

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ РЕШЕНИЯ КВАДРАТИЧНОЙ ЗАДАЧИ О НАЗНАЧЕНИЯХ

Е.М. Васильев, И.В. Крутских

Рассматриваются этапы решения квадратичной задачи о назначениях генетическим алгоритмом на основе мутагенеза. Подробно раскрыты процедуры генной и хромосомной мутаций этого алгоритма и исследовано влияние их настраиваемых параметров на сходимость поиска. В результате анализа решений тестовых задач tai12a – tai20a обсуждена проблема достижимости глобального экстремума и предложено построение алгоритма на базе матричной репликации. Представлена проверка эффективности репликационного подхода к имитации эволюционного процесса

Ключевые слова: задача о назначениях, генетический алгоритм, мутагенез, матричная репликация

1. Постановка задачи

Рассмотрим квадратичную задачу о назначениях (QAP – the Quadratic Assignment Problem) [1]:

$$Z = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{p=1}^n c_{ij} q_{kp} x_{ik} x_{jp} \rightarrow \min, \quad (1)$$

в которой: n – количество объектов и пунктов их назначения; c_{ij} – затраты на передачу единицы ресурсов из пункта i в пункт j ; q_{kp} – объём ресурсов, направляемых от объекта k к объекту p ;

$$x_{ik} = \begin{cases} 1 & \text{при назначении объекта } k \text{ в пункт } i; \\ 0, & \text{иначе;} \end{cases}$$

$$x_{jp} = \begin{cases} 1 & \text{при назначении объекта } p \text{ в пункт } j; \\ 0, & \text{иначе,} \end{cases}$$

$i = \overline{1, n}$; $k = \overline{1, n}$; $j = \overline{1, n}$; $p = \overline{1, n}$; Z – затраты, необходимые для обмена ресурсами между объектами.

Поставленная задача является математической моделью многих реальных прикладных задач, возникающих при проектировании информационных систем, размещении промышленных объектов и оборудования, выборе конфигураций сетей электропитания [2]. Явно выраженный комбинаторный характер этой задачи, общее количество вариантов решения которой определяется числом перестановок $R(n)=n!$, превосходящим по скорости роста экспоненциальную зависимость, а также наличие существенных внутрисистемных связей, выражающихся в том, что вклад назначения каждого объекта в общие затраты зависит от назначения всех других объектов, привели к тому, что точные методы решения QAP удаётся применить для размерностей n задачи не превышающей 20, и приходится использовать приближённые методы, способные за приемлемое время получать решение, близкое к оптимальному [3]. Наиболее известными из таких методов являются генетические алгоритмы, хорошо зарекомендовавшие себя при решении различных задач оптимизации [4], однако успехи их применения в QAP остаются достаточно скромными.

Причина такого положения заключается, на наш взгляд, в излишне приписываемой этим алгоритмам уникальности настроек для решения каждой конкретной задачи и, как следствие, недостатке конкретных рекомендаций по их использованию.

В предлагаемой работе подробно излагаются процедуры генетического алгоритма на примере решения QAP, количественно исследовано влияние его настраиваемых параметров на сходимость, выявлены пределы этой сходимости в задачах NP-класса и обсуждены перспективы применения.

Указанные вопросы обсуждаются в следующем порядке:

общая структура алгоритма;

исследование механизма генной мутации;

исследование хромосомной мутации;

анализ влияния на сходимость объёма популяции;

проблема достижения глобального экстремума; исследование матричной репликации.

2. Общая структура алгоритма

Поскольку каждый вариант решения задачи (1) соответствует некоторому варианту перестановки без повторений, то в генетической интерпретации многообразие этих вариантов может быть обеспечено хромосомой H^r , содержащей n генов h_i^r , $i = \overline{1, n}$, каждый из которых может занимать любой из n локусов хромосомы, причём все гены имеют по одной аллелоформе.

Таким образом, модель задачи можно представить множеством хромосом (наборов генов) $H^r = \{h_1^r, h_2^r, \dots, h_n^r\}$, совокупность $r = \overline{1, R}$ которых определяет собою общее число возможных генотипов (рис. 1).

