

Комплексная оценка параметров коагуляции у беременных женщин с помощью нечеткого классификатора

Аннотация

Описан способ оценки свертывающей системы крови у беременных женщин с помощью нечеткого классификатора. Определены информативные признаки. Описано применение гравитационного алгоритма для оптимизации параметров нечеткого классификатора. Приведены результаты тестирования.

Введение

До настоящего времени наиболее тяжелым осложнением беременности и родов во всем мире являются акушерские кровотечения. Летальность во время беременности и родов от массивной кровопотери и геморрагического шока составляет от 16 до 36 % случаев материнской смертности в разных регионах мира [1] и 18...23 % в России [2], [3]. Более 70 % всех кровотечений в акушерстве относятся к послеродовым кровотечениям [1]. Среди факторов риска развития послеродовых кровотечений выделяют нарушения системы гемостаза: существовавшие ранее заболевания (болезнь Виллебранда, гемофилия), приобретенные коагулопатии (тромбоцитопении), ДВС-синдром, применение антикоагулянтов [4]. Некоторые из этих нарушений существуют и до беременности, но с течением беременности усугубляются, другие развиваются во время беременности. В любом случае выявление женщин с нарушениями свертывания крови должно проводиться на ранних сроках беременности и в дальнейшем такие пациентки требуют постоянного контроля со стороны акушеров и гематологов.

Порядок оказания медицинской помощи по профилю «Акушерство и гинекология» (утвержден Приказом Министерства здравоохранения Российской Федерации от 1 ноября 2012 года № 572н) предусматривает, помимо скрининговых тестов, проведение тромбоэластографии (электрокоагулографии) для контроля гемостаза. Но понимание значения этого метода исследования системы гемостаза, по мнению ряда авторов [5], есть далеко не всегда. Возможно, это связано отчасти с проблемой интерпретации результатов исследования клиницистами. Одним из решений указанной проблемы является программный классификатор анализа данных исследования системы гемостаза, призванный оказать помощь практическим врачам в оценке результатов и, следовательно, ускорить процесс диагностики и лечения.

Проблема классификации относится к проблеме разделения объектов в пространстве признаков по классам или категориям на основе ретроспективных наблюдений, для которых известны значения меток классов. Цель классификации состоит в том, чтобы сопоставить входному вектору одну из меток классов. Классификатор, реализующий отображение векторов признаков в метки классов, представляет собой алгоритм классификации данных. Машины опорных векторов, искусственные нейронные сети, дискриминантный анализ, деревья решений, метод ближайших соседей – вот далеко не полный перечень методов классификации.

Методы нечетких множеств и нечеткой логики широко используются в области классификации данных. К достоинствам нечетких классификаторов можно отнести их хорошую интерпретируемость [6] и отсутствие допущений, необходимых для статистической классификации [7].

Построение нечетких классификаторов предполагает решение следующих основных задач: отбор признаков, генерация базы нечетких правил, оптимизация параметров antecedентов (ЕСЛИ-частей) правил. В рассматриваемой работе для решения первой задачи применяли общеизвестный жадный алгоритм и предложенный авторами дискретный гравитационный алгоритм. Вторая задача решалась с помощью оригинально-

го алгоритма формирования структур нечеткого классификатора на основе экстремальных значений таблицы наблюдений [8]. Оптимизация параметров antecedентов нечетких правил выполнялась непрерывным гравитационным алгоритмом, адаптированным авторами к решению этой задачи.

Целью рассматриваемой работы являлось описание способа оценки свертывающей системы крови у беременных женщин на основе нечеткого классификатора.

