

УДК 681.31.00

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МОДЕЛИРОВАНИЯ ЭВОЛЮЦИИ ДЛЯ ОПТИМИЗАЦИИ ДОКУМЕНТООБОРОТА НА ПРЕДПРИЯТИИ

Ю.О. Чернышев, А.В. Басова, П.А. Панасенко, В.И. Поляков

Предлагается новый подход к созданию информационных систем управления на предприятии. Для решения актуальных задач нелинейной оптимизации разработан генетический алгоритм, позволяющий существенно повысить качество решения и быстродействие. Описана разработанная процедура инициализации начальной популяции. Предложены модифицированные операторы кроссинговера, мутации, обоснован новый подход к отбраковке нежизнеспособных решений. Предлагается введение в генетический алгоритм макроэволюции, т.е. развития нескольких субпопуляций. Отмечено, что работа не с одной, а с несколькими начальными популяциями позволяет получить большее разнообразие генетического материала и вести эволюцию в отдельных субпопуляциях.

Ключевые слова: оптимизация, генетический алгоритм, информационные системы, транспортная задача.

Введение

Современный подход к управлению на предприятии предполагает создание информационных систем управления, включающих механизмы сбора, обработки, передачи и использования информации.

Первым шагом в решении этой проблемы является создание адекватной модели, которая упростит и удешевит процесс исследования оригинала. В качестве модели обработки и передачи информации может быть использована графо-аналитическая модель вычислительного процесса [1]. Следующий шаг – оптимизация модели и использование полученных знаний для повышения эффективности реальной работы на предприятии. Создание информационных механизмов в настоящий момент становится все более актуальным, необходимым звеном процесса принятия решений во всех областях человеческой деятельности. В основе различных методов принятия решений лежит математический аппарат, позволяющий осуществить оптимизацию и найти лучшее решение из возможных. В этой связи необходимо отметить, что выбор метода оптимизации – сложнейшая задача, и если ее решить неверно, то это повлечет за собой существенные временные и материальные затраты. В работе предложен генетический алгоритм оптимизации документооборота на предприятии.

Постановка задачи

Организационная структура работает эффективно, если решены задачи документального (информационного) обеспечения. Оценку эффективности работы такой структуры можно свести к оценке информационной загрузки отделов предприятия.

В качестве основной информационной единицы можно выбрать как стандартную единицу измерения шенноновской информации, так и более наглядную – «документ» – основную информационную единицу внутри информационной системы (ИС). Типы документов и их характеристики могут быть получены путем анализа документооборота конкретного предприятия. Функции предприятия могут быть раздроблены на более примитивные составляющие – функциональные единицы, причем множество функциональных единиц имеет взаимно однозначное соответствие множеству документов.

Основным фактором, обеспечивающим эффективность любого предприятия, является оптимизация затрат на выполнение функциональных единиц и на обеспечение документооборота.

Задачу оптимизации информационных потоков предприятия можно свести к транспортной задаче (ТЗ) путем введения условной единицы передачи документов. Основное отличие предлагаемой постановки ТЗ от классической заключается в том, что передача документов ограничена заданной структурой информационных связей предприятия. Другое отличие состоит в том, что возможная исходная несбалансированность ликвидируется не путем введения фиктивных элементов, а путем изменения фактической структуры (введение реальных потребителей документов). В рассматриваемой трактовке ТЗ распределение транспортируемых единиц (документов) определяется заданной структурой системы.

Математическая модель организации документооборота на предприятии

Имеется n источников документов и m потребителей документов. Для каждого источника i задан объем производства информации $a[i] > 0$, а для каждого потребителя j – объем потребления информации $b[j] > 0$. Стоимость передачи условной единицы документооборота между i -м источником и j -м потребителем равна $c[i, j]$. Таблица издержек $C[A, B]$ задана. Мы стремимся к тому, чтобы суммарная стоимость издержек была бы минимальной.

