

УДК 681.32

ОТБОР РЕГРЕССИОННЫХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ПОТЕРИ КРОВИ ПРИ РОДАХ С ПОМОЩЬЮ ГА**Пачаджи М.А., Васяева Т.А.**Донецкий национальный технический университет
кафедра автоматизированных систем управления
E-mail: donnty@mail.ru, vasyaeva@gmail.com**Аннотация**

Пачаджи М.А., Васяева Т.А. Отбор регрессионных моделей для прогнозирования потери крови при родах с помощью ГА. Решается задача получения знаний для медицинских систем поддержки принятия решений. Разработан подход к отбору уравнений регрессионных моделей для прогнозирования потери крови при родах на основе генетических алгоритмов (ГА).

Введение

Ежегодно в мире при родах и в послеродовом периоде умирает 350-370 женщин. Самая частая причина смертности – акушерское кровотечение. Акушерские кровотечения являются одним из самых опасных осложнений, которые могут произойти во время родов, частота их появления 2-3% из всех случаев.

Возможность прогнозирования патологической акушерской кровопотери накануне родов позволит своевременно создать запасы кровезаменителей, плазмы, препаратов крови, возможно и аутокрови, а также оказать адекватную терапию с привлечением высококвалифицированных реаниматологов, акушеров-гинекологов и сосудистых хирургов для проведения органосохраняющих операций (перевязка сосудистых пучков и подчревных артерий) в случае развития послеродового кровотечения.

Общая постановка проблемы

При построении математической функции классификации или регрессии основная задача сводится к выбору наилучшей функции из всего множества вариантов. Дело в том, что может существовать множество функций, одинаково классифицирующих одну и ту же обучающую выборку (рис. 1).

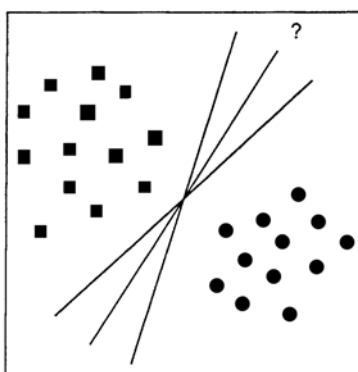


Рисунок 1 – Варианты линейного разделения обучающей выборки

Для построения регрессионной модели необходимо выполнить отбор необходимых факторов и затем рассчитать коэффициенты уравнения. Естественно выбирая тот или иной набор параметров можно получать различные регрессионные модели, причем многие из них будут показывать хорошие результаты. Таким образом, для построения оптимальной модели необходимо решить две задачи: выбрать факторы риска, которые будут переменными множественной регрессии; и рассчитать коэффициенты уравнения. Отбор факторов риска

является непростой задачей [1], которую можно решать с помощью ГА [2]. Предлагается решать эти две задачи одновременно.

Разработка метода

Стандартный ГА [3-4] предполагает следующую последовательность, рис.2.

