

Предложены и исследованы методы прогнозирования временных рядов, базирующиеся на эволюционных алгоритмах. Проанализированы проблемы применения генетических алгоритмов и разработаны меметические алгоритмы, которые являются дальнейшим развитием эволюционного подхода. Эффективность разработанных алгоритмов прогнозирования исследуется на основе анализа результатов вычислительного эксперимента.

© Л.Ф. Гуляницкий, 2012

УДК 519.8

Л.Ф. ГУЛЯНИЦКИЙ

РАЗРАБОТКА МЕТОДОВ ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ ЭВОЛЮЦИОННЫХ АЛГОРИТМОВ

Введение. Прогнозирование – это важная сфера приложения математических методов. Предметом нашего рассмотрения являются методы, использующие эволюционные алгоритмы (ЭА) для прогнозирования поведения и характеристик временных рядов, в частности, финансовых. Предлагаются и исследуются генетические алгоритмы (ГА) и меметические алгоритмы (МА) [1, 2]. Последние выбраны потому, что они позволяют соединить способность ГА к поиску глобальных экстремумов в сложных пространствах и эффективность локального поиска (ЛП) при нахождении решений, которые близки к оптимальным. Прогнозирование, базирующееся на ЭА, перспективно также и потому, что традиционные методы прогнозирования демонстрируют низкую эффективность во многих случаях, в частности, в случае процессов со скачкоподобным влиянием факторов. Традиционно, ГА используются в сфере оптимизации, но при условии некоторой модификации, могут служить и для классификации и прогнозирования [3, 4].

Исторически сложились два подхода к генетической классификации, названные в честь университетов, в которых они сформировались: мичиганский подход и питсбургский подход [1, 5]. Главное свойство, отличающее эти два подхода, – представляет ли каждый элемент популяции одно правило классификации или же целый набор таких правил. Хотя названные подходы имели и

другие отличия, будем использовать только это свойство, чтобы определять и различать их.

В мичиганском подходе каждый элемент популяции кодирует одно правило, и популяция в целом действует как концепция. В питсбургском подходе каждый элемент популяции – полная концепция, а наилучший элемент популяции, полученный в результате выполнения ГА, является заключительной концепцией, которая используется для классификации. В терминах мичиганского подхода питсбургский подход обрабатывает популяцию популяций либо популяцию множества правил.

Меметические алгоритмы. ГА моделируют принципы эволюционного отбора, наблюдаемые в природе [1]. Вычислительная схема традиционного ГА показана на рис. 1.

procedure ГА;

begin

$t := 0$;

Формирование Начальной Популяции P^0 ;

Оценивание особей из P^0 ;

repeat

$P :=$ Отбор Для Вариации (P^t);

Рекомбинация (P);

Мутация (P);

Оценивание особей из P ;

$P^{t+1} :=$ Отбор Популяции (P^t, P); $t := t+1$;

until выполнится условие завершения

end

РИС. 1. Вычислительная схема ГА

Базируясь на эволюционных принципах, ГА итерационно модифицируют популяцию абстрактных структур, используя генетические (эволюционные) операторы. ГА требуют лишь информацию о приспособленности, базирующуюся на значениях целевой функции, т.е. требуют лишь вычислимость целевой функции и не нуждаются в другой информации относительно специфики задачи. Алгоритмы, имеющие более тесную аналогию с культурной эволюцией, чем с биологической, называют гибридными ГА или меметическими (memetic) алгоритмами – МА [2]. Общий подход к разработке МА заключается в том, чтобы соединить ГА с ЛП, используя ЛП как дополнительный шаг в пределах ГА. В этих алгоритмах, в отличие от ГА, операторы рекомбинации и мутации применяются независимо друг от друга, а после этих операторов используются алгоритмы ЛП. МА может быть описан кортежем (m, k_r, k_m) , в котором m обозначает число особей в популяции, k_r – степень применения операторов рекомбинации, а k_m – степень применения операторов мутации (рис. 2).

```

procedure Memetic_Algorithm;
begin
  for  $j:= 1$  to  $m$  do
     $x :=$  ФормированиеДопустимогоВарианта;
     $x :=$  Local_Search ( $x$ );
     $P := P \cup x$ ;
  endfor;
  repeat
    for  $i:= 1$  to  $k_r$  do
      СлучайныйВыборОсобей  $x, y \in P$ ;
       $z :=$  Рекомбинация ( $x, y$ );
       $z :=$  Local_Search ( $z$ );
       $P := P \cup z$ ;
    endfor;
    for  $i:= 1$  to  $k_m$  do
      СлучайныйВыборОсобей  $x \in P$ ;
       $x :=$  Мутация ( $x$ );
       $x :=$  Local_Search ( $x$ );
       $P := P \cup z$ ;
    endfor;
     $P :=$  ОтборПопуляции ( $P$ );
  until выполнится условие завершения
end

