

3. Волков В. Т., Ягодников Д. А. Исследование и стендовая отработка ракетных двигателей на твердом топливе. – М. : Изд-во МГТУ им. Н. Э. Баумана, 2007. – 296 с.
4. Уразбахтин Ф. А., Уразбахтина А. Ю., Хмелева А. В. Критические ситуации при производстве и технической эксплуатации транспортно-пусковых контейнеров ракет. – М. ; Ижевск : Регуляр. и хаот. динамика, 2009. – 408 с.
5. ГОСТ 25136–82. Соединения трубопроводов. Методы испытаний на герметичность. – URL: <http://www.gosthelp.ru/text/GOST2513682Soedineniyatru.html> (дата обращения: 14.05.2012).
6. Баллистическая ракета на твердом топливе / А. М. Синюков, Л. И. Волков, А. И. Львов и др. – М. : Воениздат, 1972. – 512 с.

F. A. Urazbakhtin, DSc in Engineering, Professor, Votkinsk Branch of Kalashnikov Izhevsk State Technical University

A. A. Korenev, PhD, Votkinsk Branch of Kalashnikov Izhevsk State Technical University

Mathematical Model of Critical Situations According to Properties Intensity of Solid Propellant Rocket Engine

The mathematical model is considered, allowing to define critical (supernumerary) situations in operation of a solid propellant rocket engine. The model consists of modules: initial data input as parameters of state, environment influence and technical conditions; sets of criticality characteristics for separate elements and solid propellant rocket engine as a whole; analyzer, allowing to define quantitative values of parameters and characteristic of critical situations. It is possible to supplement the mathematical model with new characteristics, depending on certain parameters, including the time factor.

Keywords: critical situations, criticality indicators, parameters, solid propellant rocket engine, engine elements, mathematical model, power capacity, strength, stability, rigidity, crack resistance

Получено: 12.05.12

УДК 004.023, 004.82

А. С. Шаура, аспирант;

В. А. Тененёв, доктор физико-математических наук, профессор
Ижевский государственный технический университет имени М. Т. Калашникова

ПОСТРОЕНИЕ ДЕРЕВЬЕВ РЕШЕНИЙ С ПОМОЩЬЮ ГЕНЕТИЧЕСКОГО АЛГОРИТМА СТРУКТУРНО-ПАРАМЕТРИЧЕСКОГО СИНТЕЗА

Рассматривается задача построения деревьев решений как задача оптимального структурно-параметрического синтеза. Для ее решения предлагается генетический алгоритм, реализующий последовательное выращивание деревьев в соответствии с обучающим набором данных. Для кодирования структуры используется подход, основанный на сохранении всех структурных изменений в популяции в порядке их появления. Разработанные операторы мутации и скрещивания позволяют избежать известных проблем структурной оптимизации.

Ключевые слова: деревья решений, генетический алгоритм, структурно-параметрическая оптимизация, генетическое кодирование

Деревья решений представляют собой широко используемый в последнее время метод классификации, регрессии и анализа данных, основанный на представлении правил в иерархической, последовательной структуре [1, 2]. Для их построения обычно используют методы CART, C4.5, ID3 и др. [3], в основе которых лежит жадная стратегия поиска при выборе атрибутов для разбиения, и в случае неудачного выбора атрибута нет возможности сделать новый выбор, вернувшись на предыдущую итерацию. Разбиение множества в соответствии с выбранным атрибутом приводит к созданию подмножеств, характеризующихся минимальной статистической ошибкой, но такой подход не гарантирует, что все сформированное дерево окажется оптимальным. Как правило, уже по определению жадный алгоритм далеко не всегда позволяет построить оптимальное решение. Этим обусловлена перспективность использования эволюционных подходов или их комбинаций с указанными выше методами к построению деревьев решений [2, 4].