Материалы и методы

Исходные данные

Для сбора данных были проведены исследования на анализаторе реологических свойств крови АРП-01 «МЕДНОРД». Материалом для исследования служила цельная венозная кровь беременных женщин на различных сроках гестации с предварительно выявленными нарушениями свертывающей системы крови на скрининговых этапах обследования. В процессе регистрации определяли следующие реологические показатели (признаки) [9]: A_f – начальный показатель агрегатного состояния крови; A_r – интенсивность спонтанной агрегации тромбоцитов; r – период реакции; k – время образования сгустка; AM – фибрин-тромбоцитарная константа крови; T – время формирования фибрин-тромбоцитарной структуры сгустка; F – суммарный показатель ретракции и спонтанного лизиса сгустка.

Полученные значения показателей сравнивались с таковыми для здоровой группы лиц, и по характеру их изменения производилась оценка состояния системы гемостаза по трем составляющим: структурные характеристики, хронометрические характеристики, общее состояние.

Для каждой из трех характеристик выделяли три основных состояния свертывающей системы: нормальное (класс 1), гиперкоагуляционный сдвиг (класс 2), который характеризовался укорочением хронометрических показателей (r , k , T) и увеличением амплитудных характеристик, и гипокоагуляционный сдвиг (класс 0), выявляемый при удлинении хронометрических характеристик и снижении амплитудных параметров. Все состояния требуют определенной реакции со стороны врача-клинициста для коррекции сдвигов и предотвращения потенциальных кровотечений или тромбозов. Однако определение типа изменения зачастую представляет определенную сложность для врача. Для облегчения работы был предложен способ оценки системы гемостаза на основе реологических свойств крови в виде нечеткого классификатора.

Нечеткий классификатор

Классификатор – это функция $c: X \rightarrow C$, где $X = \{x_1, \dots, x_{|X|}\}$ – непустое конечное множество объектов (пациентов); $A = \{a_1, \dots, a_n\}$ – непустое конечное множество признаков (в нашем случае $\{A_f, A_r, k, r, AM, T, F\}$, $n = 7$), таких, что $a: X \rightarrow V_a$ для любого $a \in A$; V_a – множество допустимых значений признака a ; $C = \{c_1, c_2, \dots, c_m\}$ – непустое конечное множество меток классов (в нашем случае $\{0$ – гипокоагуляция; 1 – нормакоагуляция; 2 – гиперкоагуляция, $m = 3$). Значение $c_i = c(a; \theta, S)$ является меткой для вектора признаков a ; θ – вектор парамет-

ров классификатора; S – вектор признаков, участвующих в классификации.

На рис. 1 представлен пример, поясняющий формирование вектора θ . Здесь признак a представлен тремя термами гауссова типа, каждый из которых задается двумя параметрами (b – среднее, c – разброс), входящими в вектор $\theta = (b_{11}, c_{11}, b_{12}, c_{12}, b_{13}, c_{13}, b_{21}, c_{21}, \dots)$. Размерность вектора θ определяется числом входных признаков, участвующих в классификации, и числом термов, описывающих каждый признак.

Основой нечеткого классификатора является база правил. В работе [10] приведены три типа наиболее часто используемых нечетких правил:

- 1) правило, в консеквенте (ТО-части) которого присутствует метка класса;
- 2) правило, в консеквенте которого присутствуют метка класса и степень достоверности (вес) этого правила;
- 3) правило, в консеквенте которого присутствуют метка класса и степени достоверности всех классов.

В нашей работе используется первый тип правил как обладающий более высокой степенью интерпретируемости. Для реализации возможности отбора информативных признаков в процессе построения классификатора предлагается модифицировать продукционное правило следующим образом:

R_{ij} : ЕСЛИ $s_1 \wedge a_1$ есть A_{1i} И $s_2 \wedge a_2$ есть A_{2i} И ... И $s_n \wedge a_n$ есть A_{ni} ТО class есть c_j , где A_{li} – нечеткий терм, характеризующий l -й признак в i -м правиле ($i \in [1, R]$); R – число правил; запись $s_j \wedge a_j$ указывает на наличие ($s_j = 1$) или отсутствие ($s_j = 0$) признака в классификаторе. Вектор $S = (s_1, s_2, \dots, s_n)$ формируется алгоритмом отбора признаков.