```

РИС. 2. Меметические алгоритмы

Представление решений в задаче прогнозирования. Внутренним представлением в нашем случае выступают правила классификации [3]. То есть, каждая особь в популяции ГА является суть правилом классификации вида:

ЕСЛИ $[(v_1 R_1 x_1) \& (v_2 R_2 x_2) \& \dots \& (v_i R_i x_i) \& \dots \& (v_n R_n x_n)]$ **ТО** D ,

где v_i – значение i -го параметра правила классификации, x_i – значение i -го индикатора, отвечающего тестовому примеру, $R_i \in \{<=, >\}$ – отношение сравнения, $D \in \{Пост, Снад\}$ – класс, к которому правило классификации относит тестовый пример, $i = 1, n$.

Рассмотрим функцию, которую назовем геном правила классификации и определим ее для всех $i, i = 1, n$, так:

$$g(v_i, s_i, x_i) = \begin{cases} 1, & \text{если } v_i \leq x_i, s_i = 0, \\ 0, & \text{если } v_i > x_i, s_i = 0, \\ 1, & \text{если } v_i > x_i, s_i = 1, \\ 0, & \text{если } v_i \leq x_i, s_i = 1, \end{cases}$$

$$g(v_i, s_i, x_i) : R^1 \times \{0,1\} \times R^1 \rightarrow \{0,1\}.$$

Важной особенностью является введение механизма «выключателей», что дает возможность исключать из правил те гены, которые не являются существенными для классификации. Следовательно, определим ген с выключателем (в дальнейшем будем называть просто *геном*), так:

$$\tilde{g}(v_i, s_i, x_i, w_i) = \begin{cases} 1, & \text{если } w_i = 0, \\ g(v_i, s_i, x_i), & \text{если } w_i = 1, \end{cases}$$

$$\tilde{g}(v_i, s_i, x_i, w_i) : R^1 \times \{0,1\} \times R^1 \times \{0,1\} \rightarrow \{0,1\},$$

$$g_i(v, s, x, w) = \tilde{g}(v_i, s_i, x_i, w_i), \quad i = \overline{1, n},$$

где v – действительнзначный вектор параметров $v = (v_1, v_2, \dots, v_n)$; x – действительнзначный вектор значений индикаторов, соответствующих тестовому правилу; $x = (x_1, x_2, \dots, x_n)$, s – булевый вектор переключателей знака отношения сравнения; $s = (s_1, s_2, \dots, s_n)$, w – булевый вектор выключателей $w = (w_1, w_2, \dots, w_n)$.

Следовательно, правило классификации теперь можно переписать таким образом:

ЕСЛИ [$g_1(v, s, x, w) \& g_2(v, s, x, w) \& \dots \& g_i(v, s, x, w) \& \dots \& g_n(v, s, x, w)$] **ТО D**.

Таким образом, каждому правилу классификации можно поставить в соответствие набор из трех векторов длины n : v, s, w .

Это соответствие – одна из основных черт предлагаемого подхода. Ведь такое сочетание действительнзначных и булевых векторов в одной особи дает возможность объединить в пределах одного ЭА весь спектр эволюционных операторов и локального поиска, которые обычно разделяются на операторы для двоичных (binary EA) и непрерывных ЭА (continuous EA).

Фитнесс-функция формируется следующим образом. Первичное назначение каждому правилу приспособленности происходит согласно следующей функции:

$$f(C) = \frac{N^+}{N^+ + N^-} \cdot N^+ \cdot \frac{1}{N^-},$$

где N^+ – число охваченных правилом позитивных примеров, N^- – число охваченных правилом отрицательных примеров, C – правило или «условие».

Выражение $N^+/(N^+ + N^-)$ в вышеупомянутом выражении служит мерой точности правила; N^+ измеряет общность; $1/N^-$ – величина штрафа C за охват отрицательных примеров.