Если не ограничиваться рассмотрением бинарных деревьев решений с простейшими проверками условий в узлах типа равенств и неравенств, то необходимо решать более общую задачу построения n -арного дерева с произвольными условиями ветвления, в том числе и с проверками условий для нескольких переменных в узлах. Реализация такого дерева дает большие возможности при построении и обучении, а в ряде случаев позволяет получать более компактные по структуре решения. Такую задачу можно рассматривать как частный случай задачи оптимального структурно-параметрического синтеза, рассматривая построение только древовидных структур.

Для решения такой задачи построения оптимальных деревьев решений предлагается использовать генетический алгоритм структурно-параметрического синтеза, в котором каждая особь популяции представляет собой некоторое возможное дерево решений. Для кодирования структуры особей в генотипе алгоритм реализует идею применения «исторической» нумерации структурных новообразований (*historical markers, innovation number*) у индивидов популяции, предложенную в [5] для построения оптимальных нейронных сетей. Суть такой нумерации в том, что каждый новый узел, появившийся в структуре какого-либо дерева в популяции, получает свой уникальный порядковый номер (*innovation number*), который не изменяется в следующих поколениях и остается тем же и в генотипе потомков.

Генотип представляет собой совокупность взаимосвязанных списков, описывающих структуру фенотипа. Ключевыми в структуре являются узлы, в которых происходит ветвление по условию, в то время как хранить наравне с ними узлы-листья нет необходимости, поскольку используемое генетическое представление дуг позволяет хранить информацию и обо всех листьях дерева.

Упрощенная схема кодирования структуры представлена на рис. 1.

В соответствии с представленной схемой при кодировании дерева для каждого узла (**Node**), не являющегося листом, в генотип записывается порядковый номер появления этого узла (*InnovID*) в фенотипе индивидов популяции, тип этого узла (*Type*): *Root* – если узел является корнем дерева и *Node* – в противном случае, а также список дуг, выходящих из кодируемого узла (*Edges*). Дуги дерева (**Edge**) характеризуются входным (*In*) и выходным (*Out*) узлами и списком условий на переменные (*Variables*), при которых происходит переход по данной дуге. Условие на переменную (**Variable**) отражает, на какую именно из входных переменных в исходной задаче построения дерева решений (*VarID*) накладывается ограничение и какое это ограничение (*Value*).

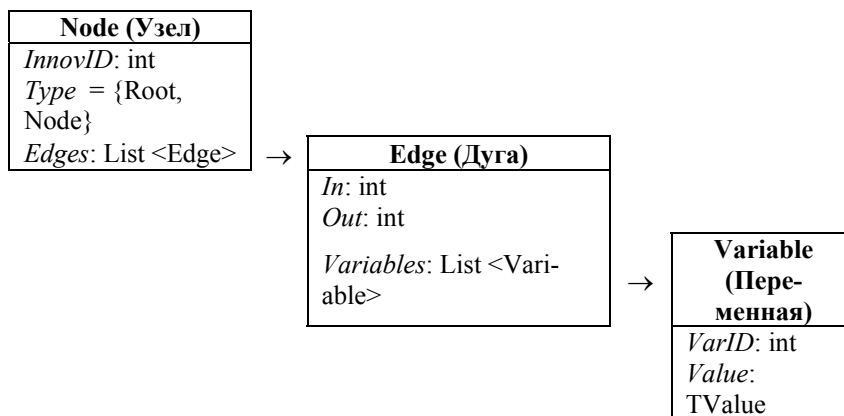


Рис. 1. Система связанных списков для кодирования фенотипа

В общем случае условия в узлах дерева и значения атрибутов на дугах могут носить произвольный характер, в качестве основных проверок условий можно перечислить: проверка равенства или неравенства, проверка булева значения конкретного оператора, проверка принадлежности к классу, проверка принадлежности к области (диапазону) значений.

Генетическое кодирование структуры в соответствии с рассмотренной схемой на примере изображенного на рис. 2 дерева решений представлено в виде системы связанных списков на рис. 3.