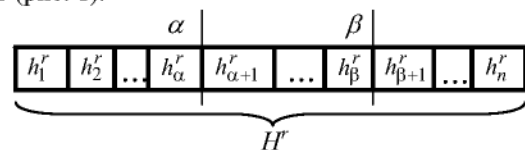


Рис.1. Генетическая модель задачи о назначениях

Различные особи не отличаются друг от друга составом своих хромосом, и для воспроизводства потомков F в данной задаче возможен единственный механизм реализации эволюции – мутагенез.

Глубина внутривхромосомной перестройки в ходе эволюции определялась количеством локусов, участвующих в мутациях:

генная мутация предусматривала случайный выбор нескольких пар локусов и обмен размещающихся в них аллелоформ;

хромосомные мутации осуществлялись выбором точки α рекомбинации и длины $\ell = \beta - \alpha$ отрезка хромосомы, подлежащего инвертированию.

Таким образом, задача построения эффективного генетического алгоритма на основе мутагенеза сводится к выбору процедур внутривхромосомных перестановок.

Наиболее общим подходом к выбору этих процедур является сочетание взаимодополняющих мутаций отдельных генов (генной мутации) и мутаций участков хромосом (хромосомной мутации), обеспечивающее как небольшие, так и крупные изменения генома. Соответствующая минимальная структура алгоритма представлена на рис. 2.

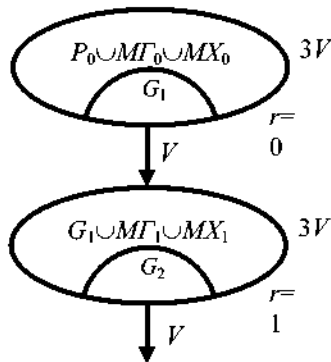


Рис. 2. Общая структура алгоритма

Содержание указанных процедур:

1. Формирование V элементов исходной (родительской) популяции P_0 , каждой особи которой соответствует одна хромосома со случайным равномерным порядком размещения генов.

2. Независимая генная мутация случайного количества n_2 ($1 \leq n_2 \leq n_{2\max}$) пар генов в хромосоме каждой исходной особи путём перестановки генов в каждой из n_2 пар. В результате применения генной мутации к V особям множества P_0 формируется множество MG_0 мутантов мощностью $|MG_0| = V$.

3. Независимая хромосомная мутация каждой исходной особи путём инверсии отрезка хромосомы длиной $2 \dots n_{\max}$ ($2 \leq n_{\max} \leq n - 2$) с его случайным расположением относительно границ хромосомы. В результате применения хромосомной мутации к каждой особи из множества P_0 формируется множество MX_0 мутантов мощностью $|MX_0| = V$.

4. Формируется множество $G_1 \subset P_0 \cup MG_0 \cup MX_0$ из V лучших особей, $|G_1| = V$.

5. Множество G_1 принимается в качестве исходного P_0 , и алгоритм повторяется с п. 2 до тех пор, пока не выполнится условие окончания поиска.

Пример поиска, реализованного описанным алгоритмом с произвольно взятыми параметрами

$V=100$, $n_{2\max}=6$, $n_{\max}=12$ для задачи (1) с $n=12$ (табл. 2а) представлен на рис. 3 (график 1).

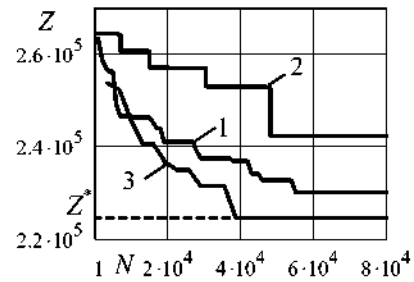


Рис. 3. Сравнение хода поиска алгоритмами: 1 – генетический с неоптимальными настройками; 2 – равномерный поиск; 3 – генетический с оптимальными настройками

