

РЕФЕРАТ

Дипломная работа: 22 с., 30 рис., 50 табл., 7 источников.

Ключевые слова: ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ, МАШИННОЕ ОБУЧЕНИЕ, НР-ПОЛНЫЕ ЗАДАЧИ, ЗАДАЧА О РЮКЗАКЕ, ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ, МНОГОМЕРНЫЙ РЮКЗАК

Объект исследования - задача о многомерном рюкзаке

Цель работы - исследование применения генетического алгоритма к задаче о многомерном рюкзаке.

Результатом работы является программа, реализующая генетический алгоритм на языке C# и его модификации. Разработанная программа позволяет найти точные максимумы для малых наборов предметов(<50 предметов) в 100% случаев и для больших(100 предметов) - в 43,3% случаев. В остальных случаях программа находит локальный максимум, различающийся с исходным не более чем на 0,21%.

Содержание

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	4
ВВЕДЕНИЕ	5
1 Теоретическая часть	6
1.1 Задача о многомерном рюкзаке	6
1.1.1 Постановка задачи	6
1.1.2 Доказательство NP-полноты	6
1.2 Генетические алгоритмы	8
1.2.1 Этапы работы алгоритма	9
1.2.2 Выбор параметров алгоритма в приложении к задаче о многомерном рюкзаке	10
1.2.3 Дополнения и улучшения, внесённые в генетический алгоритм	12
2 Практическая часть	13
2.1 Наборы тестов	13
2.2 Первый набор тестов	14
2.2.1 Мутация в одной позиции	14
2.2.2 Инверсионная мутация	15
2.2.3 Промежуточные выводы	16
2.3 Второй набор тестов	16
ЗАКЛЮЧЕНИЕ	19
ПРИЛОЖЕНИЯ	21
А Исходный код программы	22

ОБОЗНАЧЕНИЯ

В данной работе использованы следующие обозначения и сокращения.

Класс проблем NP (nondeterministic polynomial) – множество проблем разрешимости, решение которых возможно проверить на машине Тьюринга за время, не превосходящее полинома от размера входных данных.

NP-полная проблема – проблема из класса NP, к которой можно свести любую другую проблему из этого класса за полиномиальное время.

.NET Framework - программная платформа, выпущенная фирмой Microsoft. Платформа состоит из среды CLR, позволяющей исполнять программы, написанные на разных языках и собранные специально под эту среду, и библиотек классов .NET Framework.

ВВЕДЕНИЕ

В ходе развития компьютерных наук человечеством встретилось с классом NP-полных задач. Такие задачи решаются алгоритмически за недетерминированное полиномиальное время, что существенно затрудняет поиск решения таких задач в приемлемые сроки. Одной из таких задач является задача о рюкзаке (Knapsack problem) и её модификации. Объектом исследования данной работы является задача о многомерном рюкзаке (Multidimensional knapsack problem). Предмет исследования - решение задачи о многомерном рюкзаке с использованием генетического алгоритма. Многие прикладные проблемы могут быть формализованы в виде рассматриваемой задачи. Примерами таких проблем являются размещение процессоров и баз данных в системе распределенных вычислений (см. [1]), погрузка груза и контроль бюджета (см. [2]), задачи раскройки (см. [3]) и др. Решение этих проблем обуславливает актуальность решения задачи о рюкзаке.

Цель исследования - решить задачу о многомерном рюкзаке с использованием генетического алгоритма. Для достижения цели были поставлены следующие задачи:

- Исследовать генетические алгоритмы в применении к NP-полным задачам
- Спроектировать и реализовать генетический алгоритм, решающий задачу о многомерном рюкзаке
- Провести оценку эффективности генетического алгоритма в решении поставленной задачи, используя набор готовых тестов

Для реализации алгоритма был выбран язык C# ввиду удобства и простоты работы.