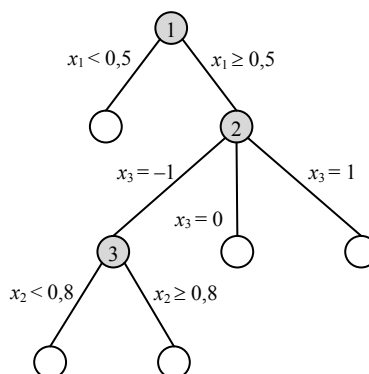


Рис. 2. Исходное дерево решений

В основе генотипа лежит перечисление узлов принятия решений исходного дерева-фенотипа с сохранением исторических номеров (*InnovID*) этих узлов. Каждому такому узлу ставится в соответствие список выходящих из него дуг, определяемых парой чисел *In* и *Out* – номерами узлов начала и конца дуги соответственно. Если дуга оканчивается листом, то *Out* = 0. Кроме соединяемых узлов в генотипе индивидов в качестве свойств дуг записаны условия в виде пар – номер переменной *VarID* и ее значение *Value*, проверяющиеся при переходе по дуге от узла *In* к узлу *Out*.

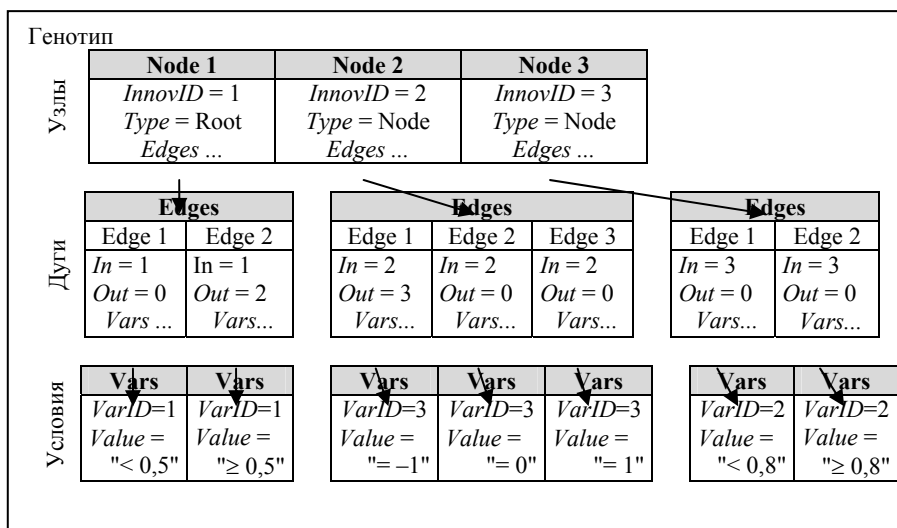


Рис. 3. Кодированное представление дерева решений

Рассматриваемый подход позволяет описать произвольное *n*-арное дерево с любыми условиями ветвления.

Начальная популяция генетического алгоритма формируется из минимально возможных деревьев, глубина которых равна 1, т. е. каждое такое дерево содержит только корень и листья. Функция приспособленности особей формируется на основании ошибки построения дерева по обучающему набору данных. В процессе работы генетического алгоритма деревья растут за счет появления новых узлов в результате мутации и скрещивания до тех пор, пока не будет выполнено условие останова алгоритма и найдено соответствующее решение.

Мутация реализует случайное появление новых узлов ветвления на месте листьев или корня дерева (рис. 4), при этом количество дуг с условиями, выходящих из нового узла, может быть зафиксировано и постоянно для всего дерева или меняться от одного узла к другому. Если новый узел возникает на месте корня дерева, то он становится новым корнем, а тот узел, который был им раньше – одним из его потомков (рис. 4, б). Если новый узел возникает на месте листа дерева, то он замещает его (рис. 4, в). При этом условия на новых дугах формируются из соответствующего числа выбранных атрибутов случайно взятой одной или нескольких переменных.

