

РЕФЕРАТ

Дипломная работа: 26 с., 0 рис., 10 табл., 7 источников. ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ, МАШИННОЕ ОБУЧЕНИЕ, NP-ПОЛНЫЕ ЗАДАЧИ, ЗАДАЧА О РЮКЗАКЕ, ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ
Объект исследования - задача о многомерном рюкзаке Цель работы - исследование применения генетического алгоритма к задаче о многомерном рюкзаке. В процессе работы был реализован генетический алгоритм на языке C# и его модификации. В результате исследования ...

ОГЛАВЛЕНИЕ

□□□.

РЕФЕРАТ	5
ВВЕДЕНИЕ	7
ГЛАВА 1 Теоретическая часть	8
1.1 Задача о многомерном рюкзаке	8
1.2 Генетические алгоритмы	9
1.2.1 Этапы работы алгоритма	10
1.2.2 Выбор этапов	11
1.2.3 Дополнения алгоритма	12
ГЛАВА 2 Практическая часть	14
2.1 Наборы тестов	14
2.2 Первый набор тестов	15
2.2.1 Мутация в одной позиции	15
2.2.2 Инверсионная мутация	15
2.2.3 Промежуточные выводы	19
2.3 Второй набор тестов	19
ЗАКЛЮЧЕНИЕ	24
□□□□□□ □□□□□□□□□□	25
ПРИЛОЖЕНИЕ А КОД ПРОГРАММНОГО ПРОДУКТА	26

ВВЕДЕНИЕ

В ходе развития компьютерных наук человечество 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. [5]), погрузка груза и контроль бюджета (см. [6]), задачи раскрой-ки (см. [7]) и др. Решение этих проблем обуславливает актуальность решения задачи о рюкзаке.

Цель исследования Для достижения цели были поставлены следующие задачи

- Исследовать
- Спроектировать
- Оценить

ГЛАВА 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. [1]) Данный же вариант задачи впервые был предложен Клиффордом Петерсеном в 1967 году.(см. [2])

Постановка задач такова

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

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

где $x_i \in \{0, 1\}$ при условии

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

И стандартная задача, и её модификация являются NP-полными задачами. Вычислительная сложность задачи такого рода при переборном решении для N предметов -

$$o(2^N)$$

, что, вкупе с NP-сложностью, делает алгоритмическое решение та-

кой 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 [3]) Алгоритм является итеративным. Генетический алгоритм моделирует естественные процессы эволюции популяции, а именно - мутацию и скрещивание. 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 однозначно моделировать сущность, рассматриваемую в задаче.

- Выбор начального приближения.

Для запуска итерационного процесса требуется создать начальное множество - пул генотипов. В зависимости от способа задания начального пула скорость поиска оптимального решения может меняться. Размер пула также оказывает влияние на эффективность алгоритма: слишком маленький пул подавляет разнообразие и приводит к попаданию в локальные максимумы, слишком большой - 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 образом изменяются их составляющие. Если мутирует слишком малая часть пула, то генетический алгоритм не сможет покинуть локальные максимумы, если же слишком большая - алгоритм теряет свойство сохранения признаков.

- Выбор механизма скрещивания (кроссинговера).

После мутации происходит создание новых генотипов из частей старых с сохранением признаков 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 пул генотипов с использованием заданного алгоритма начального приближения
- Запускается итерационный процесс

Случайным образом выбирается часть пула, которая подвергнется мутации

Выбранная часть пула генотипов мутируется с использованием заданного алгоритма мутации

Мутировавшие генотипы замещают собой исходные в пуле, немутировавшие 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 этапов

Наиболее естественным кодированием отдельного 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](#)

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 полученная конфигурация не будет удовлетворять условию корректности. После этого полученный генотип мутируется с помощью текущей мутации до заполнения пула решений.

В ходе работы было реализовано несколько алгоритмов мутации и скре-

щивания с целью сравнения их эффективности. Были реализованы следующие алгоритмы мутации:

- Мутация в одной позиции, при которой заменяется значение в одной случайно выбранной точке генотипа.
- Инверсионная мутация, при которой половина генотипа заменяется на противоположные значения.

Были реализованы следующие алгоритмы скрещивания:

- Скрещивание по 1 точке, при котором выбирается произвольная точка в последовательности генотипа, значения до точки берутся от первого генотипа, после - от второго.
- Скрещивание по двум точкам, при котором выбираются две различные произвольные точки, значения внутри интервала и в самих точках берутся из первого генотипа, вне интервала - из второго.
- Побитовое скрещивание, при котором значения на нечетных позициях берутся из первого генотипа, на четных - из второго.

