устанавливаем .
  4. *Если* , – искомое оптимальное решение.
  5. Считаем нижнюю оценку для текущего узла. Если нижняя оценка больше, чем, добавляем узел в контейнер.
  6. Если , переходим к шагу 2, иначе к шагу 2.1.

Примечания, описанные в МВИГ для ЗОР (пункт 4.1.2) также применяются и для данного алгоритма.

Производительность метода ветвей и границ зависит от класса задач, из-за чего трудно предсказать его поведение. В некоторых случаях, он может деградировать до экспоненциального времени выполнения, однако, МВИГ лучше классического метода динамического программирования для задач о рюкзаке с большими емкостями.

## 3.3 Генетический алгоритм

**Генетический алгоритм(ГА)** — это [эвристический алгоритм](https://ru.wikipedia.org/wiki/%D0%AD%D0%B2%D1%80%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC) поиска, используемый для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных [естественному отбору](https://ru.wikipedia.org/wiki/%D0%95%D1%81%D1%82%D0%B5%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BE%D1%82%D0%B1%D0%BE%D1%80) в природе. Является разновидностью [эволюционных вычислений](https://ru.wikipedia.org/wiki/%D0%AD%D0%B2%D0%BE%D0%BB%D1%8E%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5_%D0%B2%D1%8B%D1%87%D0%B8%D1%81%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F), с помощью которых решаются оптимизационные задачи с использованием методов естественной эволюции, таких как: [наследование](https://ru.wikipedia.org/wiki/%D0%9D%D0%B0%D1%81%D0%BB%D0%B5%D0%B4%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_(%D0%B1%D0%B8%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F)), [мутации](https://ru.wikipedia.org/wiki/%D0%9C%D1%83%D1%82%D0%B0%D1%86%D0%B8%D1%8F), [отбор](https://ru.wikipedia.org/wiki/%D0%95%D1%81%D1%82%D0%B5%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BE%D1%82%D0%B1%D0%BE%D1%80) и [кроссинговер](https://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%BE%D1%81%D1%81%D0%B8%D0%BD%D0%B3%D0%BE%D0%B2%D0%B5%D1%80). Отличительной особенностью генетического алгоритма является акцент на использование оператора «скрещивания», который производит операцию рекомбинации решений-кандидатов, роль которой аналогична роли скрещивания в живой природе.

Генетический алгоритм не гарантирует нахождение оптимального решения, однако показывает хорошие результаты за меньшее время по сравнению с другими алгоритмами.

Детальное описание ГА и его параметров представлено в следующей главе.

**4. Генетический алгоритм**

**4.1 Интерпретация задачи о ранце и операторов генетического алгоритма c помощью понятий популяционной генетики[[4]](#footnote-4)**

Наименьшей неделимой единицей биологического вида, подверженной действию факторов эволюции, является особь , (индекс *k* обозначает номер особи, а индекс *t* – некоторый момент времени эволюционного процесса). В качестве аналога особи *,* примем произвольное допустимое решение, которому присвоено имя . Вектор управляющих переменных – это наименьшая неделимая единица, характеризующая в задаче о рюкзаке внутренние параметры на каждом *t-*м шаге поиска оптимального решения, которые изменяют свои значения в процессе максимизации критерия оптимальности *Q*. Символьная модель задачи о ранце может быть представлена в виде множества двоичных кодировок, которые описывают конечное множество допустимых решений , принадлежащих области поиска . Для описания особей введем два типа вариабельных признаков, отражающих качественные и количественные различия между особями по степени их выраженности:

• качественные признаки – признаки, которые позволяют однозначно разделять совокупность особей на четко различимые группы;

• количественные признаки – признаки, проявляющие непрерывную изменчивость, в связи с чем степень их выраженности можно охарактеризовать числом. Качественные признаки особи , определяются из символьной модели задачи – как кодировка , соответствующая точке с именем и составляющие ее . Приведем интерпретацию этих признаков в терминах хромосомной теории наследственности. В качестве гена – единицы наследственного материала, ответственного за формирование альтернативных признаков особи, примем комбинацию , которая определяет фиксированное значение целочисленного кода управляющей переменной . Каждая особь характеризуется генами, а структуру строки можно интерпретировать хромосомой, содержащей сцепленных между собой генов, которые следуют друг за другом в строго определенной последовательности. Хромосому особи будем обозначать , т.е.

**(4.1.1)**

Согласно хромосомной теории наследственности передача генетической информации будет осуществляться через хромосомы от «родителей» к «потомкам». Местоположение определенного гена в хромосоме называется локусом, а альтернативные формы одного и того же гена, расположенные в одинаковых локусах хромосомы, называются аллелями (аллелеформами). В задаче поиска *i-*й локус соответствует *i-*й позиции в строковой кодировке *s(x)*, а аллели – это аналоги множества значений управляющих переменных.

Задача формализуется таким образом, чтобы её решение могло быть закодировано в виде [вектора](https://ru.wikipedia.org/wiki/%D0%92%D0%B5%D0%BA%D1%82%D0%BE%D1%80_(%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5)) («[генотипа](https://ru.wikipedia.org/wiki/%D0%93%D0%B5%D0%BD%D0%BE%D1%82%D0%B8%D0%BF)») генов, где каждый ген может быть [битом](https://ru.wikipedia.org/wiki/%D0%91%D0%B8%D1%82), числом или неким другим объектом. В классических реализациях генетического алгоритма (ГА) предполагается, что генотип имеет фиксированную длину. Однако существуют вариации ГА, свободные от этого ограничения.