Используемое в алгоритме скрещивание основано на том, что введенная для кодирования структуры «историческая» нумерация узлов позволяет идентифицировать предка для любого индивида популяции, а для любых двух индивидов – выявить их общую часть в структуре деревьев. Исходя из этого, при скрещивании двух выбранных индивидов общая часть структуры родительской пары сразу переходит потомку, а различные узлы структуры разыгрываются с учетом возможностей пропорционально родительской приспособленности. Пример такого скрещивания представлен на рис. 5 (общая часть родительской структуры выделена жирным), на котором видно, что наличие в генотипе индивидов одинаковых исторических номеров свидетельствует о наличии общего исторического предка в предшествующем процессе эволюционного структурного роста.

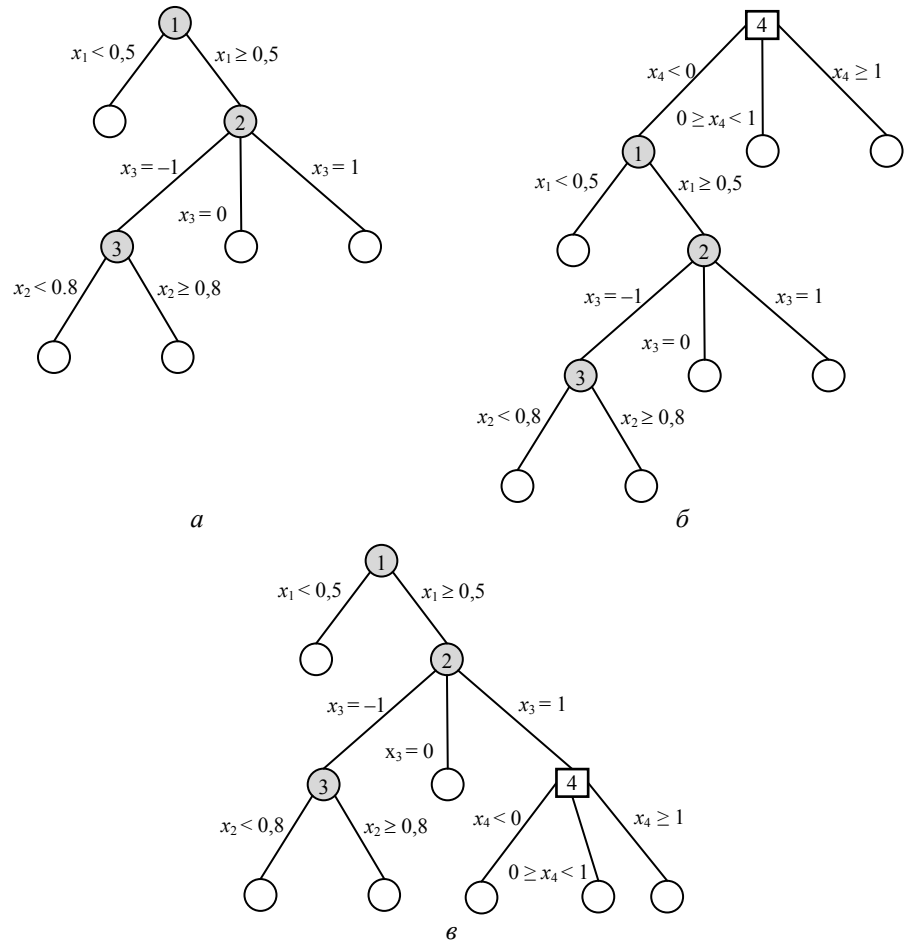


Рис. 4. Структурная мутация: а – исходное дерево; б – появление нового узла на месте корня дерева; в – появление нового узла на месте существующего листа