Поскольку структура особи в описываемой системе является дуальной, т. е. включает в себя действительнзначные и булевы векторы одновременно, то и структура эволюционного **оператора кроссинговера**, который и осуществляет рекомбинацию, является двойственной. Поэтому целесообразно рассматривать его как объединение булевой и действительнзначной частей.

Основной кроссинговер, оперирующий действительными компонентами, позволяет комбинировать значение компонент родительских особей и хранить его в генотипе потомков. Пусть P_1, P_2 – векторы параметров двух родительских особей:

$P_1 = [v_{11}, v_{12}, v_{13}, v_{14}, \dots, v_{1n}]$, $P_2 = [v_{21}, v_{22}, v_{23}, v_{24}, \dots, v_{2n}]$, а O_1, O_2 – векторы параметров двух особей потомков:

$$O_1 = [v_{11}, v_{12}, v'_{13}, v'_{14}, \dots, v'_{1n}], \quad O_2 = [v_{21}, v_{22}, v'_{23}, v'_{24}, \dots, v'_{2n}].$$

Определено, что $v'_{1i} = \beta v_{1i} + (1 - \beta)v_{2i}$, $v'_{2i} = \beta v_{2i} + v'_{1i} = \beta v_{2i} + (1 - \beta)v_{1i}$, где β – случайная величина, равномерно распределенная на $[0, 1]$.

Что касается кроссинговера булевых векторов, то это обычный одноточечный кроссинговер. Пусть P_1, P_2 – векторы булевых параметров двух родительских особей: $P_1 = [b_{11}, b_{12}, b_{13}, b_{14}, \dots, b_{1n}]$, $P_2 = [b_{21}, b_{22}, b_{23}, b_{24}, \dots, b_{2n}]$, тогда O_1, O_2 – векторы булевых параметров двух особей потомков – могут быть записаны так: $O_1 = [b_{11}, b_{12}, b_{23}, b_{24}, \dots, b_{2n}]$, $O_2 = [b_{21}, b_{22}, b_{13}, b_{14}, \dots, b_{1n}]$.

Благодаря двойственной природе особи в описываемом подходе, оператор мутации состоит из двух компонент – одноточечных мутаций соответствующих векторов. Одноточечная мутация булевого вектора – это простое инвертирование случайного «бита» (элемента вектора). Одноточечная мутация действительного вектора – замена случайной компоненты случайным значением, равномерно распределенным на интервале допустимых значений соответствующей компоненты. При выполнении мутации осуществляется модификация одной из трех ее компонент, которая выбирается случайным образом: либо мутация вектора параметров, либо вектора знаков», либо вектора выключателей.

Мутация происходит с вероятностью $p = p\{\xi \leq a\}$, где ξ – равномерно распределенная на $[0, 1]$ случайная величина, $a, a \in [0, 1]$, – параметр алгоритма.

На этапе **селекции** объединяются популяции родителей и потомков. Из этой объединенной популяции выбирается заданное число m особей с наибольшим значением приспособляемости, которые и формируют новую родительскую популяцию.

Для оптимизации подпопуляций используем метод ниш, который в данном случае будет заключаться в модификации фитнес-функции таким образом, чтобы приспособленность особей, которые скапливаются в некоторой области пространства правил (условий), уменьшалась пропорционально их количеству в данной области.

Такая модифицированная фитнес-функция называется распределенной функцией фитнеса и имеет вид:

$$S_f(C_i) = \frac{f(C_i)}{\mu(C_i)},$$

где C_i – i -я особь (правило), $f(C_i)$ – начальная фитнес-функция, $\mu(C_i)$ – функция ниш.

Для i -й особи функцию ниш предлагается вычислять следующим образом:

$$\mu(C_i) = \sum_{j=1}^N h(d_{ij}),$$

где d_{ij} – расстояние между i -й и j -й особями

$$h(d_{ij}) = \begin{cases} 1 - \left(\frac{d_{ij}}{\sigma_s}\right)^\alpha, & \text{если } d_{ij} < \sigma_s, \\ 0, & \text{в противном случае,} \end{cases}$$

а α, σ_s – параметры.