В данном случае, как упоминалось выше, рассматривается закрытая задача, в которой отсутствуют нерасширяемые очереди в структурных элементах:

$$\sum_{i=1}^n a[i] = \sum_{j=1}^m b[j].$$

На практике целевая функция зачастую является вогнутой, что связано прежде всего с так называемым законом экономии от масштаба, в соответствии с которым удельные затраты на передачу информации убывают с ростом объемов перевозок [2]. При этом целевая функция имеет вид $y(x) = A \cdot \alpha$, где A – константа, а $\alpha < 1$.

Другой причиной, обуславливающей вогнутость целевой функции, является наличие фиксированных доплат. Эта функция равна сумме постоянных и линейных затрат, зависящих от мощности для любого положительного ее аргумента, и нулевым затратам при нулевой мощности. К этому виду относятся обычно затраты на проектирование и подготовку технической документации по созданию объекта, изготовлению макетов и опытных образцов, затраты на подготовительные работы и т.д. В частности, можно оценить затраты на создание и функционирование ИС предприятия:

$$y(x) = \begin{cases} 0, & \text{если } x = 0, \\ h1 + h2 \cdot x, & \text{если } x > 0 \end{cases},$$

где $h1$ – фиксированные затраты (денежные издержки на создание ИС); $h2$ – текущие затраты.

Зачастую в реальных условиях может возникнуть комбинированная задача, объединяющая оба описанных случая. Во многих практических ситуациях нелинейность целевой функции существенна. В связи с этим аппроксимация ее линейной функцией приводит к решению, далекому от истинного оптимума, и является неприемлемой.

Таким образом, можно сделать вывод о необходимости решения нелинейной транспортной задачи

$$\sum_{i=1}^n \sum_{j=1}^m c_{ij}(x_{ij}) \longrightarrow \min,$$

где $c_{ij}(x_{ij})$ – вогнутая ценовая функция,

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix},$$

x_{ij} – количество информации, передаваемой от i -го источника j -му потребителю. Каждая матрица решения X должна удовлетворять следующим ограничениям:

$$\sum_{i=1}^n x_{ij} = b_j, \quad j = 1, 2, \dots, m,$$

$$\sum_{j=1}^m x_{ij} = a_i, \quad i = 1, 2, \dots, n,$$

$$x_{ij} \geq 0, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m.$$

Было предпринято множество попыток эффективно решить данную задачу, так как она и по сей день является актуальной с практической и интересной с теоретической точек зрения. Перспективным представляется применение для решения оптимизационных задач методов моделирования процессов, протекающих в живой природе [3–5].

Структура генетического алгоритма решения ТЗ

Опишем базисную структуру разработанного генетического алгоритма. В начале работы генерируется множество решений (популяция) случайным, направленным или случайно-направленным способом. Решения представляются в виде закодированных хромосом. Способ кодирования определяется спецификой задачи. Далее, для каждого решения с помощью моделирующей функции оценивается его качество. Эта функция также основана на знаниях из той области, где решается задача. Инструмент селекции выбирает индивидуальности для скрещивания (кроссинговера). Операция мутации применяется случайно с некоторой вероятностью P_k . Затем формируется новая популяция, и процесс повторяется.

Генетический алгоритм должен иметь критерий остановки. Можно завершить работу после фиксированного числа генераций и взять лучшее из имеющихся решений. Можно остановиться, когда все организмы в популяции тождественны, т.е. $f_i = f_j$ для всех i и j или $|f_i - f_j| < \varepsilon$, где ε – наперед заданная малая положительная величина. Выбор критерия зависит от целей и задач, стоящих перед исследователем.

Поясним приведенную схему, рассмотрев более подробно каждый ее блок.

При выборе стратегий формирования начальной популяции следует руководствоваться целями, которые обеспечили бы наилучшие условия для дальнейшего развития эволюции. Сгенерированная начальная популяция должна содержать индивиды, принадлежащие к множеству D допустимых решений задачи.