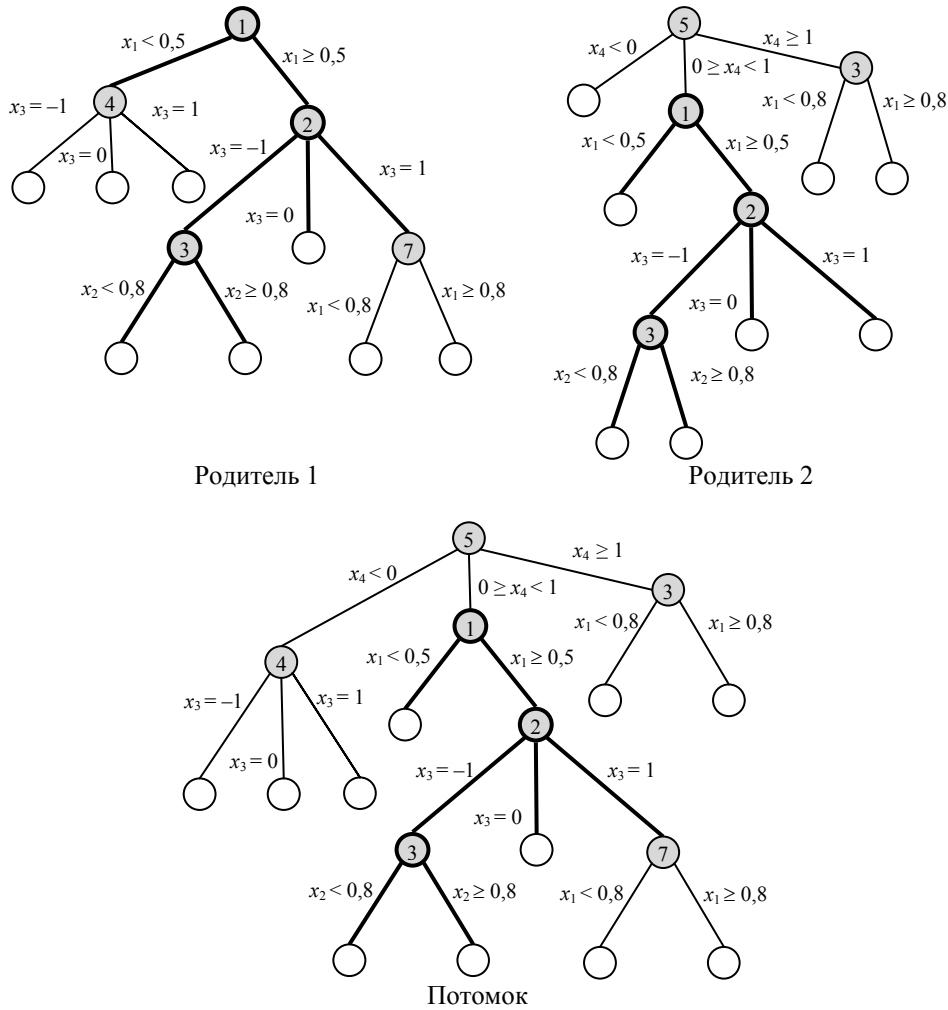


Рис. 5. Скрещивание, основанное на выявлении общей структурной части

Получаемые в результате представленных операторов скрещивания и мутации потомки являются допустимыми деревьями решений. Таким образом, рассмотренный подход к эволюционному построению структуры позволяет решить проблемы появления недопустимых особей, а применяемое кодирование – предотвратить потери генетической информации в процессе скрещивания, проблему чрезмерного усложнения структуры, а также проблему существования конкурирующих решений (competing conventions problem) в популяции [4, 5].

Рассмотрим работу алгоритма на классическом примере «Можно ли играть в гольф?» [4, 6]. Данные для построения дерева решений представлены в таблице в виде информации о погоде: наблюдение $X_1 = \{\text{Солнечно}, \text{Облачно}\}$, температура $X_2 = \{\text{Холодно}, \text{Умеренно}, \text{Жарко}\}$, влажность $X_3 = \{\text{Нормальная}, \text{Высокая}\}$, ветер $X_4 = \{\text{Да}, \text{Нет}\}$ и возможностью игры в гольф: $Y = \{\text{Да}, \text{Нет}\}$.

Наблюдения «Можно ли играть в гольф?»