Некоторым, обычно случайным, образом создаётся множество генотипов начальной популяции. Они оцениваются с использованием «функции приспособленности», в результате чего с каждым генотипом ассоциируется определённое значение («приспособленность»), которое определяет насколько хорошо [фенотип](https://ru.wikipedia.org/wiki/%D0%A4%D0%B5%D0%BD%D0%BE%D1%82%D0%B8%D0%BF), им описываемый, решает поставленную задачу. Полученные генотипы проходят несколько стадий (рисунок 6): отбор (селекция), скрещивание, мутация, создание новой популяции (нового поколения). Алгоритм повторяется заданное число итераций.



Рисунок 6. Стадии генетического алгоритма

**4.2 Представление генотипа**

Первый этап в реализации генетического алгоритма для определенной задачи – это разработка схемы представления. Для задачи о 0-1 рюкзаке в качестве гена будем рассматривать бит {0,1}, где 1 – предмет попал в рюкзак, 0 – иначе. Генотип будет являться набором битов.

Чтобы решить задачу о неограниченном рюкзаке, мы должны интерпретировать ее в стандартном 0-1 представлении. В качестве гена будет выступать набор битов размером, подходящим для двоичного представления максильмального допустимого количества предметов определенного типа. Размер вычисляется как . Например, для задачи с емкостью рюкзака 16, предмет с весом 2 может уместиться 8 раз. Соответственно, для представления числа 8 – нам нужно 4 бита. Ген для предмета с весом 2 будет состоять из четырех бит. Пример представления генотипа в задаче о неограниченном рюкзаке представлен на рисунке 7.



Рисунок 7. Представление генотипа в ЗОНР

**4.3 Операторы создания начальной популяции**

## 4.3.1 Алгоритм Данцига

Начальная популяция генерируется в зависимости от удельной стоимости. Предметы сортируются по убыванию удельной стоимости и в таком порядке помещаются в рюкзак, при этом каждый из предметов кладется с вероятностью.

Что бы получить несколько особей, введен элемент случайности – предметы с вероятностью 50% могут попасть в рюкзак.

## 4.3.2 Случайный алгоритм

Особи генерируются случайным образом с учётом функции приспособленности.

## 4.4 Операторы скрещивания[[5]](#footnote-5)

Механизм размножения с помощью операторов кроссовера прост, он содержит следующие простейшие операции:

* генерация случайных чисел;
* копирование строк фиксированной длины , являющихся «родительскими» кодировками;
* обмен кусками «родительских» кодировок, разрываемых в одной и той же точке хромосомы.

Применение этих операторов приводит к тому, что кодировки-потомки будут содержать новые сочетания аллелей генов, принадлежащих кодировкам родителей, т.е. происходит только перераспределение существующих аллелей родителей по хромосомам потомков.

## 4.4.1 Одноточечный кроссовер

Случайным образом выбирается одна точка разрыва из интервала  с вероятностью *.* Точка разрыва – участок между соседними битами в строке. Обе родительские структуры разрываются на два сегмента по этой точке. Затем, соответствующие сегменты различных родителей склеиваются и получаются два генотипа потомков.

## 4.4.2 Двуточечный кроссовер

«Родительские» кодировки разрываются в двух точках и случайным образом, выбранных с равной вероятностью без возвращения из интервала Затем, соответствующие сегменты различных родителей склеиваются, и получаются два генотипа потомков.

## 4.4.3 Однородный кроссовер

Оператор кроссовера, в котором единственный потомок воспроизводится таким образом, что аллель каждого гена копируется в хромосому потомка либо из хромосомы одного родителя, либо из хромосомы другого родителя с вероятностью .

## 4.5 Операторы мутации[[6]](#footnote-6)

## 4.5.1 Точечная мутация

Кодировка мутанта получается путем однократной перестановки аллелей в случайно выбранном локусе и соседнем с ним.

## 4.5.2 Инверсия

Кодировка разрывается в двух точках и *,* случайно выбранных с равной вероятностью из интервала . Значения аллелей во всех генах куска изменяются на обратные («1» на «0» или «0» на «1»). Кодировка мутанта образуется путем соединения вновь двух старых кусков и нового куска.

## 4.5.3 Сальтация

В кодировке случайным образом выбираются локусов, в каждой паре локусов переставляются значений аллелей.

## 4.5.4 Транслакация

В кодировке выделяется не один кусок , а сразу несколько непересекающихся между собой участков «родительской» хромосомы. Значения аллелей в мутанте внутри этих участков инвертируются.

## 4.6 Оператор обработки ограничений[[7]](#footnote-7)

## 4.6.1 Метод штрафных функций

Идея использования штрафных функций – значительное уменьшение приспособленности особей, не являющихся кодировками допустимых решений. Штрафные функции можно разделить на постоянные и адаптивные. *Постоянная* функция штрафа на все недопустимые особи накладывает одинаковый штраф, *адаптивная* накладывает штраф тем больше, чем «дальше» представляемое решение от допустимой области, например, квадратичная функция штрафа.

*В программе использована адаптивная функция шрафа:*

**(4.6.1)**

*Функция приспособленности:* **(4.6.2)**

После применения штрафа к функции приспособленности, некоторые особи, генотип которых является кодировкой недопустимого решения, могут иметь отрицательную или нулевую приспособленность. Существуют два варианта работы с недопустимыми особями:

* Первый – *удаление особей* с приспособленностью меньшей или равной нулю из популяции, но это значительно сокращает численность и генетическое разнообразие популяции.
* Второй вариант *– масштабирование приспособленностей*. В этом случае все особи сохраняются в популяции, но недопустимые особи имеют очень низкую, по сравнению с остальными особями, приспособленность. Одним из наиболее простых и часто используемых методов является *линейное динамическое масштабирование*.