В нашей работе метка класса определяется следующим образом:

$$\text{class} = c_{j^*}, \quad j^* = \arg \max_{1 \leq j \leq m} \beta_j,$$

где

$$\beta_j(a) = \prod_{R_{ij}, k=1}^n A_{ki}(a_k); \quad j = 1, 2, \dots, m.$$

Для поступившего на вход классификатора вектора a вычисляется степень его соответствия каждому правилу. Выходом классификатора является метка класса, определяемая по максимальному вычисленному значению.

На множестве обучающих данных $Tr = \{a_p, c_p\}$ определим единичную функцию

$$\text{delta}(p, \theta, S) = \begin{cases} 1, & \text{если } c_p = c(a_p, \theta, S); \\ 0, & \text{иначе,} \end{cases} \quad p = 1, 2, \dots, |Tr|,$$

тогда целевая функция или мера точности классификации может быть выражена следующим образом:

$$E(\theta, S) = \frac{\sum_{p=1}^{|Tr|} \text{delta}(p, \theta, S)}{|Tr|}.$$

Проблема построения нечеткого классификатора сводится к поиску максимума указанной функции. Для оптимизации параметров θ предлагается использовать непрерывный гравитационный алгоритм (GSA_R) оптимизации, а для отбора признаков – дискретный гравитационный алгоритм (GSA_D) и жадный алгоритм (GrA).

Гравитационный алгоритм оптимизации основан на фундаментальных законах тяготения. Популяция оптимизируемых векторов представляет собой систему частиц, между которыми действуют силы притяжения [11]. На вход алгоритму подаются параметры: количество частиц P , количество итераций T , начальное значение гравитационной постоянной G_0 , коэффициент точности поиска δ , малая константа ϵ . Значение гравитационной постоянной рассчитывается на каждой итерации на основе монотонно убывающей функции. Для каждой i -й частицы из популяции θ на t -й итерации рассчитываются следующие физические характеристики: $m_i(t)$ – масса; $a_i(t)$ – ускорение; $V_i(t)$ – скорость. На последнем шаге происходит обновление позиции частицы путем изменения текущих координат на величину, пропорциональную скорости. Расчеты проводятся до истечения итераций T , затем на выход подается вектор с наибольшим значением θ_{best} .

Состав входного вектора различается от вида алгоритма. В непрерывном алгоритме оптимизируется положение термов, поэтому исходный вектор состоит из параметров антецедентов (в нашем случае размерность вектора равна $2mn$, так как в термах гауссова типа два параметра). Остальные частицы являются дубликатами исходного вектора, сгенерированными с помощью нормального отклонения. Ниже приведен псевдокод непрерывного гравитационного алгоритма.

Вход: $P, T, G_0, \delta, \epsilon, S$.

Выход: значение θ_{best} .

$Population = \{\theta_1, \theta_2, \dots, \theta_P\}$;

$t := 1$;

цикл пока ($t \neq T$)

$G[t] := G_0 * (t / T)^\delta$;

цикл по i от 0 до P

$m[i] := (E[\theta_i, S] - E[\theta_{worst}, S]) / (E[\theta_{best}, S] - E[\theta_{worst}, S])$;

цикл по j от 0 до P

$R[i, j] := \|\theta_j - \theta_i\|$;

цикл по d от 1 до $|\theta_i|$

$a_i^d := a_i^d + \text{rand} \cdot G[t] \cdot M[j] \cdot (\theta_j^d - \theta_i^d) / R[i, j] + \epsilon$;