№ п/п	X_1	X_2	X_3	X_4	Y
	Наблюдение	Температура	Влажность	Ветер	Играть
1	Солнечно	Жарко	Высокая	Нет	Нет
2	Солнечно	Жарко	Высокая	Да	Нет
3	Облачно	Жарко	Высокая	Нет	Да
4	Дождливо	Умеренно	Высокая	Нет	Да
5	Дождливо	Холодно	Нормальная	Нет	Да
6	Дождливо	Холодно	Нормальная	Да	Нет
7	Облачно	Холодно	Нормальная	Да	Да
8	Солнечно	Умеренно	Высокая	Нет	Нет
9	Солнечно	Холодно	Нормальная	Нет	Да
10	Дождливо	Умеренно	Нормальная	Нет	Да
11	Солнечно	Умеренно	Нормальная	Да	Да
12	Облачно	Умеренно	Высокая	Да	Да
13	Облачно	Жарко	Нормальная	Нет	Да
14	Дождливо	Умеренно	Высокая	Да	Нет

Результатом решения задачи должно быть наименьшее дерево, построенное по данным таблицы, позволяющее принять решение о том, возможно ли при заданной погоде (X_1, X_2, X_3, X_4) играть в гольф.

Работа алгоритма начинается с популяции, заполненной деревьями только с одним узлом ветвления (корнем), который для каждого индивида случайно выбирается из множества атрибутов {Наблюдение, Температура, Влажность, Ветер}. Каждый индивид оценивается по качеству классификации данных из исходной таблицы: приспособленность *Fitness* определяется разностью между общим количеством наблюдений в таблице – 14 и количеством неправильно классифицированных наблюдений *Misclassification*:

$$Fitness = 14 - Misclassification.$$

$Fitness = 14$ означает, что все наблюдения классифицированы верно, а $Fitness = 0$ – что нет ни одного верно классифицированного. Таким образом, чем выше приспособленность индивида, тем лучше его дерево описывает исходный набор данных.

Одно из возможных деревьев решений для задачи «Можно ли играть в гольф?», построенное по данным таблицы, представлено на рис. 6. Однако нетрудно заметить, что оно не является наилучшим вариантом решения. Оптимальное дерево для этой задачи, которое находит предложенный генетический алгоритм структурно-параметрического синтеза не более чем за 10 поколений, представлено на рис. 7. В отличие от дерева, представленного на рис. 6, оно имеет глубину, равную 2 и только 3 узла ветвления, но при этом также описывает данные таблицы.