В реализации генетического алгоритма был использован второй вариант, *метод линейного динамического* масштабирования для избежания сокращения численности и генетического разнообразия:

**(4.6.3)**

где ( для задачи максимизации); – коэффициент масштабирования.

Коэффициенты и выбираются в каждом *t-*м поколении по информации о значениях функции приспособленности для обрабатываемых кодировок, исходя из следующих двух условий.

* Среднее значение масштабированной функции приспособленности для множества

**(4.6.4)**

должно равняться среднему значению необработанной функции приспособленности для множества

**(4.6.5)**

Учитывая **(4.6.1)** получим:

**(4.6.6)**

Выполнение этого равенства гарантирует, что всякая кодировка , имеющая значение функции приспособленности совпадающее со значением средней приспособленности по популяции, при использовании схемы пропорциональной селекции репродуцирует в среднем одну ожидаемую копию в популяцию .

* Наибольшее значение масштабируемой функции приспособленности должно находиться в следующем отношении со средним значением необработанной функции приспособленности:

**(4.6.7)**

где – ожидаемое число копий для особи, которая имеет максимальное значение необработанной функции приспособленности в . Выполнение равенства гарантирует, что «наилучшая» кодировка множества репродуцирует при использовании схемы пропорциональной селекции в среднем ожидаемых копий в популяцию .

Таким образом, решая уравнения **(4.6.5) – (4.6.7)** получаем, что для масштабируемой функции приспособленности **(4.6.3)** в каждом t-м поколении коэффициенты и должны вычисляться согласно следующим выражениям:

**(4.6.8)**

Здесь – значение функции приспособленности «наилучшей» особи в .



Рисунок 8. Зависимость значений масштабируемой функции приспособленности от для

и

Одно из требований, предъявляемых к необработанной функции приспособленности *m*, состоит в том, что она должна принимать неотрицательные значения для всех . Это требование должно сохраняться и для масштабируемой функции приспособленности:

**(4.6.9)**

Однако для некоторых репродукционных множеств масштабируемая функция приспособленности **(4.6.3)** с коэффициентами и из выражений **(4.6.8)** может принимать отрицательные значения. Такая ситуация возникает обычно, когда большинство кодировок репродукционного множества являются наиболее приспособленными, и их значения необработанной функции приспособленности отличаются друг от друга очень мало, в то время как ряд кодировок этого же множества имеет значения необработанной функции

приспособленности значительно ниже среднего значения приспособленности *.*

Рисунок 9. Зависимость значений масштабируемой функции приспособленности от для и

Для решения проблемы исключения отрицательных значений масштабируемой функции приспособленности **(4.6.3)** потребуем, чтобы она удовлетворяла следующим двум условиям:

2. – значение масштабированной функции приспособленности «наихудшей» кодировки. С учетом этих условий находим, что коэффициенты и , обеспечивающие неотрицательные значения масштабируемой функции приспособленности, должны вычисляться согласно следующим выражениям:

**(4.6.10)**

*Зависимость значений масштабированной функции приспособленности от значений необработанной функции приспособленности приведена на данном рисунке.*

**

Рисунок 10. Зависимость значений масштабируемой функции приспособленности от для и

Учитывая выражение **(4.6.3)**, условие **(4.6.9)** для «наихудшей» кодировки из можно записать в следующем виде:

**(4.6.11)**

Тогда, подставляя в неравенство **(4.6.11)** значения , из выражений **(4.6.8)**, получаем неравенство:

**(4.6.12)**

Таким образом, при линейном динамическом масштабировании **(4.6.3)**, если неравенство **(4.6.12)** для заданного значения выполняется, то коэффициенты и вычисляются с помощью выражений **(4.6.8)**, в противном случае коэффициенты и вычисляются с помощью выражений **(4.6.10)**. В нашем случае, коэффициенты вычисляются по формуле **(4.6.10)**.

## 4.7 Операторы селекции[[8]](#footnote-8)

## 4.7.1 Линейно-ранговая схема селекции

Схема линейной ранговой селекции основана на том, что в ней вместо значений функции приспособленности используются ранги :

**(4.7.1)**

|  |  |
| --- | --- |
|  |  |

– ранг *i-*й кодировки, когда все особи репродукционного множества , отсортированы в порядке монотонного неубывания значений функции приспособленности.

– нижняя граница ожидаемого числа копий для «наихудшей» особи с наименьшим значением функции приспособленности, имеющей ранг ;

– верхняя граница ожидаемого числа копий для «наилучшей» кодировки с наибольшим значением функции приспособленности, имеющей ранг.

Для того чтобы определить численные значения граничных оценок ожидаемого числа копийи, требуется, чтобы значения,определяемые формулой **(4.7.1)** удовлетворяли следующим условиям:

* значениядолжны монотонно увеличиватьсяпо отношению к возрастанию значений функции приспособленности *;*
* общая сумма ожидаемого числа копий решений, репродуцируемых в популяцию , должна равняться численности n этой популяции.

выбирается из интервала1,2случайным образом или детерминированно, авычисляется по формуле: .

На втором этапе селекции был использован алгоритм вероятностного отбора –*остаточный стохастический выбор с возвращением:*

* Пусть целая часть числа, а – дробная часть числа,являются действительными числами, определенными на первом этапе оператора селекции.

*Целочисленная фаза.*

1. Детерминированно в популяцию репродуцируется копий для каждого решения из репродукционного множества .
2. Определяется число недостающих копий ,необходимых для заполнения популяции полностью копиями:

**(4.7.2)**

1. Если*,*то алгоритм завершает свою работу.