В качестве функции оценки используется стоимость всех предметов, содержащихся в рюкзаке, соответствующем конфигурации.

1.2.3 Дополнения алгоритма

В связи со спецификой задачи в алгоритм были внесены дополнения.

Были введены проверки генотипов на корректность после мутации и скрещивания. Если генотип не удовлетворяет условию корректности, то значения начиная с первой позиции начинают зануляться до достижения генотипом корректности.

Были введены дополнительные пулы лучших конфигураций за время работы алгоритма. Такие пулы решают одновременно несколько задач

- Недопущение сильного ухудшения результатов решения вследствие случайных мутаций. При достаточно сильном ухудшении средней стоимости пула текущий пуз замещается текущим пулом лучших конфигураций
- Возможность сохранения результатов при перезапуске алгоритма с другим начальным приближением. Такой перезапуск оправдан при получении генотипа - локального максимума, покинуть который с помощью мутации невозможно. Перезапуск алгоритма происходит при продолжительном сохранении лучшего значения в пуле конфигураций неизменным.
- Возможность сравнения решений после окончания работы алгоритма

ГЛАВА 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. [2]) и содержит 7 задач. Условия этих задач(см таблицу 2.1) позволяют проверить эффективность простейшей версии генетического алгоритма без модификаций и оценить её эффективность.

□□□□□□□ 2.1 --- Параметры первого набора тестов

№ 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

Второй набор тестов взят из статьи Чу (см [4]) 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 [2.2](#), [2.3](#), [2.4](#)) При сравнении результатов скрещивания можно заметить, что алгоритм побитового скрещивания в большинстве случаев несколько уступает в скорости другим алгоритмам.

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 [2.5](#), [2.6](#), [2.7](#)). Неэффективность такого варианта мутации может быть обусловлена применением

□□□□□□□ 2.2 --- Одноточечная мутация, одноточечное скрещивание

№ 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

□□□□□□□ 2.3 --- Одноточечная мутация, двуточечное скрещивание

№ теста	Итерации	Время, мс
1	21	9
2	70930	1697
3	901	22
4	2092	67
5	2672	95
6	23261	818
7	4770456	199461

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.

□□□□□□ 2.4 --- Одноточечная мутация, побитовое 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

□□□□□□ 2.5 --- Инверсионная мутация, одноточечное скрещивание

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

□□□□□□ 2.6 --- Инверсионная мутация, двуточечное скрещивание

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

□□□□□□ 2.7 --- Инверсионная мутация, побитовое скрещивание

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

2.2.3 Промежуточные выводы

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

2.3 Второй набор тестов

Для проверки эффективности второго набора тестов были выбраны параметры запуска алгоритма, указанные в таблице 2.8.

□□□□□□□ 2.8 --- Параметры запуска второго набора тестов

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

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. [?])

□□□□□□□ 2.9 --- Временная эффективность алгоритма

№ 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

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

□□□□□□ 2.10 --- Эффективность в достижении максимума задачи

№ 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

+ улучшения

ЗАКЛЮЧЕНИЕ

В ходе работы был реализован алгоритм, решающий задачу... получили такие-то результаты, такие-то методические рекомендации

□□□□□□ □□□□□□□□□□

1. Mathews, G. B. On the partition of numbers / G.B. Mathews / / Proceedings of the London Mathematical Society. 28: C. 486–490.
2. C.C.Petersen "Computational experience / with variants of the Balas algorithm applied to the selection of R&D projects" Management Science 13(9) (1967) 736-750.
3. Fraser Alex. Computer Models in Genetics. — New York: McGraw-Hill, 1970. — ISBN 0-07-021904-4.
4. P.C.Chu and J.E.Beasley "A genetic algorithm for the multidimensional knapsack problem-/ Journal of Heuristics, vol. 4, 1998, C. 63-86.
5. Gavish, B. and H. Pirkul. (1982). "Allocation of Databases and Processors in a Distributed Computing System." In J. Akoka (ed.) Management of Distributed Data Processing, North-Holland, c. 215–231.
6. Shih, W. (1979). "A Branch and Bound Method for the Multiconstraint Zero-One Knapsack Problem," Journal of the Operational Research Society 30, 369–378.
7. Gilmore, P.C. and R.E. Gomory. (1966). "The Theory and Computation of Knapsack Functions," Operations Research 14, 1045–1075.

ПРИЛОЖЕНИЕ А

КОД ПРОГРАММНОГО ПРОДУКТА

```
namespace Hellenist {  
    public enum Case {  
        Nominative = 1,  
        Vocative = 2,  
        Accusative = 3,  
        Genitive = 4,  
        Dative = 5  
    }  
}
```