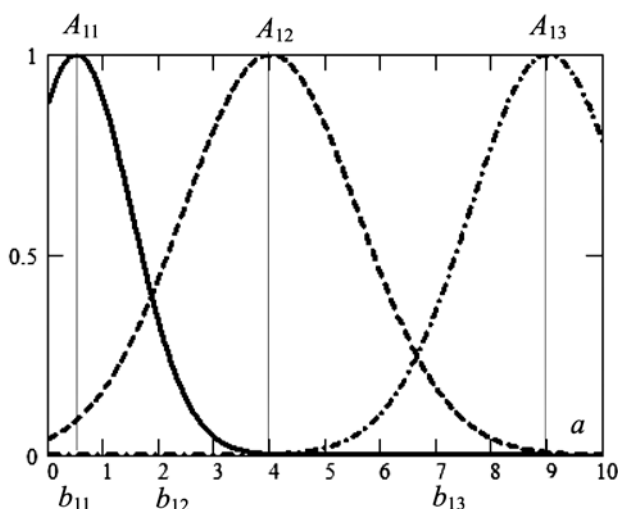


Рис. 1. Нечеткое разбиение признака a

$$V_i^d[t+1] := \text{rand} \cdot V_i^d[t] + a_i^d[t];$$

$$\theta_i^d[t+1] := \theta_i^d[t] + V_i^d[t+1];$$

конец цикла

конец цикла

конец цикла

$t := t + 1;$

конец цикла

вывод $\theta_{best} := \text{Search_best}(\text{Population}).$

Дискретный алгоритм отбирает информативные признаки, поэтому вектор является бинарным: 1 указывает на наличие признака в нечетком правиле, 0 указывает на его отсутствие. Размерность вектора равняется количеству признаков в исходной таблице наблюдений. Первая популяция векторов для дискретного алгоритма генерируется случайно, затем итерационно повторяются те же вычисления, что и в непрерывном алгоритме, за исключением этапа обновления вектора.

Вход: $P, T, G_0, b, e.$

Выход: значение $S_{best}.$

$\text{Population} = \{S_1, S_2, \dots, S_P\};$

$\theta = \{\theta_1, \theta_2, \dots, \theta_P\};$

$t := 1;$

цикл пока ($t \neq T$)

$G[t] := G_0 \cdot (t / T)^b;$

цикл по i от 0 до P

$m[i] := (E[\theta_i, S] - E[\theta_{worst}, S_{worst}]) / (E[\theta_{best}, S_{best}] - E[\theta_{worst}, S_{worst}]);$

цикл по j от 0 до P

$R[i, j] := \|S_j - S_i\|;$

цикл по d от 1 до $|S_i|$

$a_i^d := a_i^d + \text{rand} * G[t] * M[j] * (S_j^d - S_i^d) / R[i, j] + \varepsilon;$

$V_i^d[t+1] := \text{rand} * V_i^d[t] + a_i^d[t];$

$S_i^d[t+1] := 1 / \{1 + \exp(-V_i^d[t+1])\};$

если ($\text{rand}(0,1) < S_i^d[t+1]$) **то** $S_i^d[t+1] := 1;$

иначе $S_i^d[t+1] := 0;$

конец цикла

конец цикла

конец цикла

$t := t + 1;$

вывод $S_{best} := \text{Search_best}(\text{Population}).$

Алгоритм формирования структуры классификатора

Формирование структуры нечеткого классификатора в нашем исследовании ведется на основе экстремумов таблицы наблюдений [8]. Алгоритм создает базу правил, в которой содержится одно правило на каждый класс. Ниже представлено пошаговое описание алгоритма.