*Дробная фаза.*

1. Определяется
2. Осуществляется отображение всех решений из в смежные отрезки на оси действительных чисел так, чтобы каждый сегмент *i-*го решения равнялся дробной частиего ожидаемого числа копий .
3. Случайным образом с равной вероятностью генерируется число.
4. В качестве копии выбирается то решение, в отрезок которого попало число .
5. Процедура стохастического выбора с возвращением повторяется с шага 6 до тех пор, пока не будет получено копий.

## 4.5.2турнир

Схема -турнирной селекции организована следующим образом:

Случайно, с равной вероятностью из выбирается группа из особей. Среди выбранных в группу особей определяется «наилучшая» особь, копия которой репродуцируется в популяцию следующего поколения. Группа особей возвращается в . Процедура выбора копии «наилучших» особей повторяется требуемое число раз. Число решений называется размером турнира. (см. приложение)

Наиболее простым является бинарный турнир, проводимый между двумя решениями *2*

## 4.6 Оператор восстановления

Попадаются задачи, с которыми метод штрафов не справляется - *незначительно* уменьшает функцию приспособленности у недопустимых кодировок, в результате чего алгоритм может сойтись к такой кодировке. Например, случай, когда у особи, не являющейся кодировкой допустимого решения, вес на единицу превосходит ограничение, а функция приспособленности близка к максимальной.

Во избежание такой ситуации, после селекции применяется *оператор восстановления*. Оператор восстановления берет за основу удельную стоимость , увеличивает или уменьшает значение гена для i-го предмета на основе . Чем больше удельная стоимость, тем выше шанс, что ген будет иметь максимальное допустимое значение для i-го предмета. Оператор восстановления состоит из двух фаз:

* В *первой фазе* анализируется каждая переменная(ген) в порядке неубывания и уменьшается, пока решение не становится допустимым.
* Во *второй фазе* анализируется каждая переменная в порядке невозрастания и увеличивается, пока решение может быть допустимым.

Цель первой фазы – получить допустимое решение из недопустимого, в то время как вторая фаза направлена на улучшение пригодности допустимого решения.

Оператор восстановления гарантированно всегда создает допустимое решение, независимо от исходного решения.

# **5. Классы тестовых задач**

Задача о рюкзаке имеет особенности в виде корреляций между стоимостью и весом, поэтому исследования проводятся для 4 классов тестовых задач:

**5.1 Задачи без корреляции**

В этом случае не существует корреляции между стоимостью и весом элемента. Такие примеры иллюстрируют ситуации, когда вес и стоимость независимы. Некоррелированые экземпляры, как правило, легко решить, так как существует большая разница между весами, из-за чего легко заполнить ранец.

случайно выбираются из диапазона

**5.2 Задачи с корреляцией**

Стоимость коррелирует с весом. Как правило, стоимость отличается от веса на пару процентов. Такие случаи – одни из наиболее реалистичных в управлении (возврат инвестиций, как правило, пропорционален вложенной сумме в некоторых небольших вариациях).

****

**5.3 Задачи с сильной корреляцией**

Подобные ЗОР трудно решить по двум причинам:

* Все аллели вокруг элемента разрыва имеют сходные веса, а это значит, что трудно объединить их таким образом, чтобы заполнить ранец.
* Существует относительно большая потеря при удалении мелких взвешенных элементов, что значит, что мы вообще не можем удалить любые мелкие предметы, чтобы освободить место для большого предмета.



Такие задачи используются в качестве меры возможности алгоритма для решения сложных проблем. Сортировка по невозрастанию относительной стоимости

соответствует упорядочению согласно неубывания весов.

Так для *0-1* рюкзака относительно легко ввести дополнительное ограничение:

, *b –* индекс элемента разрыва **(5.3.1)**

Ограничение говорит, что ни одно решение не будет содержать более предметов, и эти предметы будут самыми лёгкими. Любое решение, содержащее более тяжелые предметы, будет содержать меньшее количество предметов. Ограничения с противоложным неравенством могут быть наложены в том случае, если имеет место быть отрицательная стоимость.

На основании этого наблюдения Martello, Toth разработали эффективный алгоритм для решения сильно коррелированных задач. Но очевидно, что если хотя бы несколько предметов не последуют структуре сильно коррелированной задачи, то ограничение **(5.3.1)** будет слишком слабым, чтобы иметь какой-нибудь эффект.

**5.4 Задачи с подсуммами**

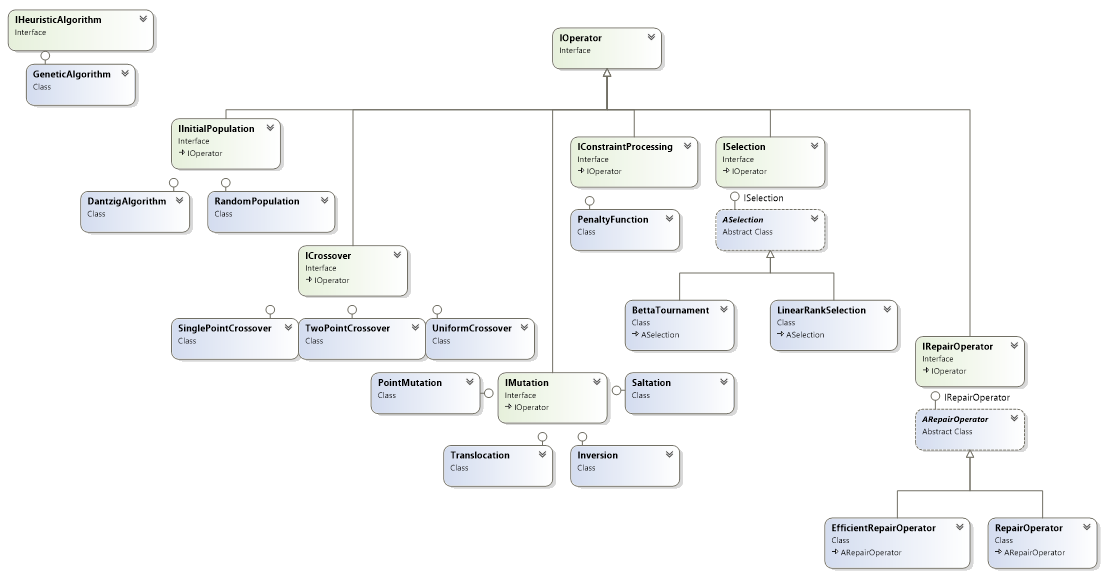
Они отражают ситуацию, когда прибыль каждого элемента является линейной функцией веса. Задачи с подсуммами нетрудно решаемы.