Сравнение хода поиска (рис. 3 (график 1)) с движением к экстремуму $Z^* = 224416$ при равномерном случайном формировании особей (рис. 3 (график 2)) подтверждает эффективность генетического алгоритма даже с указанными произвольными настройками. Однако, учитывая, что в результате 15-кратного повторения поиска с количеством вычислений функции цели $N = 20 \cdot 10^4$ экстремум Z^* достигался всего 2...4 раза, была поставлена задача улучшить сходимость алгоритма путём оптимизации его параметров:

- максимальное количество пар генов $n_{2\max}$, участвующих в генной мутации;
- максимальное количество генов n_{\max} , участвующих в хромосомной мутации;
- объём V популяции.

3. Исследование механизма генной мутации

Генные мутации осуществляются равномерным случайным выбором номеров генов, обмениваемых местами в хромосоме (рис. 4).

Процедура генной мутации оставляет возможность выбора оптимального параметра её проведения: количества пар n_2 генов, участвующих в мутагенезе на каждом шаге. Наличие такой возможности объясняется для генной мутации, тем, что с ростом n_2 возрастает число обменов и генные мутации по своему эффекту начинают приближаться к хромосомной, т.е. утрачивают своё предназначение.

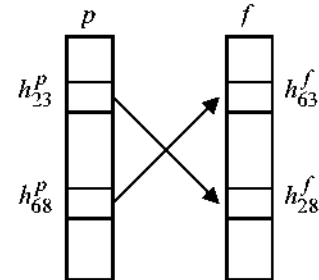


Рис. 4. Иллюстрация генной мутации

На рис. 5 показаны значения функции цели Z , полученные при поиске экстремума для целочисленных комбинаций n_2 и n_{\max} с разным количеством

N вычислений критерия Z . Размерность задачи $n=15$ (tail5a); для каждой комбинации n_z и $n_{\text{хmax}}$ поиск повторялся 15 раз с последующим вычислением математического ожидания критерия.

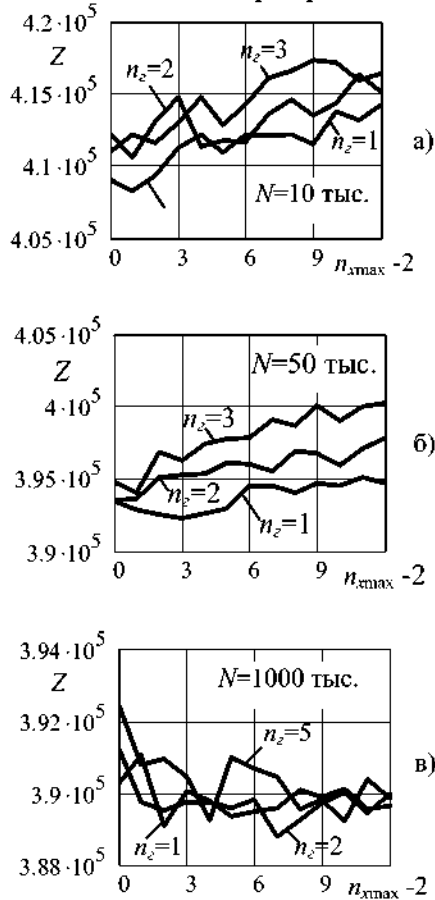


Рис. 5. Определение оптимальных параметров генной мутации

Анализ рис. 5 показывает, что по всей глубине поиска достаточна генная мутация одной пары генов.

4. Исследование хромосомной мутации

Иллюстрация механизма инвертирования при хромосомных мутациях представлена на рис. 6.