В разработанном алгоритме (рисунок) формирование стартовой популяции решений $X^0 = x_1^0 \{x_1^0, x_2^0, \dots, x_n^0\}$ осуществляется с помощью механизма последовательного насыщения строк и столбцов, используемом в методе северо-западного угла и других методах [6]. Предложенная процедура отличается случайным выбором номера первого элемента, с которого начинается очередной цикл алгоритма инициализации, не опираясь на его северо-западность.

Пусть заданы a_1, a_2, \dots, a_n – объемы источников документов и b_1, b_2, \dots, b_m – объемы потребления документов. Пусть множество E_0 состоит из элементов $\{1, 2, \dots, n\}$, а множество F_0 – из элементов $\{1, 2, \dots, m\}$.

Функция оценки годности (моделирующая функция) естественным образом записывается как целевая функция исходной задачи

$$F(x) = \sum_{i=1}^n \sum_{j=1}^m c_{ij} \cdot x_{ij},$$

где $c_{ij} \in C(x)$, а $C(x)$ – ценовая функция (вогнутая).

В процессе решения задачи необходимо минимизировать моделирующую функцию. Чем меньше значение функции $F(x)$, тем более ценным является решение x .

Сортировка популяции производится на основе полученных знаний. После того, как процедура сортировки будет закончена, получаем популяцию

$$X = \{X1, X2, X3, \dots, XS\},$$

в которой $F(X1) \leq F(X2) \leq \dots \leq F(XS)$, где S – размер популяции.

Отсев неудачных особей после получения потомков является необходимой процедурой. Он формирует из общего множества родителей и потомков популяцию следующего поколения более высокого качества и направляет весь процесс эволюции в нужное русло. Основной классической процедурой отбора является пропорциональный отбор по значению целевой функции индивидов. Основную проблему отсева представляет нахождение компромисса между разнообразием генетического материала и качеством индивидов в популяции.