Вычислительный эксперимент. Задача заключается в том, чтобы базирясь на массиве исторических данных, касающихся определенных акций, предусмотреть будущие их характеристики или динамику их изменения. В частности, прогнозируем изменение относительной прибыли акций. Относительная прибыль вычисляется в соответствии с некоторым индексом – обычно, взвешенного среднего значения прибылей за подобные ценные бумаги. Например, если акции IBM выросли на 5% за определенный период, а индекс S&P 500 – на 3% в течение того же периода, то относительная прибыль IBM составляет + 2% в течение периода (относительно индекса S&P 500)

Анализ работ [3–5] показывает, что для наших целей необходимо рассматривать 10–15 атрибутов. Поэтому было использовано 10 атрибутов, которые представляют как техническую, так и фундаментальную информацию относительно каждой акции. Примеры состоят из значений этих атрибутов в определенные моменты времени и значений относительной прибыли акции через определенный промежуток времени.

Для прогнозирования были избраны акции десяти больших и известных компаний, которые торгуются на Нью-Йоркской фондовой бирже (NYSE). Основой для расчета относительной прибыли выступал биржевой индекс Standard & Poog's – S&P 500.

База данных примеров для акций каждой из исследуемых компаний содержит по 500 примеров. Мы включили 80% примеров в учебную выборку для ЭА – они используются для расчета приспособленности. 20% примеров используются, чтобы оценить результат – они составляют тестовую выборку. В качестве исследуемого периода выступает базовый период в 1 день.

В список атрибутов, использовавшихся при прогнозировании, входят:

1. Величина относительной прибыли акции за базовый период.
2. Цена акции на момент окончания базового периода (цена закрытия, Close).
3. Объем торговли акцией в течение базового периода.
4. Значения простого скользящего среднего цены акции, рассчитанного по значениям цен закрытия предыдущих периодов.

5. Значения экспоненциального скользящего среднего цены акции, рассчитанного по значениям цен закрытия.

6. Значения технического индикатора восхождения-расхождения скользящих средних (MACD, Moving Average Convergence/Divergence), рассчитанного по ценам закрытия.

7. Значения технического индикатора MACD Signal.

8. Разность значений индикаторов, описанных в пунктах 6 и 7.

9. Значения технического индикатора относительной силы (RSI, Relative Strength Index), рассчитанного по ценам закрытия.

10. Значения закрытия биржевого индекса S&P 500.

Каждый алгоритм прогнозирования возвращает один из трех возможных прогнозов для каждого примера из тестового множества – «рост», «спад» или «нет прогноза». В табл. 1 отображены некоторые результаты экспериментов.

ТАБЛИЦА 1. Результаты эксперимента – доли прогнозов

№	Компания	Правильный прогноз, %		Нет прогноза, %		Ошибочный прогноз, %		Верные прогнозы, %	
		ГА	МА	ГА	МА	ГА	МА	ГА	МА
1	General Electric	49,7	37,1	33,1	35,1	17,2	27,8	74,3	57,1
2	Bank of America	47,1	41,7	45,1	25,1	7,8	33,2	85,9	55,6
3	JPMorgan Chase	44,1	53,8	34,7	27,2	21,2	19,1	67,5	73,8
4	ExxonMobil	47,2	41,9	50,2	32,9	2,6	25,2	94,8	62,4
5	Berkshire Hathaway	45,9	42,6	49,1	29,8	5,0	27,6	90,2	60,7
6	AT&T	45,4	50,8	51,0	29,4	3,6	19,9	92,6	71,9
7	Wal-Mart Stores	50,3	55,3	41,3	24,9	8,4	19,8	85,7	73,7
8	Citigroup	42,8	59,8	37,9	27,7	19,3	12,5	68,9	82,7
9	Procter & Gamble	43,0	57,7	34,9	21,9	22,0	20,4	66,1	73,9
10	IBM	47,2	49,1	45,5	29,2	7,3	21,7	86,6	69,4

С помощью ЭА можно делать прогноз для каждого примера тестового множества, используя алгоритмы, о которых шла речь выше. Однако, вариант «нет прогноза» позволяет ЭА указывать на случаи, когда акция может как вырасти, так и упасть почти с одинаковой вероятностью.

Отметим, что вычисляя результат по тестовому множеству, в среднем ГА правильно спрогнозировали динамику акций исследуемых компаний относительно рынка в 46,3% случаев, не осуществили прогнозирование в 42,3% случаев и, соответственно, ошибочно предсказали динамику в 11,4% случаев.

На тех же данных МА выдали правильный прогноз в 49,0 % случаев, не выдали прогноз в 28,3 % случаев и сгенерировали ошибочный прогноз в 22,7 % случаев. Следовательно, можно сделать вывод, что в контексте задачи, которая рассматривается, прогнозирование с помощью МА является более рискованным, чем за ГА, хотя МА выдавали правильный прогноз на 2,7% чаще, чем ГА.