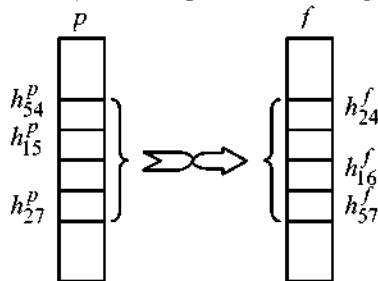


Рис. 6. Иллюстрация хромосомной мутации

Возможность выбора некоторого оптимального числа $n_{\text{хmax}}$ генов, участвующих в хромосомных мутациях, определяется тем обстоятельством, что очень малое или очень большое число $n_{\text{хmax}}$ в первом случае неэффективно, а во втором – слишком сильно расшатывает наследственные признаки. Так, из

рис. 5а-в следует, что длину $n_{\text{хmax}}$ хромосомной мутации целесообразно увеличивать по мере приближения к экстремуму, например, для задачи tail5a с $n_{\text{хmax}}=2$ до $n_{\text{хmax}}=8$. Для каждого значения $n_{\text{хmax}}$ проведено 15 экспериментов, выявлено влияние $n_{\text{хmax}}$ на сходимость (рис. 7) и найден оптимальный диапазон $n_{\text{хmax}}=5...6$.

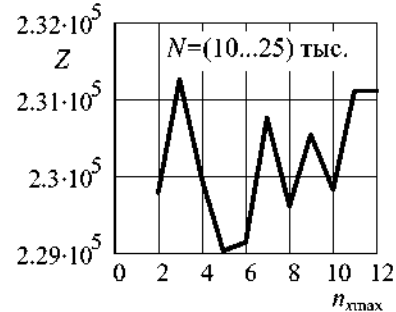


Рис. 7. Исследование влияния параметров хромосомной мутации

5. Анализ влияния объема популяции

Для оптимальных значений параметров $n_{\text{хmax}}=1$ и $n_{\text{хmax}}=5$ для tail5a проведён анализ влияния на сходимость объема V популяции (рис. 8).

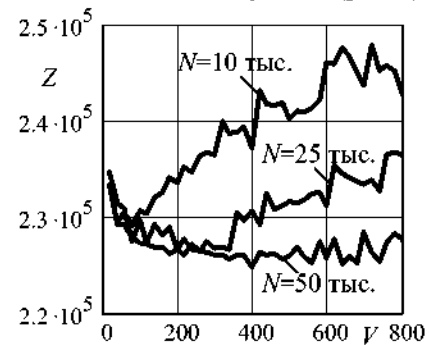


Рис. 8. Анализ влияния объема V популяции

Из рис. 8 следует, что оптимальное значение V связано с ограничением на количество вычислений N : при недостаточно большом N алгоритм не в состоянии реализовать необходимое для достижения экстремума число эволюций. Для указанных настроек следует принять $V=400$, при этом количество вычислений должно быть не менее 100 в расчёте на одну особь, т.е. при популяции V особей, получим необходимое количество вычислений $N \geq 100V$.

При найденных оптимальных настройках глобальный экстремум в задачах с $n \leq 15$ достигался в 80...100% поисков, однако с повышением размерности до $n=20$ этот показатель не превышал 5% ни при каких значениях указанных параметров, и возникла необходимость обсуждения проблемы достижения глобального экстремума.

6. Проблема достижения глобального экстремума

Малые изменения генома в результате генных мутаций не обязательно соответствуют малым

изменениям фенотипа (критерия задачи), так же, как и инверсия большого участка хромосомы не всегда приводит к существенным изменениям критерия.

Для количественной оценки изменения генома удобно использовать кодовое расстояние между исходной хромосомой p и её мутантом f :

$$d = \sum_{i=1}^n \sum_{j=1}^n w_{ij}; \quad w_{ij} = \begin{cases} 1, & \text{если } h_{ij}^p \neq h_{ij}^f; \\ 0, & \text{иначе,} \end{cases}$$

где h_{ij}^p , h_{ij}^f – гены i в локусе j хромосомы p и f .