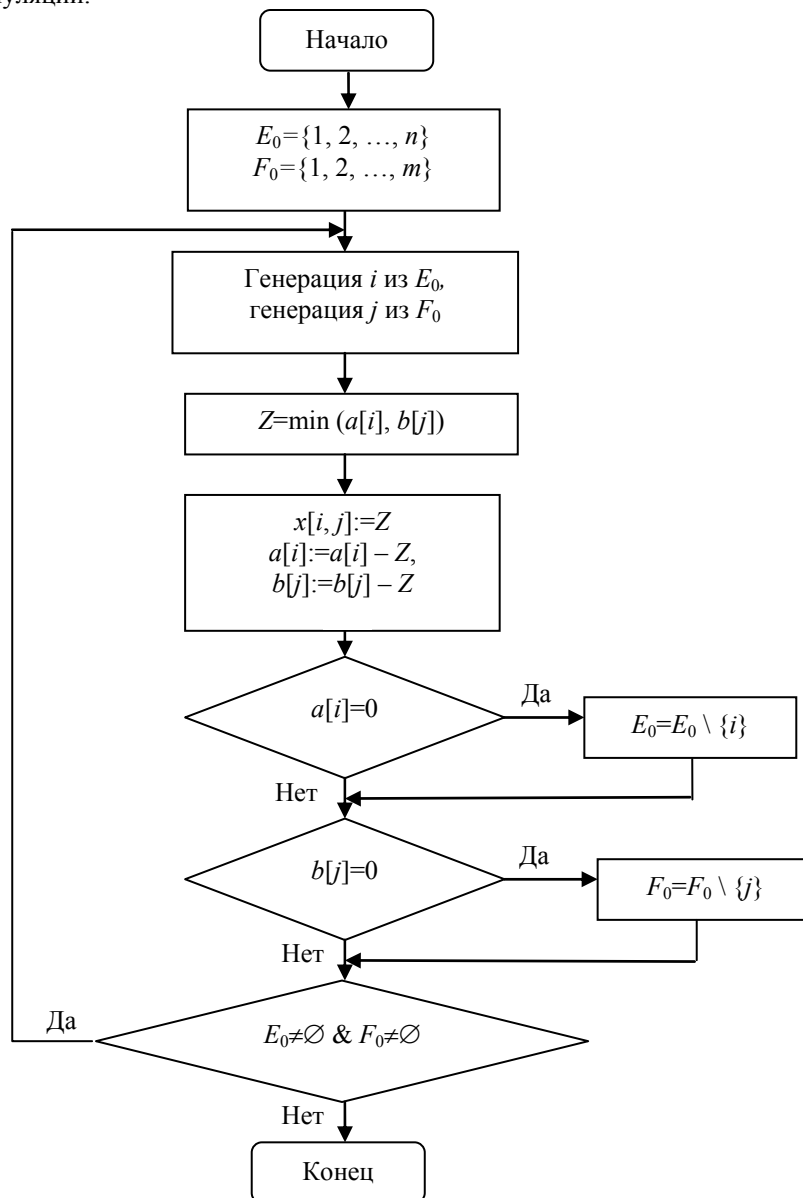


Рисунок. Блок-схема алгоритма формирования начальной популяции

В настоящей работе предлагается введение в генетический алгоритм макроэволюции, т.е. развития нескольких субпопуляций. Сложность этого подхода заключается в том, что после некоторого числа поколений хромосомы в отдельной популяции становятся очень похожими. Работа не с одной, а с несколькими начальными популяциями позволяет получить большее разнообразие генетического материала и возможность вести эволюцию в отдельных субпопуляциях, в каждой из которых можно получить решения, недостижимые в одной популяции.

Если в пределах популяции свободное скрещивание особей нарушается на длительное время, то в процесс эволюции включается третий фактор – изоляция. В результате действия изоляции образуются генотипически отличающиеся друг от друга субпопуляции. Скрещивание представителей этих субпопуляций оказывает положительный эффект на процесс эволюции. Включение в схему генетического алгоритма факторов изоляции и миграции при решении задач оптимизации позволит во многих случаях выходить из локальных оптимумов.

Апробация разработанного алгоритма

Для оценки эффективности разработанного алгоритма сравним полученные данные с результатами работы алгоритмов, созданных авторами [7–13]. Сравнение алгоритмов проводилось на тестовом примере при $n = 10$ источников и $m = 20$ потребителей документов. Результаты сравнения приведены в таблице.

Алгоритм	Возможность работы с вогнутыми функциями	Найденный оптимум	Вычислительная сложность	Источник
Йенсена-Барнеса	да	1635	$o(\exp(n))$	[7]
С.С. Лебедева	нет	–	$o(n^2) - o(n^3)$	[8]
Л.П. Падалко	нет	–	$o(n^2)$	[9]
С.Л. Уткина	да	1644	$o(n^3 - n^4)$	[10]
В.Р.Хачатурова, В.М. Монтлевича	да	1601	$o(n^3 - n^4)$	[11]
В.Р.Хачатурова, С.Л. Уткина	да	1538	$o(n^4)$	[12]
Хоанга Туй	да	1548	$o(n^3)$	[13]
Разработанный	да	1537	$o(n^2)$	[14–17]

Таблица. Сравнение различных алгоритмов решения нелинейной ТЗ

Заключение

Разработан генетический алгоритм решения задачи оптимизации документооборота на предприятии, позволяющий находить хорошее приближенное решение и имеющий приемлемую вычислительную сложность. Разработаны модифицированные операторы инициализации, кроссинговера и мутации, позволяющие получать допустимые решения исходной задачи. Предложенный алгоритм можно использовать для оптимизации документооборота при создании информационных систем управления предприятия.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты № 12-01-00474а и № 12-07-00376а).