Применяется и другой критерий качества результата, который называется средний итог. Он принимает во внимание как правдивость прогноза, так и абсолютное значение прибыли с акции. Если ЭА правильно предугадывает относительное направление динамики тестового примера, то к итогу добавляется фактическая относительная прибыль соответствующей акции. Если прогноз неправилен, значение фактической относительной прибыли вычитается из итога. Результаты экспериментов в терминах среднего итога приведены в табл. 2.

ТАБЛИЦА 2. Результаты эксперимента – средние итоги (%)

№	Компания	ГА	МА
1	General Electric	12,8	13,8
2	Bank of America	12,6	12,9
3	JPMorgan Chase	11,6	12,3
4	ExxonMobil	12,7	11,8
5	Berkshire Hathaway	11,0	10,3
6	AT&T	11,9	13,8
7	Wal-Mart Stores	10,0	12,4
8	Citigroup	10,9	13,0
9	Procter & Gamble	12,4	13,9
10	IBM	11,5	12,9
Среднее по всем компаниям		11,7	12,7

Вычисленный итог усредняется по всем примерам, а в дальнейшем – и по всем экспериментам. По итогам всех расчетов, точность ГА достигает среднего итога в +11,7 %, а МА – в +12,7%. То есть МА показывает в среднем на 8,54% лучшие чем ГА результаты, если подсчитывать успешность по средним итогам.

Заключение. Рассмотрено применение эволюционных алгоритмов – ГА и МА в нетрадиционной для них сфере – прогнозировании временных рядов. Особое внимание уделено не только разработке метода прогнозирования на основе МА, но и сравнению его с предложенным ГА. Также проведен вычислительный эксперимент по прогнозированию ряда акций ведущих компаний, котирующихся на Нью-Йоркской фондовой бирже. По его результатам МА показал в среднем на 8,54% лучшие результаты, чем ГА. Это подтверждает целесообразность разработки и внедрения усовершенствованных локальных операторов в структуру ЭА прогнозирования.

Л.Ф. Гуляницький

РОЗРОБКА МЕТОДІВ ПРОГНОЗУВАННЯ НА ОСНОВІ ЕВОЛЮЦІЙНИХ АЛГОРИТМІВ

Запропоновано і досліджено методи прогнозування часових рядів, які ґрунтуються на еволюційних алгоритмах. Проаналізовані проблеми застосування генетичних алгоритмів і розроблені меметичні алгоритми, які є подальшим розвитком еволюційного підходу. Ефективність розроблених алгоритмів прогнозування досліджена на основі аналізу результатів обчислювального експерименту.

L.F. Hulianytskyi

DEVELOPMENT OF FORECASTING METHODS ON THE BASIS OF EVOLUTIONARY ALGORITHMS

Methods of time series forecasting based on evolutionary algorithms are proposed and investigated. Problems of application of genetic algorithms are analyzed and memetic algorithms, which are the development of the evolutionary approach, are built. Efficiency of the developed forecasting algorithms is investigated based on the analysis of results of numerical experiment.

1. *Holland J.H.* Adaptation in natural and artificial systems. – Cambridge, Mass.: MIT Press. – 1992.
2. *Moscato P.* Memetic algorithms: a short introduction // New ideas in optimization (Eds. Corne D, Dorigo M, Glover F.). – NY: McGraw-Hill, 1999. – P. 219–234.
3. *Mahfoud S., Mani G.* Financial forecasting using genetic algorithms // Applied Artificial Intelligence. – 1996. – 10, N 6. – P. 543–565.
4. *Nicholls J.F., Malan K.M., Engelbrecht, A.P.* Comparison of trade decision strategies in an equity market GA trader // Proc. IEEE Symp. «Computational Intelligence for Financial Engineering and Economics (CIFEr)» (11–15 April 2011, Paris). – Paris, 2011. – P. 1–8.
5. *Luke S.* Essentials of Metaheuristics (First Edition). – Lulu, 2011. Available at <http://cs.gmu.edu/~sean/book/metaheuristics/>.

Получено 25.12.2011

Об авторе:

Гуляницький Леонід Федорович,

доктор технических наук, заведующий отделом
Института кибернетики имени В.М. Глушкова НАН Украины.
leonhul.icyb@google.com