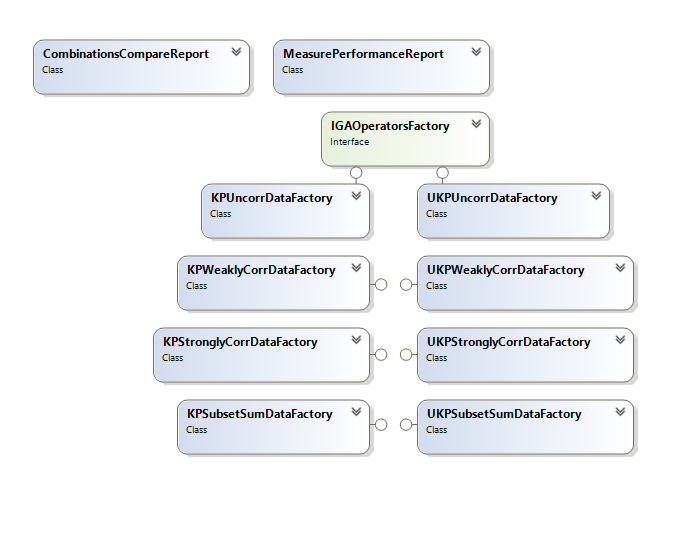
Для ЗОР емкость рюкзака вычисляется по формуле: , диапазон значений выбирается следующий: R1 = 1, R2 = 100.

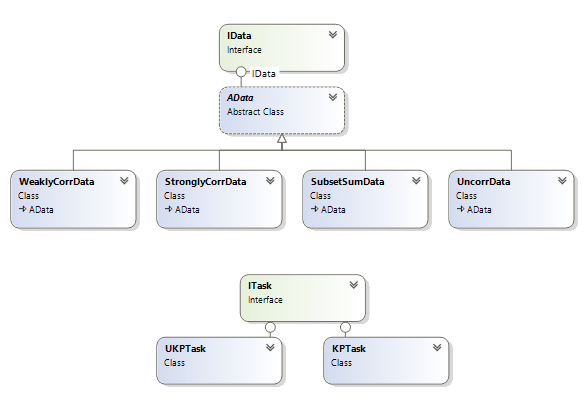
Для ЗОНР емкость рюкзака выбирается случайным образом из [10000, 100000], удовлетворяя условию, что , диапазон значений: R1 = 10, R2 = 9999.

# **6. Программный модуль**









# **Эксперименты**

В планах исследовательской работы предполагается проведение следующих экспериментов:

**Э1**. Поиск «эффективных» комбинаций параметров ГА, дающих наиболее выгодное решение для каждого типа задачи о рюкзаке.

**Э2**. Исследование производительности и качества рассмотренных точных и приближенных алгоритмов, оценка ошибки приближения.

## 7.1 Описание эксперимента по поиску эффективных комбинаций параметров ГА

Для подбора параметров ГА, влияющих на наиболее эффективный поиск оптимального решения был реализован алгоритм проведения «исследования ГА».

Для проведения исследования задаются начальные данные:

* Количество особей.
* Количество итераций (поколений).
* Так как перебираются все возможные комбинации операторов генетического алгоритма, учитывается задание значения (для -турнира, по умолчанию = 2).
* Количество экспериментов для конкретной задачи из класса тестовых задач. Один эксперимент – запуск программы с различными комбинациями операторов генетического алгоритма (всего 48 комбинаций = 48 запусков ГА для одного эксперимента).
* Количество запускаемых задач для каждого класса.

Все отчеты представляют собой Excel таблицу.

Для каждой задачи создается отдельная книга, название которой состоит из типа класса тестовых задач и индекса самой задачи, например, “uncorreled\_1.xlsx”, в ней:

* Для каждого эксперимента создается отдельный лист, в котором столбцы – комбинации операторов, а в строках выводится следующая информация:
* Значение функции приспобленности для каждого поколения.
* Максимальная функция приспособленности среди всех поколений.
* Количество итераций до получения к лучшей особи.
* Последний лист – общий результат экспериментов по данной задаче. В нем столбцы – комбинации операторов, а строки содержат следующую информацию:
* Максимальное значение функций приспособленности среди всех экспериментов.
* Вероятность получения лучшей особи среди всех экспериментов.
* Среднее значение количества итераций до получения лучшей особи среди всех экспериментов.