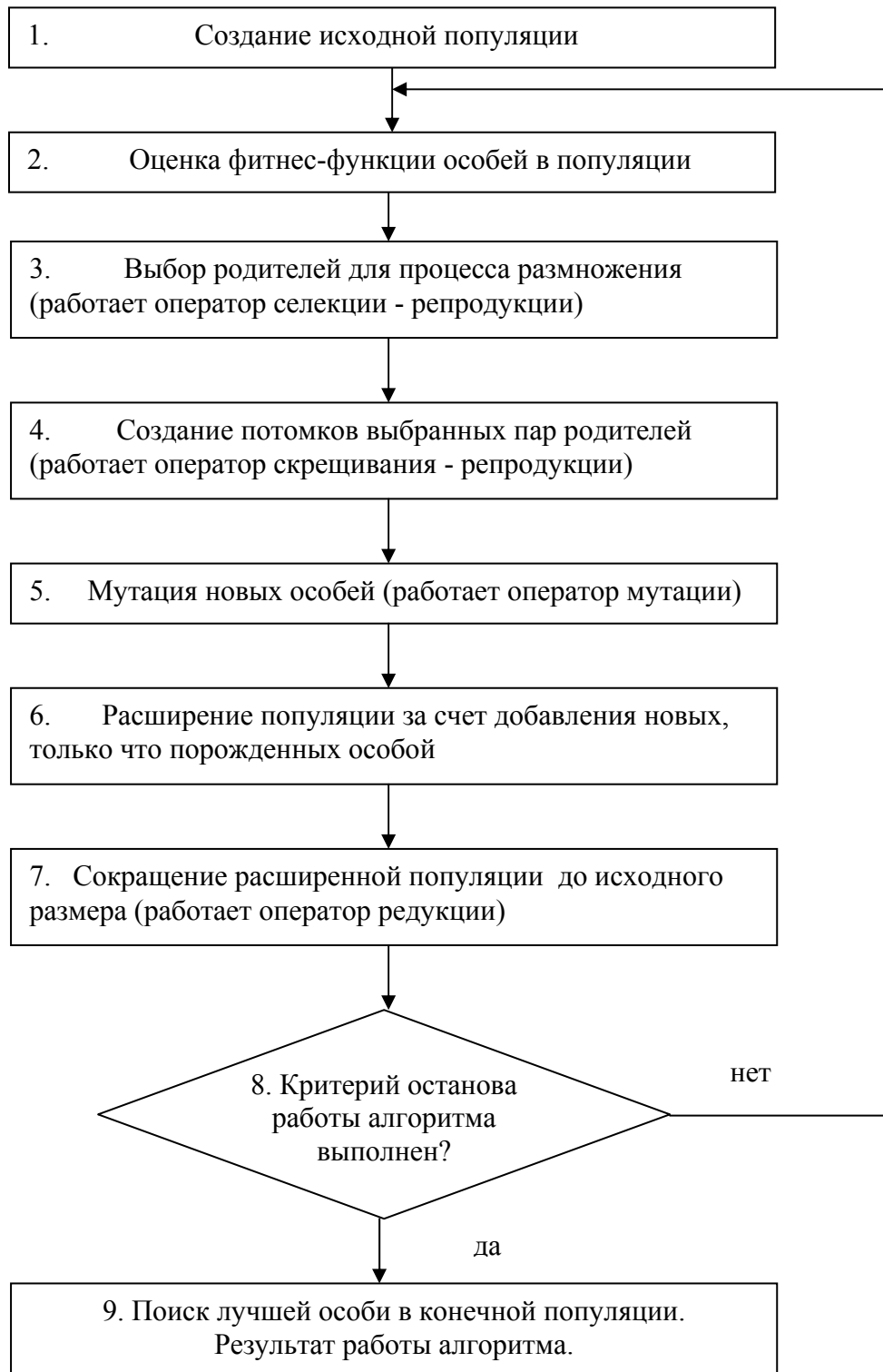


Рисунок 2 – Простой генетический алгоритм

1. На первом этапе выполняется генерация начальной популяции, где случайным образом генерируется некоторое количество особей. Важным этапом является разработка

способа кодирования хромосомы, так как каждая особь это возможное решение задачи (уравнение множественной регрессии). Мы предлагаем следующую структуру хромосомы рис. 3. Каждая особь представляется последовательностью определенного количества битов (определяется максимально возможным количеством факторов риска). Значение каждого бита может быть равно «1», если фактор с соответствующим номером присутствует в данном уравнении регрессии, и «0», если этот фактор отсутствует. Таким образом, каждой хромосоме соответствует регрессионная модель, построенная методом наименьших квадратов (МНК), с соответствующим ее структуре набором факторов риска.

Факторы риска

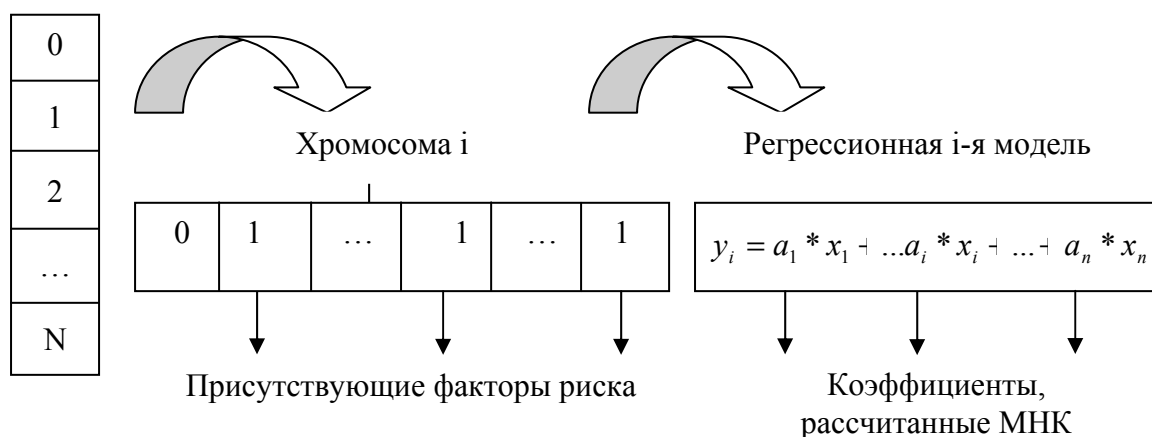


Рисунок 3 – Кодирование хромосомы

2. Второй этап подразумевает расчет фитнес-функции. В качестве фитнес-функции выступает ошибка регрессионной модели, которая может рассчитываться по формуле (1):

$$E = \frac{1}{N} \cdot \sum_{i=1}^N (y - y')^2 \quad (1)$$

где N – количество примеров в обучении, y – полученный результат регрессионной модели, y' – действительный результат.

Рассчитанная ошибка сопоставляется каждой хромосоме. Лучшая особь будет иметь минимальную ошибку, то есть минимальную фитнес-функцию.

Третий этап является реализацией генетического алгоритма. Генетические операторы выполняются стандартные.

3. Наиболее распространенный метод реализации оператора репродукции (ОР) – построение колеса рулетки, в котором каждой хромосоме соответствует сектор, пропорциональный ее значению фитнес-функции, что обеспечивает большую вероятность выбора лучших особей для оператора кроссинговера.

4. Применяется простой оператор кроссинговера (ОК), который выполняется в 3 этапа:

- выбираются две хромосомы (родители) из текущей популяции;
- случайно выбирается точка скрещивания – число k из диапазона $[1, 2, \dots, n-1]$, где n – длина хромосомы;
- две новых хромосомы A' , B' (потомки) формируются из A и B путем обмена подстрокой после точки скрещивания (рис. 4).

ОК выполняется с заданной вероятностью P_c (отобранные два родителя не обязательно производят потомков).

$$\begin{array}{ccc}
 A = a_1 a_2 \dots a_k a_{k+1} \dots a_l & & A' = a_1 a_2 \dots a_k b_{k+1} \dots b_l \\
 B = b_1 b_2 \dots b_k b_{k+1} \dots b_l & \longrightarrow & B' = b_1 b_2 \dots b_k a_{k+1} \dots a_l
 \end{array}$$

Рисунок 4 – Формирование хромосом оператором кроссинговера

5. Оператор мутации (ОМ) играет вторичную роль и его вероятность обычно мала. Оператор мутации выполняется в 2 этапа:

- в хромосоме $A = a_1 a_2 \dots a_k \dots a_n$ случайно выбирается k -ая позиция (бит) мутации ($1 \leq k \leq n$).
- Производится инверсия значения гена в k -й позиции, формула (2).

$$a'_k = \overline{a_k} \quad (2)$$

6. Следующий этап алгоритма предполагает расширение популяции за счет добавления новых, только что созданных особей.

7. Затем полученную популяцию сокращаем до исходных размеров (работает оператор редукции), путем отбора лучших особей.

8. Критериев останова работы алгоритма будет два: максимально заданное количество шагов; незначительное изменение значения фитнес-функции.

9. После выполнения критерия останова выполняется поиск лучшей особи в конечной популяции, что является результатом работы алгоритма (уравнение множественной регрессии). В случае, не выполнения критерия останова, переходим на пункт 2.

Выводы

Рассмотрена актуальная задача выбора лучшей регрессионной модели, для прогнозирования потери крови при родах. В дальнейшем планируется реализовать рассмотренный математический аппарат и провести эксперименты на реальных медицинских данных, предоставленными сотрудниками центра материнства и детства. Планируется разработка и внедрение системы поддержки принятия решений (СППР) прогнозирования акушерских кровотечений.

Список литературы

1. Васяева Т.А. Анализ методов отбора факторов риска развития патологий в акушерстве и гинекологии / Васяева Т.А., Иванов Д.Е., Соков И.В., Сокова А.С. // Збірка матеріалів II Всеукраїнської науково-технічної конференції студентів, аспірантів та молодих вчених. ІУС КМ-2011 11-13 квітня 2011р., Донецьк, ДонНТУ, 2011, Т.1.-209-212с.
2. Васяева Т.А. Отбор факторов риска потери крови при родах / Т.А. Васяева, Д.Е. Иванов, И.В. Соков. // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту: Матеріали міжнародної наукової конференції. Том 1. - Херсон: ХНТУ, 2011. – 472 с.
3. Скобцов Ю. О. Основы эволюционных обчислень / Скобцов Ю. О. – Донецьк: ДонНТУ, 2009. – 316 с.
4. Рутковская Д., Пилиньский М., Л. Рутковский Л. // Нейронные сети, генетические алгоритмы и нечеткие системы. М.: 2004. – 452с.