Максимальное количество вариантов Per_{1z} , Per_{1x} перестановок, возможных в результате одной геной и одной хромосомной мутации, одинаково:

$$Per_{1z} = Per_{1x} = \frac{n(n-1)}{2}.$$

За один цикл последовательных процедур геной и хромосомной мутаций возможен выбор одного из Per_1 вариантов; $Per_1 = Per_{1z} Per_{1x}$:

$$Per_1 = \left(\frac{n(n-1)}{2} \right)^2,$$

которые составляют малую часть общего многообразия $R=n!$. Например, для $n=10$ эта часть равна $Per_1/R=0,00056$. При таком соотношении Per_1 и R достижение экстремума возможно только за счёт направленности поиска, реализуемого отбором и сохранением лучших особей с их последующим использованием в качестве исходного генетического материала для мутагенеза. Вместе с тем, направленный поиск сам по себе не обеспечивает достижение глобального экстремума. Если рассматривать проблему сходимости к глобальному экстремуму на уровне процедур алгоритма, то можно выделить два обстоятельства, препятствующие её разрешению.

1. Полное вытеснение локально лучшими вариантами решений худших на текущем этапе поиска, но перспективных вариантов. Для смягчения этого обстоятельства можно увеличить объём популяции V , а также исключить из её состава особи с одинаковыми значениями Z .

2. Варианты решений с малым расстоянием d от оптимального могут характеризоваться значительным удалением критерия Z от Z^* . На рис. 9 представлены значения Z в задаче tai20a для 190 вариантов, отстоящих от оптимального на расстояние $d=2$.



Рис. 9. Анализ отклонения критерия Z от оптимального значения $Z^*=703482$ в вариантах решений с $d=2$

Значительная удалённость Z от экстремального значения Z^* , достигающая 10% от Z^* , не только способствует вытеснению этих вариантов из популяции V , но делает бессмысленным понятие направленного движения к оптимуму в обширной по своим размерам окрестности глобального экстремума (движение при $d=2$ соответствует одному шагу геной мутации, в которой выбор пары переставляемых генов осуществляется равновероятным образом).

Указанные обстоятельства процедурного уровня носят принципиально неустранимый характер, и для их преодоления необходимо перейти на другой – информационный уровень рассмотрения процесса мутагенеза.

Воспользуемся данным Г. Кастлером [5] определением эволюционного процесса как процесса создания информации путём запоминания случайного выбора. Из такого понимания процесса эволюции вытекает необходимость пространственного разделения особей для фиксации их генотипов, т.е. обеспечения условий их независимого развития.

Такой подход к процессу эволюции как к процессу упорядочения – создания информации получил своё развитие в работе М. Эйгена [6], в которой выдвигается тезис о том, что любая последовательность эволюционирующих особей обладает способностью инструктировать своё дальнейшее построение. Этот механизм получил название матричной репликации, позволяющей объяснить самоорганизацию материи на таком высоком уровне энтропии, как уровень возникновения жизни.

Применим его к рассматриваемой задаче.

7. Исследование матричной репликации

В рамках решаемой задачи механизм матричной репликации был реализован следующим образом.

1. Выбирается размер r матрицы репликации, т.е. количество генов, положение которых в хромосоме заведомо совпадает с оптимальным. Это значение определяет собою начальную селективную избирательность особей и размеры популяции V .

2. Выбираются, (в общем случае, произвольно), номера генов, включаемых в матрицу репликации. Если в QAP есть некоторые объекты, выделяющиеся по своей значимости из их общего перечня, т.е. обладающие высокими структурообразующими свойствами, то эти объекты целесообразно включить в матрицу репликации.

3. Регулярным образом формируются особи, содержащие выбранную матрицу. Например, для задачи с $n=20$ и размером матрицы $r=1$ потребуется 20 особей, в одной из которых выбранный ген обязательно займёт оптимальный локус. Такие особи составят исходную популяцию V , объём которой определится выражением:

$$V = \frac{n!}{(n-r)!}.$$

Очевидно, что с ростом r повышается селективная избирательность матрицированных особей,

приближая их к так называемой точке невозврата, но растёт объём V и, как следствие, возрастает трудоёмкость алгоритма. Например, для $n=20$ и $r=5$ получим $V=1860480$ особей, каждая из которых должна будет пройти одинаковое число шагов эволюции.

Остальные $n-r$ локусов хромосомы заполняются равновероятным случайным образом генами, не вошедшими в матрицу.