По всем сумммарным отчётам одного класса тестовых задач делается общий отчёт в отдельной книге “(type)\_summary.xlsx” (type – тип класса тестовых задач), в котором в строках выводятся комбинации операторов ГА, а в столбцах:

* Минимальное отклонение значения функции приспособленности в процентном соотношении от оптимума (для нахождения оптимума использован метод динамического программирования).
* Средняя вероятность получения лучшей особи в процентах.
* Cреднее количество итераций среди всех задач в процентом соотношении от наибольшего среднего количества итераций до получения лучшей особи (средняя скорость до наилучшего решения).

**Для одного запуска генетического алгоритма взяты следующие параметры:**

* 30 особей
* 40 поколений
* = 14 (для бетта-турнирной селекции)

**Для каждого типа задачи о рюкзаке проведено по 600 экспериментов (28800 запусков ГА), из них:**

* 30 экспериментов на 1 задачу
* По 5 задач на конкретный класс тестовых задач
* Рассмотрено 4 класса тестовых задач

Для генерации задачи о 0-1 рюкзаке диапазон значений веса и цены принимает следующее значение: , а для генерации задачи о неограниченном рюкзаке – *.*

## Описание эксперимента по исследованию производительности и качества алгоримтов

Произведено исследование производительности (скорости выполнения) и качества среди реализованных алгоритмов, чтобы подобрать к каждому классу тестовых задач наилучший подход к решению.

Рассматриваемые алгоритмы для решения задачи о 0-1 рюкзаке:

1. Табличный метод динамического программирования
2. Рекуррентный метод динаического программирования
3. Метод ветвей и границ
4. Генетический алгоритм

Рассматриваемые алгоритмы для решения задачи о неограниченном рюкзаке:

1. Классический метод динамического программирования
2. Метод эффективного динамического программирования
3. Метод ветвей и границ с использованием верхней границы по трем предметам с лучшим соотношением цены и веса (U3)
4. Генетический алгоритм

Входные данные:

1. Размер экземпляров задач
2. Количество запусков алгоритма для каждого экземпляра для нахождения медианы/среднего значения
3. Количество экземпляров на каждый тестовый класс задач
4. Требуемые входные параметры для ГА

В качестве операторов ГА для каждого класса тестовых задач выбраны эффективные комбинации, полученные из результатов эксперимента Э1.

Для каждого класса тестовых задач создается отчет в виде Excel таблицы, столбцы которой содержат названия исследуемых алгоритмов, а в строках для каждого прорешенного экземпляра задачи выводится следующая информация:

* Медиана времени выполнения алгоритма по заданному количеству запусков
* Среднее отклонение от оптимума % по заданному количеству запусков

Среди всех алгоритмов для каждого экземпляра класса находятся:

* Минимальная медиана времени выполнения
* Минимальное среднее отклонение от оптимума %

Алгоритм с данными показателем считается лучшим для текущего экземпляра класса тестовых задач.

Среди всех лучших алгоритмов для текущего экземпляра находится лучший алгоритм для конкретного класса задач по максимальному количеству минимальных медиан времени выполнения и максимальному количеству минимальных средний отклонений.

**Для одного запуска генетического алгоритма взяты следующие параметры:**

* 30 особей
* 40 поколений
* = 14 (для бетта-турнирной селекции)

**Для каждого типа задачи о рюкзаке проведены эксперименты со следующими параметрами:**

* Размеры задачи: 50, 100, 300 предметов
* По 30 запусков каждого алгоритма для решения одного экземпляра задачи
* По 30 экземпляров на конкретный класс тестовых задач
* Рассмотрено 4 класса тестовых задач

Для генерации ЗОР и ЗОНР диапазон значений веса и цены принимает значение: *.*

## Результаты

**7.3.1 Результаты эксперимента по поиску эффективных комбинаций параметров ГА**

По проведенному исследованию генетического алгоритма для каждого класса тестовых задач выбрано по 5 эффективных комбинаций.

*Эффективные (лучшие)* комбинации – комбинации операторов, влияющих на наиболее эффективный поиск решения.

Критериями отбора являлись:

* Вероятность нахождения глобального оптимума
* Скорость нахождения глобального оптимума

Для всех классов тестовых задач выбраны три эффективные комбинации. Критерий отбора – количество классов, содержащих комбинацию.

Ниже представлены таблицы с комбинациями параметров ГА, дающими «лучшие» решения.

Результаты для **Э1** задачи **ЗОР** предоставлены в таблицах 3 – 6 со следующими значениями строк:

1 – *набор параметров*

2 – *вероятность нахождения глобального оптимума %*

3 – *среднее количество итераций % (скорость нахождения глобального оптимума)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Алгоритм Данцига Однородный кроссовер Точечная мутация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Транслокация Линейная ранговая схема | Алгоритм Данцига  Однородный кроссовер Инверсия Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Сальтация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Инверсия Линейная ранговая схема |
| 72% | 69,33% | 69,33% | 69,33% | 68,67% |
| 17,95% | 16,93% | 17,17% | 17,4% | 17,62% |
|  |  |  |  |  |

Таблица 3. Задачи без корреляции

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Случайный алгоритм Однородный кроссовер Точечная мутация Линейная ранговая схема | Алгоритм Данцига Однородный кроссовер Сальтация Линейная ранговая схема | Алгоритм Данцига Однородный кроссовер  Инверсия Линейная ранговая схема | Алгоритм Данцига Однородный кроссовер Транслокация  Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Сальтация Линейная ранговая схема |
| 22,67% | 20,67% | 20,66% | 18,67% | 17,33% |
| 17,57% | 16,6% | 18,32% | 17,32% | 14,82% |

