

Генетический алгоритм формирования оптимального подмножества диагностических тестов

Янковская А. Е., Цой Ю. Р.

yank@tsuab.ru, qai@mail.ru

Томск, Томский архитектурно-строительный университет,
Томский политехнический университет

Рассматривается проблема выбора оптимального подмножества безызбыточных безусловных диагностических тестов (ББДТ) с использованием эволюционного подхода. Для повышения качества результатов и обеспечения параллельного поиска предлагается использование нишинга (niching).

Введение

Выбор «хороших» безызбыточных диагностических тестов является одним из наиболее важных при принятии решений в интеллектуальных системах, поскольку от свойств используемых тестов существенно зависит качество получаемых решений. Однако этот выбор не всегда приводит к оптимальному решению, поскольку общее количество признаков в выбранном множестве тестов может быть слишком большим, также как временные и стоимостные затраты или ущерб (риск) [1], наносимый в результате выявления значений признаков исследуемого объекта, например, при решении геоэкологических или биомедицинских задач.

Впервые критерии оптимальности и безусловно актуальная задача поиска оптимального подмножества ББДТ поставлена в статье [2]. В статье [3] сформулированы критерии оптимальности, а в статье [4] предложено три алгоритма, обеспечивающие выполнение этих критериев: логико-комбинаторный, алгоритм на основе метода анализа иерархии и генетический алгоритм (ГА).

Постановка задачи

Кратко сформулируем рассматриваемую задачу. Пусть $\mathbf{T} = \{t_{ij}\}$, $i = 1, \dots, n$, $j = 1, \dots, m$ — матрица ББДТ, n — количество ББДТ, m — количество характеристических признаков, \mathbf{T}_i соответствует i -му ББДТ (i -я строка матрицы \mathbf{T}). Обозначим через $\mathbf{z} = \{z_j : j = 1, \dots, m\}$ множество характеристических признаков, причем $t_{ij} = 1 \leftrightarrow z_j \in \mathbf{T}_i$. Для каждого признака z_j зададим весовой коэффициент w_j и коэффициенты стоимости w'_j и ущерба (риска) w''_j [1].

Далее для краткости будем использовать термины «вес», «стоимость» и «ущерб» признака вместо соответственно «весовой коэффициент», «коэффициент стоимости» и «коэффициент ущерба».

Для данной матрицы \mathbf{T} с заданными весами, стоимостью и ущербами признаков, необходимо выделить такую подматрицу \mathbf{T}_0 , содержащую

n_0 строк, чтобы соответствующее ей множество тестов \mathbf{N}^0 обеспечивало выполнение следующих критериев, в порядке их следования [4]:

1. В выбранном множестве тестов \mathbf{N}^0 мощности n_0 должно содержаться максимальное число псевдообязательных признаков.
2. Множество \mathbf{N}^0 должно содержать минимальное общее число признаков.
3. Множество \mathbf{N}^0 должно иметь максимальный суммарный вес.
4. Множество \mathbf{N}^0 должно иметь наименьшую суммарную стоимость.
5. Множество \mathbf{N}^0 должно иметь наименьший суммарный ущерб.

Генетический алгоритм

Для решения поставленной задачи предлагается использовать ГА [5], представляющий итерационный вероятностный эвристический алгоритм поиска. Отличительной особенностью ГА является одновременная работа со множеством точек (популяцией) из пространства потенциальных решений.

Для решения рассматриваемой задачи каждое возможное решение представлено бинарной хромосомой (строкой) длины n , каждый i -й символ которой кодирует включение i -го диагностического теста в итоговое подмножество. В результате процесса искусственной эволюции, включающего отбор, рекомбинацию и вариацию (мутацию) «хороших», с точки зрения сформулированного критерия оценки [1], решений, качество решений в популяции постепенно улучшается. Окончанием генетического поиска, как правило, является нахождение субоптимального решения.

Для повышения качества результатов работы ГА предлагается использование нишинга [6, 7], для которого характерно сохранение в одной популяции как можно большего количества недоминируемых (несравнимых, nondominated) решений, соответствующих различным точкам на границе Парето, что дает возможность организовать параллельный поиск нескольких решений, в отличие от ГА без нишинга, в котором вся популяция сходится к одному решению.

Наиболее известны следующие разновидности нишинга:

1. Разделение приспособленности (fitness sharing) [6], при котором приспособленность хромосомы пересчитывается в зависимости от количества схожих с ней хромосом (в соответствии с выбранной метрикой) в данной популяции и их приспособленности.
2. «Перенаселение» (crowding) [7]. При использовании данной разновидности нишинга хромосомы-потомки замещают в популяции только наиболее близкие к ним хромосомы, что препятствует сходимости популяции к одному единственному решению.

Дальнейшие исследования будут посвящены исследованию влияния различных видов нишинга на качество результатов при решении поставленной задачи. Программный модуль, реализующий ГА будет включен в состав интеллектуального инструментального средства ИМСЛОГ [8].

Работа выполнена при поддержке РФФИ, проект № 07-01-00452, и РГНФ, проект № 06-06-12603В.

Литература

- [1] *Yankovskaya A. E., Tsoy Y. R.* Optimization of a set of tests selection satisfying the criteria prescribed using compensatory genetic algorithm // Proc. of IEEE East-West Design & Test Workshop. Odessa, Ukraine, September 15–19, 2005. — Pp. 123–126.
- [2] *Янковская А. Е.* Построение логических тестов с заданными свойствами и логико-комбинаторное распознавание на них // Интеллектуализация обработки информации (ИОИ-2002). — Симферополь, 2002. — С. 100–102.
- [3] *Yankovskaya A. E., Mozheiko V. I.* Optimization of a set of tests selection satisfying the criteria prescribed // 7th Int. Conf. on Pattern Recognition and Image Analysis. — St. Petersburg: SPbETU, 2004. — Vol. I. — Pp. 145–148.
- [4] *Колесникова С. И., Можсейко В. И., Цой Ю. Р., Янковская А. Е.* Алгоритмы выбора оптимального множества безызбыточных диагностических тестов в интеллектуальных системах поддержки принятия решений // Межд. конф. «Системный анализ и информационные технологии» САИТ-2005, Переславль-Залесский. — М.: КомКнига, 2005. — Т. 1. — С. 256–262.
- [5] *Емельянов В. В., Курейчик В. М., Курейчик В. В.* Теория и практика эволюционного моделирования. — М.: ФИЗМАТЛИТ, 2003. — 432 с.
- [6] *Goldberg D. E., Richardson J.* Genetic Algorithms with sharing for multimodal optimization // Proc. of the 2nd Int. Conf. on Genetic Alg., 1987. — P. 41–49.
- [7] *De Jong K.* An analysis of the behavior of a class of genetic adaptive systems. PhD thesis. — University of Michigan, Ann Arbor, 1975.
- [8] *Yankovskaya A. E., Gedike A. I., Ametov R. V., Bleikher A. M.* IMSLOG-2002 Software Tool for Supporting Information Technologies of Test Pattern Recognition // Pattern Recognition and Image Analysis. — 2003. — Vol. 13, № 4. — P. 650–657.