4. Для каждой особи популяции V осуществляются независимые последовательности процедур генной мутации, реализующие их раздельную эволюцию.

При детерминированном подходе к процессу мутагенеза гены матрицы репликации не подвергаются мутациям.

Однако на практике, в условиях ограниченных вычислительных и временных ресурсов оказалось, что частота завершения эволюций глобальным экстремумом в несколько раз выше, если допустить участие матричных генов в общем процессе мутагенеза.

Этот результат очевидно вытекает из того обстоятельства, что из всей начальной популяции V только одна полная матрица длиной r занимает оптимальное расположение в хромосоме, и её сохранение возможно только при условии сохранения матриц в других особях – с неоптимальным расположением. Это приводит к тому, что для всех начальных особей, кроме одной – заведомо оптимальной, заранее предопределено неудачное завершение эволюционного процесса. В результате, эффективность алгоритма в статистическом смысле существенно снижается.

Поскольку процесс репликации по своей природе является случайным, то при отборе одного из вариантов решения, соответствующего родителю или мутанту, допускалась малая уступка E_z – замена родителя со значением Z_p мутантом с худшим значением $Z_j < Z_p$, при условии, что $Z_p - Z_j \leq E_z$.

В таблице представлены результаты сравнительного исследования обычного мутагенеза и различных модификаций алгоритма с матричной репликацией.

В каждом исследовании была реализована эволюция равных по объёму популяций, содержащих 332560 особей с $n=20$. В алгоритмах с матричной репликацией выбран размер матрицы $r=4$.

Исследование алгоритма с матричной репликацией

Модификация алгоритма		Количество оптимальных особей	
Общий мутагенез без репликации		2	
Матричная репликация	Общий мутагенез	$E_z=0$	11
		$E_z=0,001Z$	7
		$E_z=0,01Z$	1
	Генная мутация	$E_z=0$	8
		$E_z=0,001Z$	22
		$E_z=0,01Z$	6

Из таблицы следует, что применение механизма матричной репликации даже без введения уступки E_z обеспечивает рост числа сформированных глобально оптимальных особей с 2 до 11. Введение малой уступки $E_z=0,001Z$ увеличивает это преимущество вдвое.

Литература

1. Pardalos P. The quadratic assignment problem: A survey and recent developments / P. Pardalos, F. Rendl, H. Wolkowicz // DIMACS. – 1994. – Vol. 16. – P. 1 – 42.
2. Brixius N.W. The Steinberg Wiring Problem / N.W. Brixius, K.M. Anstreicher // Fest-Schrift in Honor of Manfred Padberg's 60th Birthday, SIAM. – 2003. – P. 331 – 348.
3. Романовский И.В. Дискретный анализ / И.В. Романовский. – С.-Пб.: Невский диалект, 2000. – 240 с.
4. Батищев Д.И. Генетические алгоритмы решения экстремальных задач / Д.И. Батищев. – Воронеж: ВГТУ, 1995. – 69 с.
5. Кастлер Г. Возникновение биологической организации / Г. Кастлер. – М.: Мир, 1967. – 90 с.
6. Эйген М. Самоорганизация материи и эволюция биологических макромолекул / М. Эйген. – М.: Мир, 1973. – 224 с.

Воронежский государственный технический университет

GENETIC ALGORITHM FOR THE QUADRATIC ASSIGNMENT PROBLEM

E.M. Vasilyev, I.V. Krutskih

Stages of the decision of the quadratic assignment problem by genetic algorithm for a basis mutagenesis are considered. Procedures of gene and chromosomal mutations of this algorithm are in detail opened and influence of their adjusted parameters on convergence of search is investigated. As a result of the analysis of decisions of test tasks tai12a - tai20a the problem of accessibility of a global extremum is discussed and construction of algorithm is offered based on matrix replication. Check of efficiency of the offered approach to imitation of evolutionary process is submitted

Key words: quadratic assignment problem, genetic algorithm, mutagenesis, matrix replication