Таблица 4. Задачи с корреляцией

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Алгоритм Данцига Двуточечный кроссовер Инверсия Линейная ранговая схема | Алгоритм Данцига Одноточечный кроссовер  Инверсия Линейная ранговая схема | Случайный алгоритм  Однородный кроссовер Сальтация Линейная ранговая схема | Случайный алгоритм Двуточечный кроссовер Точечная мутация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Транслокация Линейная ранговая схема |
| 22,67% | 17,33% | 16% | 16% | 15,33% |
| 15,62% | 15,72% | 14,03% | 17,13% | 13,18% |

Таблица 5. Задачи с сильной корреляцией

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Алгоритм Данцига Однородный кроссовер Сальтация Линейная ранговая схема | Алгоритм Данцига Однородный кроссовер Инверсия Линейная ранговая схема | Случайный алгоритм Однородный кроссовер  Точечная мутация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Сальтация  Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Транслокация Линейная ранговая схема |
| 84,67% | 80,67% | 78% | 76,66% | 74% |
| 13,1% | 13,78% | 14,62% | 13,23% | 13,01% |

Таблица 6. Задачи с подсуммами

|  |  |  |  |
| --- | --- | --- | --- |
| Набор параметров | Случайный алгоритм  Однородный кроссовер Сальтация Линейная ранговая схема | Алгоритм Данцига  Однородный кроссовер  Инверсия Линейная ранговая схема | Случайный алгоритм  Однородный кроссовер Сальтация Линейная ранговая схема |
| Количество классов, содержащих комбинацию | 4 | 3 | 3 |

Таблица 7. По всем классам задач

Результаты для **Э2** задачи **ЗОНР** предоставлены в таблицах 8 – 11 со следующими значениями строк:

1 – *набор параметров*

2 – *вероятность достижения верхней границы %*

3 – *среднее количество итераций % (скорость достижения верхней границы)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cлучайный алгоритм Одноточечный кроссовер Сальтация Линейная ранговая схема | Случайный алгоритм Двуточечный кроссовер Сальтация Линейная ранговая схема | Случайный алгоритм  Двуточечный кроссовер Точечная мутация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Сальтация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Точечная мутация Линейная ранговая схема |
| 60% | 60% | 60% | 60% | 60% |
| 32,176% | 29,95% | 29,35% | 18,274% | 13,972% |
|  |  |  |  |  |

Таблица 8. Задачи без корреляции

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Случайный алгоритм Одноточечный кроссовер Точечная мутация Бетта-Турнир | Случайный алгоритм Однородный кроссовер Точечная мутация Бетта-Турнир | Алгоритм Данцига Двуточечный кроссовер  Точечная мутация Бетта-Турнир | Алгоритм Данцига Одноточечный кроссовер Точечная мутация Бетта-Турнир | Случайный алгоритм Одноточечный кроссовер Инверсия Линейная ранговая схема |
| 20% | 20% | 20% | 20% | 20% |
| 9,524% | 6,05% | 4,9% | 3,5% | 2,25% |

Таблица 9. Задачи с корреляцией

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Случайный алгоритм Одноточечный кроссовер Сальтация Линейная ранговая схема | Случайный алгоритм Двуточечный кроссовер  Транслокация Линейная ранговая схема | Случайный алгоритм  Однородный кроссовер Точечная мутация Линейная ранговая схема | Случайный алгоритм Одноточечный кроссовер Транслокация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Транслокация Линейная ранговая схема |
| 20% | 20% | 20% | 20% | 20% |
| 13,876% | 9,35% | 9,276% | 7,8% | 6,8% |

Таблица 10. Задачи с сильной корреляцией

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Случайный алгоритм Однородный кроссовер Сальтация Линейная ранговая схема | Случайный алгоритм Однородный кроссовер Точечная мутация Линейная ранговая схема | Алгоритм Данцига  Однородный кроссовер  Точечная мутация Линейная ранговая схема | Случайный алгоритм Двуточечный кроссовер Точечная мутация  Линейная ранговая схема | Случайный алгоритм Одноточечный кроссовер Точечная мутация Линейная ранговая схема |
| 73% | 70% | 70% | 69% | 67% |
| 33,35% | 23,748% | 22,372% | 21,726% | 20,75% |

Таблица 11. Задачи с подсуммами

|  |  |  |  |
| --- | --- | --- | --- |
| Набор параметров | Случайный алгоритм  Однородный кроссовер Точечная мутация Линейная ранговая схема | Случайный алгоритм  Одноточечный кроссовер  Сальтация Линейная ранговая схема | Случайный алгоритм  Однородный кроссовер Сальтация Линейная ранговая схема |
| Количество классов, содержащих комбинацию | 3 | 2 | 2 |

Таблица 12. По всем классам задач

Из результатов, видно, что в большинстве оптимальных комбинаций в качестве оператора селекции выступает линейная ранговая схема. Это подтверждает теорию о том, что данная схема предотвращает преждевременную сходимость и приводит к наиболее лучшему решению.

Для задач с корреляцией и с сильной корреляцией вероятность нахождения глобального оптимума (верхней границы для ЗОНР) не превышает , что подтверждает теорию о труднорешаемости таких задач, напротив, для задач без корреляции и с подсуммами вероятность нахождения глобального оптимума (верхней границы для ЗОНР) превышает , что говорит о менее трудном нахождении их решения.

# **8. Заключение**

**Итоги о проделанной работе:**

* Изучен материал о классической задачи о ранце, варианты ее постановки, области применения, вычислительная сложность, известные методы решения.
* В качестве метода решения и исcледуемого объекта выбран генетический алгоритм.
* Рассмотрены классы тестовых задач для задачи формирования инвестиционного портфеля:
* Задачи без корреляции
* Задачи с корреляцией
* Задачи с сильной корреляцией
* Задачи с подсуммами
* Pеализован генетический алгоритм для классической задачи о ранце с различными параметрами (2 оператора начальной популяции, 3 оператора кроссовера, 4 оператора мутации, 2 оператора селекции, 1 оператор обработки ограничений).
* Для увеличения генетического разнообразия в поколении реализованы линейное динамическое масштабирование в операторе обработки ограничений и модификация генотипа, применяющаяся после оператора селекции.
* Проведено «исследование алгоритма» (подбор параметров ГА) для каждого класса тестовых задач на выявление эффективных комбинаций из 48 вариантов, описанное в главе 6.
* Из полученых результатов исследования для каждого класса тестовых задач выбрано по 5 эффективных комбинаций. Критериями отбора являлись:
* Вероятность нахождения глобального оптимума
* Скорость нахождения глобального оптимума
* Для всех классов тестовых задач выбраны три эффективные комбинации. Критерий отбора - количество классов, содержащих комбинацию.
* Данные результаты так же проанализированы и сделаны выводы, подтверждающие теоретический материал:
* Во всех оптимальных комбинациях в качестве оператора селекции выступает линейная ранговая схема. Это подтверждает теорию о том, что данная схема предотвращает преждевременную сходимость и приводит к наиболее лучшему решению.
* Для классов тестовых задач с корреляцией и с сильной корреляцией вероятность нахождения глобального оптимума не превышает 22,67%, что подтверждает теорию о труднорешаемости таких задач, напротив, для классов тестовых задач без корряции и с подсуммами вероятность нахождения глобального оптимума достигает 72% и 84,67%, соответственно, что говорит о менее трудном нахождении их решения.

**Практическая значимость исследования** состоит в том, что подобраные эффективные наборы параметров генетического алгоритма для задачи формирования инвестиционного портфеля ускорят процесс и дадут наиболее точное решение для выбора самых выгодных акций по ограниченной цене.

Так как проблемы, решаемые генетическим алгоритмом, являются актуальными и по сей день, в дальнейшем можно продолжить его исследование исходя из других свойств:

* Метод кодировки решений.
* Настройки параметров, такие как:
  + количество особей,
  + количество итераций.
* Частный критерий успеха.

# **9. Литература**

1. Батищев, «Применение генетических алгоритмов к решению задач дискретной оптимизации» / Д.И. Батищев, Е. А. Неймарк, Н. В. Старостин. – Н.Н : ННГУ им. Лобачевского, 2007.
2. Пападимитриу, «Комбинаторная оптимизация. Алгоритмы и сложность» / Х. Пападимитриу, К. Стайглиц. – М.: Мир, 1985.
3. Гери, «Вычислительные машины и трудноҏешаемые задачи» / М. Гери, Д. Джонсон. – М.: Мир, 1982.
4. David Persinger, Algorithms for Knapsack Problems, Ph.D. thesis, February 1995, Dept, of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen, Denmark
5. И. Сигал, «Введение в прикладное дискретное программирование: модели и вычислительные алгоритмы» / И. Сигал, А. Иванова – 2 изд. – М.: Физматлит, 2007 – 304 с.
6. [A.N.M. Bazlur Rashid](http://www.amazon.com/s/ref=dp_byline_sr_book_1?ie=UTF8&text=A.N.M.+Bazlur+Rashid&search-alias=books&field-author=A.N.M.+Bazlur+Rashid&sort=relevancerank), The 0-1 Knapsack Problem: A solution by Genetic Algorithm Paperback – October 1, 2010
7. KEN-LI LI, GUANG-MING DAI, QING-HUA LI, A Genetic Algorithm for the Unbounded Knapsack Problem, November 2003, Computer School, Huazhong University of Science and Technology, Wuhan, 430074, China
8. LEONARDO FERNANDO DOS SANTOS MOURA, An Efficient Dynamic Programming Algorithm For The Unbounded Knapsack Problem, Porto Alege, December 14th, 2012
9. Vincent Poirriez, Nicola Yanev, Rumen Andonov, A Hybrid Algorithm gor the Unbounded Knapsack Problem
10. Swarna Chitra Lyer, A Complementary Heuristic for the Unbounded Knapsack Problem, Department of Computer and Mathematical Sciences Victoria University of Technology December, 199

**9. Приложение**

1. dic.academic.ru/dic.nsf/ruwiki/107129 – словари и энциклопедии на Академике [↑](#footnote-ref-1)
2. dic.academic.ru/dic.nsf/ruwiki/107129 – словари и энциклопедии на Академике [↑](#footnote-ref-2)
3. dic.academic.ru/dic.nsf/ruwiki/107129 – словари и энциклопедии на Академике [↑](#footnote-ref-3)
4. Д.И. Батищев, Е. А. Неймарк, Н. В. Старостин «Применение генетических алгоритмов к решению задач дискретной оптимизации»,Н.Н,2007, c.33 - 34 [↑](#footnote-ref-4)
5. Д.И. Батищев, Е. А. Неймарк, Н. В. Старостин «Применение генетических алгоритмов к решению задач дискретной оптимизации»,Н.Н,2007, с. 73 - 75 [↑](#footnote-ref-5)
6. Д.И. Батищев, Е. А. Неймарк, Н. В. Старостин «Применение генетических алгоритмов к решению задач дискретной оптимизации»,Н.Н,2007, с. 77 - 78 [↑](#footnote-ref-6)
7. Аналогично ссылке, с. 27 – 28, 65 - 71 [↑](#footnote-ref-7)
8. Д.И. Батищев, Е. А. Неймарк, Н. В. Старостин «Применение генетических алгоритмов к решению задач дискретной оптимизации»,Н.Н,2007, с. 55 - 58 [↑](#footnote-ref-8)