Литература

1. Зыков А.Г., Безруков А.В., Немолочнов О.Ф., Поляков В.И., Андронов А.В. Графо-аналитические модели вычислительных процессов в САПР // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 4 (74). – С. 116–120.
2. Райзберг Б.А., Лозовский Л.Ш., Стародубцева Е.Б. Современный экономический словарь. – 5-е изд., перераб. и доп. – М.: ИНФРА-М, 2006. – 495 с.
3. Курейчик В.М. Биоинспирированный поиск с использованием сценарного подхода // Известия Южного федерального университета. Технические науки. – 2010. – Т. 108. – № 7. – С. 7–13.
4. Курейчик В.М., Кажаров А.А. Применение пчелиного алгоритма для раскраски графов // Известия Южного федерального университета. Технические науки. – 2010. – Т. 113. – № 12. – С. 30–36.
5. Курейчик В.М., Кажаров А.А. Использование роевого интеллекта в решении NP-трудных задач // Известия Южного федерального университета. Технические науки. – 2010. – Т. 120. – № 7. – С. 30–36.
6. Лунгу К.Н. Линейное программирование. Руководство к решению задач. – М.: ФИЗМАТЛИТ, 2005. – 128 с.
7. Йенсен П., Барнес Д. Потокосое программирование. – М.: Радио и связь, 1984. – 392 с.
8. Лебедев С.С. Конечный метод решения нелинейных задач транспортного типа // Экономика и математические методы. – М.: Наука, 1965. – Т. 1. – Вып. 1. – С. 71–82.
9. Падалко Л.П. Алгоритм решения нелинейной транспортной задачи // Экономика и математические методы. – М.: Наука, 1968. – Т. 4. – Вып. 6. – С. 962–964.
10. Уткин С.Л. Вычислительные алгоритмы и опыт минимизации вогнутой функции на выпуклом многограннике. – М.: ВЦ АН СССР, 1990. – 29 с.
11. Хачатуров В.Р., Монтлевич В.М. Решение нелинейных производственно-транспортных задач с неделимыми потребителями. – М.: ВЦ АН СССР, 1987. – 22 с.
12. Хачатуров В.Р., Уткин С.Л. Решение многоэкстремальных задач вогнутого программирования аппроксимационно-комбинативным методом. – М.: ВЦ АН СССР, 1988. – 38 с.
13. Хоанг Туй. Вогнутое программирование при линейных ограничениях // Доклады Академии наук СССР. – М., 1964. – Т. 159. – С. 32–35.
14. Чернышев Ю.О., Басова А.В. Теоретическое обоснование возможности решения вогнутой распределительной задачи с помощью методов эволюционного моделирования // Информационные техно-

- гии в науке, образовании, телекоммуникации и бизнесе. Материалы юбилейной XXX международной конференции. – Украина, Ялта–Гурзуф, 2003. – С. 147–149.
15. Басова А.В. Генетический метод отыскания глобального минимума многоэкстремальных задач // Математические методы в технике и технологиях. Сборник трудов XVI международной научной конференции. – СПб, 2003. – Т. 2. – С. 141–143.
16. Басова А.В. Новые методы отыскания глобального минимума многоэкстремальных задач // Фундаментальные и прикладные проблемы приборостроения, информатики, экономики и права. Научные труды VI международной научно-практической конференции. Книга «Информатика». – М., 2003. – С. 31–35.
17. Басова А.В., Белявский П.Г. Генетические методы решения задачи оптимального планирования грузовых морских перевозок // Вестник Донского государственного технического университета. – 2011. – Т. 11. – № 5. (56). – С. 630–633.

- Чернышев Юрий Олегович** – Филиал военной академии связи (г. Краснодар), доктор технических наук, профессор, myvnn@list.ru
- Басова Алина Викторовна** – Донской государственный технический университет, кандидат технических наук, доцент, зав. сектором, abasova@donstu.ru
- Панасенко Павел Александрович** – Филиал военной академии связи (г. Краснодар), офицер службы защиты государственной тайны, We_panasenko_777@mail.ru
- Поляков Владимир Иванович** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, v_i_polyakov@mail.ru