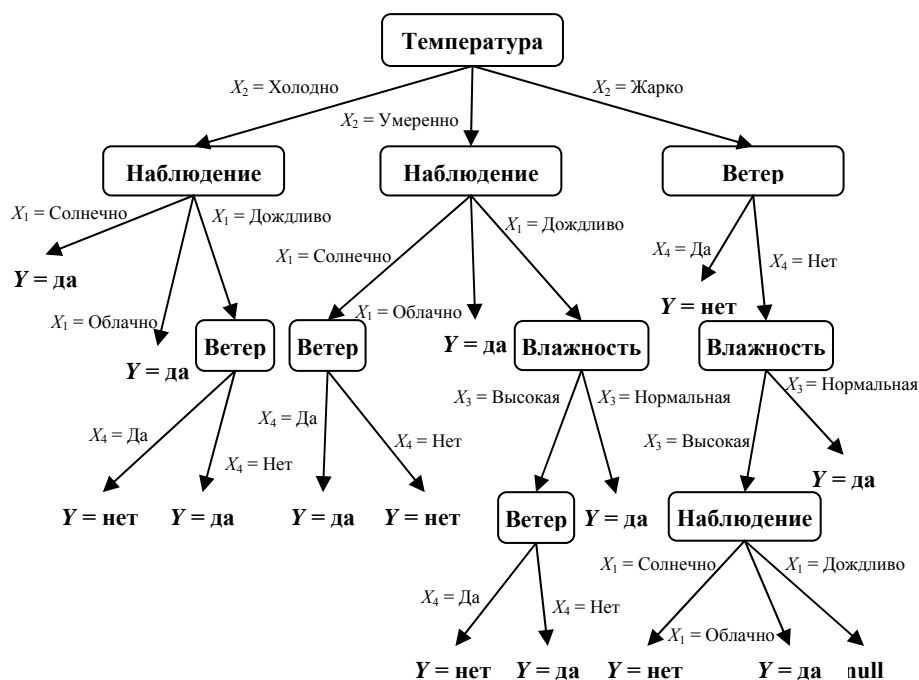


Рис. 6. Произвольное дерево решений

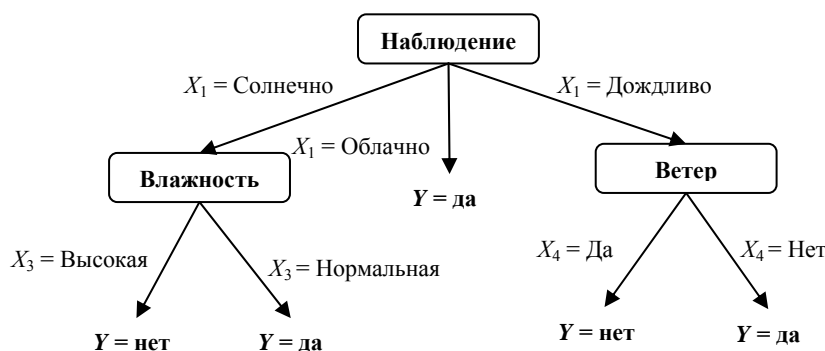


Рис. 7. Оптимальное дерево решений

Наименьший размер получаемого решения обеспечивается в результате того, что в предложенном алгоритме синтеза поиск начинается с минимально возможных решений, которые постепенно усложняются, в результате чего происходит последовательное исследование пространства поиска.

Рассмотренный подход применим и к решению задачи структурно-параметрической оптимизации в общем виде, если рассматривать построение произвольных структур, а не только деревьев.

Библиографические ссылки

1. *Тенев В. А., Якимович Б. А.* Генетические алгоритмы в моделировании систем. – Ижевск : Изд-во ИЖГТУ, 2010. – 308 с.
2. *Wo-Chiang Lee.* Genetic Programming Decision Tree for Bankruptcy Prediction // Atlantis Press, Proceedings of the 2006 Joint Conference on Information Sciences JCIS 2006. – Pp. 4-7. – URL: http://www.atlantis-press.com/php/download_paper.php?id=8 (дата обращения: 24.05.2012).
3. Classification and regression trees / L. Breiman, J. H. Friedman, R. A. Olshen et al. – California : Wadsworth & Brooks, 1984. – 368 p.
4. *Koza, J. R.* Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems). – MIT Press, 1992. – 819 p.
5. *Stanley, K. O., Miikkulainen, R.* Evolving Neural Networks through Augmenting Topologies // Evolutionary Computation. – 2002. – Vol. 10, Nr 2. – Pp. 99-127.
6. *Quinlan, J. R.* Induction of Decision Trees // Machine Learning. – Vol. 1, Iss. 1. – Pp. 81-106. – URL: <http://www.dmi.unict.it/~apulvirenti/agd/Qui86.pdf> (дата обращения: 24.05.2012).

A. S. Shaura, Post-graduate, Kalashnikov Izhevsk State Technical University

V. A. Tennev, DSc (Physics and Mathematics), Professor, Kalashnikov Izhevsk State Technical University;

Construction of decision trees applying genetic algorithm of structural and parametric synthesis

Construction of decision trees is considered as the problem of optimal structural and parametric synthesis. A genetic algorithm is proposed to solve this problem, implementing a consequent trees growing according to the learning dataset. The approach is used for structure encoding based on saving all structural variations in population in the order as they appear. Developed crossing and mutation operators allow avoiding known problems of structural optimization.

Keywords: decision trees, genetic algorithm, structural and parametric optimization, genetic encoding

Получено: 23.05.12