1 Теоретическая часть

1.1 Задача о многомерном рюкзаке

В данной работе рассматривается задача о многомерном рюкзаке (Multidimensional 0-1 knapsack problem, МКР). Эта задача является модификацией классической задачи о рюкзаке, поставленной в 19 веке Джорджем Мэттьюсоном. (см. [4]) Данный же вариант задачи впервые был предложен Клиффордом Петерсеном в 1967 году. (см. [5])

1.1.1 Постановка задачи

Постановка задачи такова:

Пусть существует набор S , состоящий из N предметов, каждый из которых имеет стоимость $c_i \in \mathbb{R}^+$ и набор размеров S $s_{ij} \in \mathbb{R}^+$, где $i \in 1, 2, \dots, N$, $j \in 1, 2, \dots, M$, $M, N \in \mathbb{N}$. Пусть также существует рюкзак с ограничениями по вместимости по измерениям $r_j \in \mathbb{R}^+$. Требуется максимизировать сумму

$$\sum_{i=1}^N c_i x_i$$

где $x_i \in \{0, 1\}$ при условии

$$\sum_{i=1}^N s_{ij} x_i < r_j \forall j \in \{1, 2, \dots, M\} \quad (1.1)$$

Обычную задачу о рюкзаке можно рассматривать как частный случай при $M = 1$. И обычная задача, и её модификация являются NP-полными задачами.

1.1.2 Доказательство NP-полноты

Для доказательства NP-полноты рассматриваемой задачи достаточно доказать NP-полноту обычной задачи о рюкзаке. Сначала покажем, что задача принадлежит к классу NP. Входными данными является набор предметов S , процесс проверки

правильности решения заключается в вычислении сумм

$$\sum_{i \in S} s_i x_i$$

и

$$\sum_{i \in S} c_i x_i$$

Очевидно, такое вычисление требует затрат полиномиального времени в зависимости от вводимых данных.

Затем покажем что задача о рюкзаке является NP-полной. Для этого воспользуемся задачей разбиения множества чисел, для которой доказана NP-полнота. Требуется показать, что существует такая полиномиальновременная редукция $Q()$, что $Q(X)$ дает ответ «Да» для задачи о рюкзаке тогда, и только тогда, когда X даёт ответ «Да» для задачи разбиения.

Обозначим ёмкость рюкзака как R , максимум, для которого рассматривается задача как V . Предположим, что существует набор a_i для задачи разбиения, рассмотрим соответствующую ему задачу о рюкзаке: $c_i = a_i, s_i = a_i \forall i \in \{1, 2, \dots, N\}$, при этом

$$R = V = \frac{1}{2} \sum_{i=1}^n a_i$$

. $Q()$ здесь - процесс, преобразующий задачу о разбиении к задаче о рюкзаке. Очевидно, что этот процесс является полиномиальным от входных данных.

Если X является тем случаем, для которого задача о разбиении имеет ответ «Да», то существуют такие множества S и T , что

$$\sum_{i \in S} a_i = \sum_{i \in T} a_i = \frac{1}{2} \sum_{i=1}^n a_i$$

. Пусть наш рюкзак содержит предметы из S , из чего следует, что

$$\sum_{i \in S} s_i = \sum_{i \in S} a_i = R$$

и

$$\sum_{i \in S} c_i = \sum_{i \in S} a_i = V$$

. Следовательно, $Q(X)$ является случаем с ответом «Да» для задачи о рюкзаке. Наоборот, если $Q(X)$ является случаем с ответом «Да» для задачи о рюкзаке, рассмотрим для выбранного набора S такое T , что $T = \{1, 2, \dots, n\} \sim S$. Имеем

$$\sum_{i \in S} s_i = \sum_{i \in S} a_i \leq R = \frac{1}{2} \sum_{i=1}^n a_i$$

и также

$$\sum_{i \in S} c_i = \sum_{i \in S} a_i \geq V = \frac{1}{2} \sum_{i=1}^n a_i$$

Отсюда получаем, что

$$\sum_{i \in S} a_i = \frac{1}{2} \sum_{i=1}^n a_i$$

и

$$\sum_{i \in T} a_i = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n a_i = \frac{1}{2} \sum_{i=1}^n a_i$$

Следовательно, $\{S, T\}$ – искомое разбиение, и X является случаем с ответом «Да» для задачи о разбиении. Таким образом, NP-полнота задачи о рюкзаке доказана.

Вычислительная сложность задачи такого рода при переборном решении для N предметов - $O(2^M N^3)$, что, вкупе с NP-сложностью, делает алгоритмическое решение такой задачи неэффективным для больших N . Однако такие задачи могут быть решены эвристическими алгоритмами, то есть алгоритмами, для которых их корректность строго не доказана.

1.2 Генетические алгоритмы

Генетические алгоритмы являются семейством в множестве эвристических алгоритмов. Впервые такой алгоритм был предложен А. Фразером. (см [6]) В генетическом алгоритме рассматривается популяция возможных решений (особей, фенотипов), которая итерационно подвергается эволюционному процессу, направленному на получение лучших решений. У каждого возможного решения существует набор свойств (генотип), который может быть изменен.

Решение задачи с помощью такого алгоритма требует нескольких предварительных этапов:

- Выбор кодирования генотипа.

На этом этапе нужно выбрать способ кодирования генотипа, который будет эффективен для данной задачи. Такой генотип должен позволять однозначно восстановить соответствующий ему фенотип.

- Выбор начального приближения.

Для запуска итерационного процесса требуется создать начальное множество генотипов - нулевое поколение. В зависимости от способа задания начального поколения скорость поиска оптимального решения может меняться. Размер поколения также оказывает влияние на эффективность алгоритма: слишком маленькое поколение подавляет разнообразие и приводит к попаданию в локальные максимумы, слишком большое - увеличивает число операций на каждой итерации, чем замедляет работу алгоритма.

- Выбор мутации.

На каждой итерации алгоритма заранее определенная часть пула генотипов подвергнется мутациям, то есть определенным случайным образом изменяются их составляющие. Если мутации подвергается слишком малая часть пула, то генетический алгоритм не сможет покинуть локальные максимумы, если же слишком большая - алгоритм потеряет свойство сохранения признаков фенотипа.

- Выбор механизма скрещивания (кроссинговера).

После мутации происходит создание новых генотипов из частей старых, что обеспечивает сохранение признаков фенотипа родителя. Алгоритм скрещивания позволяет получить из двух родительских генотипов два различных дочерних генотипа.

- Выбор функции оценки(фитнесс-функции).

Такая функция позволяет оценивать генотипы с точки зрения их близости к оптимальному решению и отбирать из них лучшие на каждой итерации.

1.2.1 Этапы работы алгоритма

- Создается нулевое поколение генотипов с использованием заданного алгоритма начального приближения
- Запускается итерационный процесс

Случайным образом выбирается часть поколения, которая подвергнется мутации

Выбранная часть пула генотипов мутируется с использованием заданного алгоритма мутации

Мутировавшие генотипы замещают собой исходные в поколении, не мутировавшие – остаются без изменений

Из пула генотипов выбираются пары для скрещивания

Производится скрещивание с использованием заданного алгоритма

С использованием заданной функции оценки из результатов скрещивания выбираются лучшие

Если выполнено условие останова - например, достигнут предел числа итераций или известный максимум, то итерационный процесс завершается, в противном случае начинается следующая итерация.

- Результат итерационного процесса отдается пользователю

1.2.2 Выбор параметров алгоритма в приложении к задаче о многомерном рюкзаке

Наиболее естественным кодированием возможного решения задачи о рюкзаке в генотип является бинарная последовательность длины N , состоящая из нулей и единиц. Каждый i -й элемент такой последовательности является индикатором вхождения i -го предмета в текущее решение. Такая модель требует наличия проверки корректности генотипа - соблюдения условия 1.1. Этот генотип позволяет однозначно восстановить фенотип и провести оценку.

Для генерации начального приближения был использован жадный алгоритм. Сначала создается генотип из единиц, соответствующий конфигурации рюкзака, в который положены все предметы. Затем в случайном порядке единицы заменяются на нули, пока полученный генотип не будет удовлетворять условию корректности 1.1. Если полученный генотип состоит только из нулей или только из единиц (отсутствие решения или тривиальный случай), то эволюционный процесс не запускается. В противном случае полученный генотип подвергается мутации до заполнения пула решений.

В ходе работы было реализовано несколько алгоритмов мутации и скрещивания с целью сравнения их эффективности. Были реализованы следующие алгоритмы мутации:

- Мутация в одной позиции, при которой инвертируется значение в одной случайно выбранной позиции генотипа (см. 1). Данный алгоритм наиболее прост в реализации и не требует дополнительных условий.
- Инверсионная мутация, при которой случайно выбранная половина генотипа заменяется на противоположные значения (см. 2). Данный алгоритм обеспечивает высокую изменчивость, что может ускорить процесс перебора возможных решений. С другой стороны, высокая изменчивость может стать причиной неустойчивости эволюционного процесса.

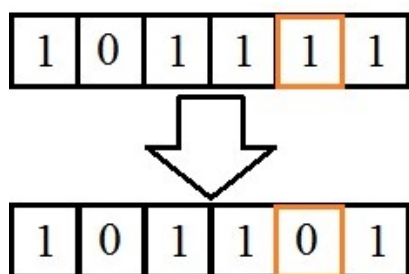


Рис. 1: Одноточечная мутация

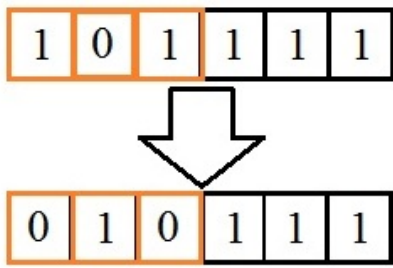


Рис. 2: Инверсионная мутация

Были реализованы следующие алгоритмы скрещивания:

- Скрещивание по одной точке, при котором выбирается произвольная позиция в последовательности генотипа, значения до позиции берутся из первого генотипа, на позиции и после - из второго (см. 3). Наиболее простое из возможных скрещиваний, позволяющее получить разнообразные особи при повторном применении для одинаковых родительских генотипов.
- Скрещивание по двум точкам, при котором выбираются две различные произвольные точки, значения внутри полученного интервала и в самих точках берутся из первого генотипа, вне интервала - из второго (см. 4). Обеспечивает ещё большую вариативность, чем одноточечное скрещивание, но требует проверки на расположение границ интервала друг относительно друга.
- Побитовое скрещивание, при котором значения на нечетных позициях берутся из первого генотипа, на четных - из второго (см. 5). Не требует проверок, но и не дает разнообразия генотипов при повторном применении для одинаковых родительских генотипов.

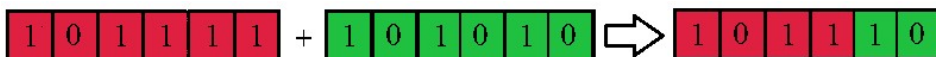


Рис. 3: Скрещивание по одной точке

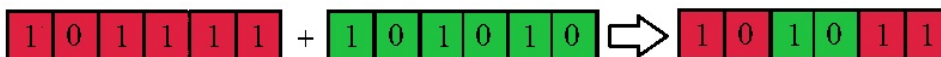


Рис. 4: Скрещивание по двум точкам



Рис. 5: Побитовое скрещивание

В качестве функции оценки используется стоимость всех предметов, содержащихся в рюкзаке-фенотипе, соответствующем генотипу.

1.2.3 Дополнения и улучшения, внесённые в генетический алгоритм

В связи со спецификой задачи и сложностями, возникшими при тестировании на практике, в алгоритм были внесены дополнения и улучшения.

Важнейшее дополнение – проверка и исправление генотипов на корректность после мутации и скрещивания. Если соответствующий генотипу рюкзак не удовлетворяет условию корректности 1.1, то производится принудительная коррекция генотипа: значения индикаторов начинают последовательно зануляться до достижения фенотипом корректности. Такая операция позволяет эволюционному процессу функционировать, но существенно замедляет процесс работы генетического алгоритма. Также эта коррекция негативным образом влияет на свойство наследственности генетического алгоритма, так как скорректированный генотип по определению не может кодировать решение лучше решения, заданного исходным генотипом.

Были реализованы два принципиальных улучшения алгоритма: пулы лучших генотипов и механика перезапуск эволюционного процесса.

Пул лучших генотипов для текущего эволюционного процесса хранит в себе лучшие с точки зрения оценки генотипы с начала текущей эволюции. Этот пул позволяет реализовать возврат к лучшим конфигурациям при существенном ухудшении стоимости текущего поколения, что происходит, в связи с необходимостью коррекций, достаточно часто.

Эволюционный процесс часто приводит к такой ситуации, когда разнообразие внутри поколения мало, а лучшая оценка генотипа внутри поколения не меняется – попадание в локальный максимум функции оценки. В немодифицированном генетическом алгоритме эта проблема решается за счет мутации, однако из-за различных причин такое решение может быть недостаточно эффективным. Одним из способов покинуть локальный максимум является перезапуск алгоритма. При таком перезапуске текущее поколение заменяется новым, созданным с использованием алгоритма начального приближения, после чего эволюционный процесс запускается вновь. Текущий пул лучших конфигураций при этом сохраняется отдельно – в глобальном пуле, который не взаимодействует с эволюционным процессом. При множественных перезапусках глобальный пул хранит информацию о лучших решениях за все время работы генетического алгоритма, что позволяет при необходимости получить неоптимальные решения, близкие к искомому.

2 Практическая часть

2.1 Наборы тестов

Для оценки эффективности работы алгоритма было проведено его тестирование на различных наборах тестов. Первый набор тестов взят из книги Петерсена(см. [5]) и содержит 7 задач. Условия этих задач(см таблицу 1) позволяют проверить эффективность простейшей версии генетического алгоритма без модификаций и оценить её эффективность.

№ задачи	Размерность	Количество предметов
1	6	10
2	10	10
3	15	10
4	20	10
5	28	10
6	39	5
7	50	5

Таблица 1: Параметры первого набора тестов

Второй набор тестов взят из статьи Чу (см [7]) и содержит в себе 30 задач с одинаковыми параметрами: размерность рюкзака равна 5, рассматривается 100 различных предметов. Для каждого набора тестов известно лучшее решение, эффективность алгоритма оценивалась по скорости поиска решения и числу итераций.

Для тестирования использовался компьютер с следующей аппаратно-программной базой:

- Процессор: Intel(R) Core(TM) i5-3230M CPU @ 2.60GHz (4 CPUs), 2.6GHz
- Количество оперативной памяти: 6144MB RAM
- Операционная система: Windows 10 Домашняя для одного языка 64-bit (10.0, Build 15063) (15063.rs2_release.170317-1834)

Для сборки использовалась программная платформа .NET Framework 4.5.2.

2.2 Первый набор тестов

Рассмотрим результаты работы алгоритма без модификаций на первом наборе тестов. Экспериментальным образом были подобраны количество конфигураций в пуле(10) и процент мутирующих особей(20)

2.2.1 Мутация в одной позиции

Данная мутация показала себя эффективной с точки зрения решения задачи - для всех возможных вариантов скрещивания во всех тестах найден оптимальный результат(см таблицы 2, 3, 4) При сравнении результатов скрещивания можно заметить, что алгоритм побитового скрещивания в большинстве случаев несколько уступает в скорости другим алгоритмам.

№ теста	Итерации	Время, мс
1	6	8
2	114340	2694
3	712	18
4	4022	92
5	3049	97
6	400974	13481
7	9343432	396644

Таблица 2: Одноточечная мутация, одноточечное скрещивание

№ теста	Итерации	Время, мс
1	21	9
2	70930	1697
3	901	22
4	2092	67
5	2672	95
6	23261	8181
7	4770456	199461

Таблица 3: Одноточечная мутация, двуточечное скрещивание

№ теста	Итерации	Время. мс
1	3	9
2	182633	3301
3	1455	49
4	7869	219
5	2198	74
6	289379	100133
7	348826	139403

Таблица 4: Одноточечная мутация, побитовое скрещивание

2.2.2 Инверсионная мутация

Инверсионная мутация в контексте данной задачи оказалась менее эффективной, для тестов 5-7 решения не были найдены за 20 минут для всех вариантов скрещивания (см табл 5, 6, 7). Неэффективность такого варианта мутации может быть обусловлена применением алгоритма коррекции после мутации, что сводит на нет какой-либо положительный эффект от её применения. Также можно отметить существенно более высокую скорость решения в сравнении с алгоритмом одноточечной мутации для теста 2.

№ теста	Итерации	Время. мс
1	3	6
2	10	2
3	58	2
4	9538	560
5	-	-
6	-	-
7	-	-

Таблица 5: Инверсионная мутация, одноточечное скрещивание

Таблица 6: Инверсионная мутация, двуточечное скрещивание

№ теста	Итерации	Время. мс
1	2	10
2	16	1
3	356	19
4	7161	422
5	-	-
6	-	-
7	-	-

№ теста	Итерации	Время. мс
1	10	10
2	23	1
3	1409	52
4	246001	12189
5	-	-
6	-	-
7	-	-

Таблица 7: Инверсионная мутация, побитовое скрещивание

2.2.3 Промежуточные выводы

Исходя из приведенных тестов, можно считать подтвержденной принципиальную возможность реализованного алгоритма быть использованным для решения задачи о многомерном рюкзаке. Наилолее эффективной себя показало двуточечное скрещивание.

2.3 Второй набор тестов

Для проверки эффективности второго набора тестов были выбраны параметры запуска алгоритма, указанные в таблице 8.

Вид скрещивания	Двуточечное
Вид мутации	Одноточечная
Размер пула генотипов	10
Процент мутирующих особей	20%
Максимальное число итераций	2000000
Величина отклонения для сброса алгоритма	0.01
Количество итераций до полного перезапуска	200000

Таблица 8: Параметры запуска второго набора тестов

Максимальное число итераций было введено с целью ограничить время работы алгоритма. В случае, если лучшее решение не найдено, программа выдает на печать расхождение лучшего из найденных решений с текущим.

Рассмотрим результаты работы программы на втором наборе тестов. Для успешных решений среднее время составило 315 с, среднее число итераций - 652 007 (см. [BigResults1])

№ теста	Количество итераций	Время, с
1	-	888
2	197915	44
3	-	915
4	-	812
5	-	860
6	-	970
7	1253381	619
8	83207	39
9	262645	255
10	-	963
11	189635	83
12	-	912
13	-	912
14	-	825
15	-	881
16	942021	484
17	-	864
18	4386	5
19	-	880
20	-	865
21	-	734
22	1735194	752
23	-	779
24	845324	428
25	-	812
26	908217	454
27	594220	281
28	36434	35
29	-	895
30	1423522	625

Таблица 9: Временная эффективность алгоритма

В 43.3% случаев генетический алгоритм решил задачу и во всех случаях расхождение найденного и действительного максимумов расходится не более чем на 0,0021 с оптимальным результатом в относительных значениях(см [BigResults2]).

№ теста	Относительное расхождение с максимумом
1	0,0021
2	0
3	0,0005
4	0.0022
5	0.0018
6	0.0014
7	0
8	0
9	0
10	0,0031
11	0
12	0,0020
13	0,0008
14	0,0001
15	0,0017
16	0
17	0,0006
18	0
19	0,0018
20	0,0005
21	0,0016
22	0
23	0,0008
24	0
25	0,0007
26	0
27	0
28	0
29	0,0006
30	0

Таблица 10: Эффективность в достижении максимума задачи

ЗАКЛЮЧЕНИЕ

В ходе работы были выполнены все поставленные задачи. Была создана программа, реализующая генетический алгоритм с требуемыми для решения задачи модификациями. Было проведено тестирование, показавшее высокую эффективность программы на различных наборах тестов. Для малых наборов предметов (50 предметов и менее) оптимальное решение было достигнуто в 100. Для больших наборов (100 предметов) оптимум был получен в 43,3. При этом среднее время поиска решения составляет 5 минут 25 секунд. Таким образом, алгоритм может эффективно применяться в решении практических проблем.

Список

ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Gavish B., Pirkul H.* Allocation of Databases and Processors in a Distributed Computing System // Management of Distributed Data Processing. – 1982. – № 31. – С. 215–231.
2. *Shih W.* A Branch and Bound Method for the Multiconstraint Zero-One Knapsack Problem // Journal of the Operational Research Society. – 1979. – № 30. – С. 369–378.
3. *Gilmore P., Gomory R.* The Theory and Computation of Knapsack Functions // Operations Research. – 1966. – № 14.
4. *Mathews G.* On the partition of numbers // Proceedings of the London Mathematical Society. – 1897. – № 28. – С. 486–490.
5. *C.C.Petersen* Computational experience with variants of the Balas algorithm applied to the selection of R&D projects // Management Science. – 1967. – 13(9). – С. 736–750.
6. *Alex F.* Computer Models in Genetics. – McGraw-Hill, 1970.
7. *P.C.Chu, J.E.Beasley* A genetic algorithm for the multidimensional knapsack problem // Journal of Heuristics. – 1998. – Т. 4. – С. 63–86.

ПРИЛОЖЕНИЯ

А Исходный код программы