Вход: Число классов m , вектор признаков S , таблица наблюдений $\{a_{pi}, c_p\}.$

Выход: База правил классификатора $\Omega.$

Алгоритм:

Шаг 1. $\Omega = \emptyset;$

Шаг 2. Для каждого q -го класса **Do**:

Шаг 2.1. Для каждого i -го признака, у которого $s_i = 1$, **Do**:

Шаг 2.1.1.

$$\min \text{class}_{qi} = \min_p (a_{pi});$$

Шаг 2.1.2.

$$\max \text{class}_{qi} = \max_p (a_{pi});$$

Шаг 2.1.3. Создание симметричного нечеткого термина A_{iq} , накрывающего интервал $[\min \text{class}_{qi}, \max \text{class}_{qi}];$

End do (i);

Шаг 2.2. Создание правила

R_q : ЕСЛИ $s_1 \wedge a_1$ есть A_{1q} И $s_2 \wedge a_2$ есть A_{2q} И ... И $s_n \wedge a_n$ есть A_{nq} **ТО** $\text{class} = c_q;$

Шаг 2.3. $\Omega := \Omega \cup \{R_q\};$

End do (q).

Этапы построения нечеткого классификатора включали в себя применение алгоритмов отбора информативных признаков, формирования структуры и оптимизации параметров классификатора, а также использование критерия выбора классификатора, оптимального по точности и сложности.

На первом этапе жадным или бинарным гравитационным алгоритмом были сформированы группы информативных признаков. Поскольку присутствующая в гравитационном алгоритме случайность способна привести к отбору не самых информативных признаков, формировались несколько групп признаков.

На следующем этапе производилось построение нечетких классификаторов на каждой отобранной группе признаков. Из построенных классификаторов был выбран оптимальный классификатор относительно точности и количества используемых признаков с помощью статистического информационного критерия Акаике [12]. Вычисление критерия осуществлялось по формуле, адаптированной для нахождения компромисса между точностью и сложностью (количеством используемых признаков), приведенной ниже:

$$AIC = \ln[1 - E(\theta, S)] + \frac{2}{F_{\max}} (1 + cF_s),$$

где $E(q, S)$ – мера точности классификатора; F_s – число признаков, используемых классификатором; $F_{\max}$ – максимальное число признаков; c – эмпирический коэффициент штрафа. Если необходимо, чтобы критерий учитывал в большей степени точность, нежели сложность, то коэффициент c должен принимать значение меньше 1; если же наоборот, то значение c должно быть больше 1; если предпочтений нет или они равнозначны, то значение коэффициента равно 1.

Как было отмечено выше, на каждой отобранной группе признаков был построен классификатор, для которого вычислялся AIC . Оптимальный классификатор обладал наименьшим значением AIC .

Результаты

В качестве обучающих использовались данные исследования реологических свойств крови у беременных женщин, состоящих на наблюдении в ОГАУЗ «Родильный дом № 1» г. Томска (всего 500 пациенток).

Входные параметры классификатора – это семь показателей, регистрируемых прибором; показатели A_f, A_r, AM являются целыми числами, показатели r, k, T, F – вещественными.

На основе обучающих данных были построены три классификатора для определения различных характеристик системы свертывания крови – оценки структурных и хронометрических признаков, а также общего ее состояния. Структурные признаки характеризуют вязкость крови, свойство образующегося сгустка, функциональную активность тромбоцитов; хронометрические – первую и вторую фазы свертывания крови. Комплексная оценка является обобщением, кратким заключением по системе свертывания в целом.

Каждый классификатор оперировал с тремя классами, соответствующими состоянию системы гемостаза: гиперкоагуляция, нормокоагуляция и гипокоагуляция.

Каждый классификатор тестировался на различных комбинациях признаков, полученных с помощью жадного или дискретного гравитационного алгоритмов. В *табл. 1-3* приведены результаты работы классификаторов на тестовых выборках после оптимизации непрерывным гравитационным алгоритмом (точность после оптимизации GSA_R). Набор признаков, соответствующий оптимальному классификатору, выделен жирным шрифтом.

Структурные признаки отражают реологические свойства образовавшегося сгустка (*табл. 1*).

Пример базы правил для классификатора, обученного на трех признаках r, AM и T , приведен ниже:

- R_1 : ЕСЛИ r есть $A_3(6,75; 2,05)$ И AM есть $B_1(372,6; 43,8)$ И T есть $C_1(53,78; 12,78)$, **ТО** класс есть 0 (гипокоагуляция);
- R_2 : ЕСЛИ r есть $A_2(8,23; 2,59)$ И AM есть $B_2(684; 62,66)$ И T есть $C_3(55,3; 12,33)$, **ТО** класс есть 1 (нормокоагуляция);

- R_3 : ЕСЛИ r есть $A_1(12; 3,23)$ И AM есть $B_3(802,75; 97,25)$ И T есть $C_2(63,5; 13,63)$, ТО класс есть 2 (гиперкоагуляция).
Здесь нечеткие термы A_i , B_j , C_k задаются функциями гауссова типа с двумя параметрами (b – среднее; c – разброс), например

$$A_3(6,75; 2,05) = \exp \frac{-(r-6,75)^2}{2,05^2}.$$

Таблица 1

Результаты тестирования классификатора для оценки структурных признаков

Алгоритм отбора признаков	GSA_D	GSA_D	GrA
Набор признаков	r, AM, T	k, AM	AM
Точность после GSA_D	95,92	95,92	97,84
A/C	-2,0562	-2,3419	-3,2636

Хронометрические признаки характеризуют процесс образования тромба (табл. 2).

Таблица 2

Результаты тестирования классификатора для оценки хронометрических признаков

Алгоритм отбора признаков	GSA_D	GrA	GrA
Набор признаков	A_i, k, T	r, k, T	r, k
Точность после GSA_D	82,23	93,42	94,22
A/C	-0,5848	-1,5783	-1,9936

Конечным результатом служит комплексная оценка системы гемостаза (табл. 3).

Таблица 3

Результаты тестирования классификатора для оценки общего состояния свертывающей системы

Алгоритм отбора признаков	GSA_D	GSA_D	GrA
Набор признаков	r, k, AM, T	A_i, r, k, AM	r, k, AM
Точность после GSA_D	90,87	91,5	93,15
A/C	-0,965	-1,0365	-1,5381

Заключение

Результаты тестирования нечетких классификаторов позволяют сделать следующие выводы:

- нечеткие классификаторы, настроенные приведенными алгоритмами, имеют хорошие способности к обучению (высокий процент правильной классификации на обучающих выборках);
- для построения нечеткого классификатора необходима лишь часть реологических показателей крови беременных женщин; информативными являются: r – период реакции; k – время образования сгустка; AM – фибрин-тромбоцитарная константа крови;
- предложенный подход может применяться для оценки свертывающей системы крови у беременных женщин.

Работа выполнена в рамках базовой части государственно-го задания Министерства образования и науки Российской Федерации на 2017-2019 гг. Номер 8.9628.2017/БЧ.

Список литературы:

1. Say L., Chou D., Gemmill A., Tunçalp Ö., Moller A.-B., Daniels J., Gulmezoglu A M., Temmerman M., Alkema L. Global causes of maternal death: A WHO systematic analysis // The Lancet Global Health. 2014. Vol. 2. № 6. PP. e323-e333.
2. Куликов А.В., Мартыросян С.В., Обоскалова Т.А. Протокол неотложной помощи при кровотечении в акушерстве. Методические рекомендации. – Екатеринбург, 2010. 38 с.
3. Айламазян Э.К., Репина М.А., Кузьминых Т.У. Еще один взгляд на проблему акушерских кровотечений // Журнал акушерства и женских болезней. 2008. Т. LVII. Вып. 3. С. 3-11.
4. Шифман Е.М., Куликов А.В., Беломестнов С.Р. Интенсивная терапия и анестезия при кровопотере в акушерстве. Клинические рекомендации // Анестезиология и реаниматология. 2014. № 1. С. 76-78.
5. Беломестнов С., Жилин А., Матковский А., Быков А., Куликов А. Интраоперационный контроль гемостаза в акушерстве // Медицина: целевые проекты. 2014. № 18. С. 24-25.
6. