

ОТБОР ФАКТОРОВ РИСКА ПОТЕРИ КРОВИ ПРИ РОДАХ

Т. А. Васяева, Д. Е. Иванов, И. В. Соков

Донецкий национальный технический университет 83000, Донецк, ул. Артема 58
Институт прикладной математики и механики НАН Украины 83114, Донецк, ул. Р. Люксембург, 74
vasyaeva@gmail.com, ivanov@iamm.ac.donetsk.ua, sokov_ivan@mail.ru

Ежедневно в мире от осложнений, связанных с беременностью и родами, умирает 1500 женщин. По оценкам экспертов большинство этих случаев можно было предотвратить. Около 80% смертей обусловлено прямыми причинами: кровотечения, инфекции, гестозы, патологические роды. И лидирует в этом списке смертность в результате массивной кровопотери. При этом если даже удастся предотвратить гибель женщины, то следствием перенесенного критического состояния является развитие какой-либо тяжелой хронической патологии - вплоть до инвалидности. Таким образом данная проблема носит не только медицинский, но и выраженный социальный характер.

Не вызывает сомнения, что наиболее действенный путь снижения частоты кровотечений лежит в разработке программ прогнозирования индивидуального объема потери крови при родах. Сложность их разработки заключается в необходимости научного анализа большого количества клинических и лабораторных показателей, которые находятся в сложной зависимости друг от друга и не всегда поддаются количественной оценке. Наиболее перспективным методом для реализации этой задачи является прогноз данного вида патологии на основе интеллектуального анализа данных с применением современного программного обеспечения. При этом часть анализируемых параметров имеет принципиальное значение для прогнозирования, другая же - не столь важна и может игнорироваться, поэтому на первом этапе необходимо определить факторы риска, которые вызывают развитие акушерских кровотечений.

Как правило, первоначально отбор данных для анализа выполняется врачом по следующему принципу: сначала осуществляется выделение факторов риска относящихся к данной патологии в основном по формальному принципу: отягощенный семейный анамнез, перинатальные факторы, заболевания в анамнезе, неблагоприятная макросоциальная среда, социально-гигиенические факторы и т.д. Практически каждый врач, при постановке диагноза, всегда старается анализировать весь доступный комплекс сведений о пациенте, поэтому предоставленный перечень собранных факторов для прогнозирования объема потери крови при родах, является очень большим. В него входят медицинские и социально-демографические факторы. Перечень факторов риска частично представлен:

- возраст моложе 18 или старше 35 лет;
- рост менее 155 см и вес до беременности на 20% ниже или выше нормы для данного роста;
- пятая и последующая беременность, особенно если беременная старше 35 лет;
- злоупотребление курением;
- многоплодная беременность;
- отсутствие прибавки в весе или минимальная прибавка;
- срок беременности более 42 недель;
- и многие другие.

Выполним отбор информативных параметров с помощью нейронных сетей (НС) [1,2] и генетических алгоритмов (ГА) [2,3]. Такой подход (совместное использование НС и ГА) позволит не просто выполнить отбор факторов риска, а выбрать информативный набор данных, сохранив взаимосвязанные переменные.

Для реализации такого подхода сначала необходимо разработать НС для определения риска развития акушерских кровотечений. Для практики самым важным является определение риска потери крови при родах более чем 0,5% от массы тела, с целью принятия соответствующих мер. В этом случае, по сути, выполняется классификация на два класса: патологическое кровотечение и допустимое. Для реализации поставленной задачи в такой форме целесообразно использовать многослойную НС прямого распространения. Такой тип НС показывает хорошие результаты при обучении с учителем, а так как у нас имеется обучающая выборка с реальными медицинскими данными, то выберем многослойную НС для классификации на патологическую и обычную потерю крови при родах. Таким образом, на первом этапе необходимо выбрать архитектуру нейронной сети и обучить ее. Учитывая то, что в дальнейшем (при отборе факторов риска) на каждом шаге выполнения ГА будет происходить обучение НС, целесообразно выбирать архитектуру сети с минимальной сложностью, что позволит уменьшить время выполнения программы. Под сложностью сети будем предполагать вычислительную сложность алгоритма обучения НС. Под вычислительной сложностью алгоритма обучения будем понимать количество операций за один шаг обучения. Известно, что для алгоритма обратного распространения количество операций связано линейной зависимостью с синаптическими весами (включая пороги) [1]. Т. е. вычислительную сложность определяет число скрытых слоев и число нейронов на каждом из них. В результате проведенных экспериментов выбрана архитектура НС, представленная на рис. 1. Количество входов обусловлено максимальным количеством факторов риска (после кодирования их получили 99). Выбранные активационные функции - гиперболический тангенс и линейная функция.

После того как определена и успешно обучена НС, выполняется ГА, который подает различные комбинации факторов риска на входы НС и затем выполняется попытка обучить сеть при такой комбинации отобранных факторов риска. Принцип реализации кодирования хромосомы представлен на рис. 2. Каждая хромосома представляет собой последовательность определенного количества битов (определяется максимальным количеством факторов риска). Значение каждого бита равно 1, если фактор с соответствующим номером присутствует в данном наборе, и нулю, если этот фактор отсутствует. Предусмотрены различные варианты параметров ГА в качестве метода селекции можно выбирать колесо рулетки или турнир (с указанием количества особей в туре); в качестве метода редукции предусмотрена элитарная стратегия, полная замена, частичная замена популяции (с указанием процента заменяемых особей); можно задавать вероятность операции мутации и скрещивания (предусмотрен одноточечный и двухточечный оператор скрещивания). В качестве критерия останова можно использовать определенное количество итераций или указать количество повторений результата.

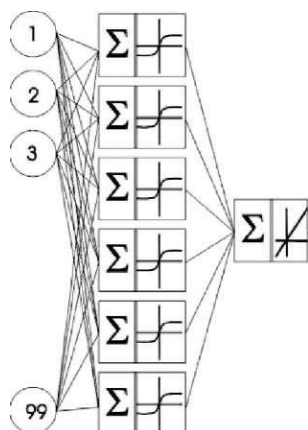


Рисунок 1. Архитектура нейронной сети

Факторы риска

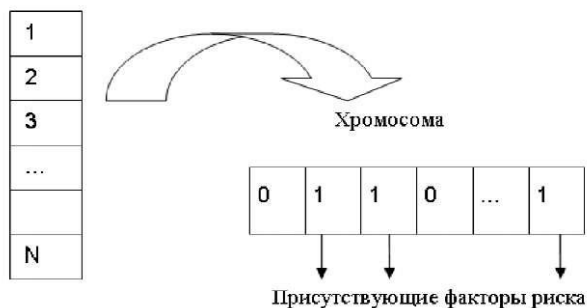


Рисунок 2. Кодирование хромосомы

В предлагаемом подходе для выделения признаков использован генетический алгоритм, который имеет модифицированную схему реализации применительно к задаче многокритериальной оптимизации.

В то время как большинство подобных методов, используемых для решения таких задач, использует единый, составной оптимизируемый критерий [4]. Решением задачи в данном случае является несколько недоминируемых подмножеств признаков. Разработана фитнес-функция (1), которая предполагает поиск решения, являющегося оптимальным согласно двум критериям - учитывается точность классификации и количество факторов риска. К тому же данная фитнес-функция позволяет задавать желаемое соотношение точности классификации и количества факторов риска.

$$(1) \quad F = \left(\frac{X_i}{X_n} \right) \cdot Q_1 + \left(\frac{E_i - E_n}{E_n} \right) \cdot Q_2$$

где X_i - количество присутствующих факторов риска в i -й хромосоме, X_n - максимальное количество факторов риска, E_i - ошибка обучения НС для i -й хромосомы, E_n - ошибка обучения НС при использовании максимального количества факторов, Q_1 и Q_2 - коэффициенты, с помощью которых регулируется соотношение между точностью классификации и числом факторов риска. Рекомендуется выбирать диапазон значений (2), и придерживаться условия (3).

$$Q_1, Q_2 \in (0,1) \quad (2)$$

$$Q_1 + Q_2 = 1 \quad (3)$$

После этапа подготовки данных получили обучающий массив размером 99 x 100, где 99 входных параметров и 100 обучающих примеров. Данная выборка разделена на две - по 50 примеров в каждой. По первой выборке проводилось обучение нейронной сети и отбор признаков с помощью генетических алгоритмов. В результате количество факторов уменьшилось практически в два раза. Затем, на второй выборке успешно протестирована нейронная сеть, с использованием только выделенных факторов. Таким образом, результат можно считать положительным и переходить к этапу построения и обучения аналитической системы прогнозирования потери крови при родах.

ЛИТЕРАТУРА

1. Саймон Хайкин Нейронные сети: полный курс, 2-е издание. : Пер. с англ. - М. : Издательский дом «Вильямс», 2006. - 1104 с.
2. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Рутковская Д., Пилинский М., Рутковский Л. - М.: Горячая линия - Телеком, 2006. - 452 с.
3. Скобцов Ю. О. Основы эволюционных обчислень / Скобцов Ю. О. - Донецьк: ДонНТУ, 2009. - 316 с.
4. Соков И.В. Анализ методов отбора факторов риска развития патологий в акушерстве и гинекологии / И. В. Соков, А. С. Сокова, Т. А. Васяева, Д. Е. Иванов // 1УС КМ - 2011: II Всеукраїнська науково-технічна конференція студентів, аспірантів та молодих вчених, 11-13 квітня 2011р., Донецьк: ДонНТУ, 2011. -Т.1. - 297с.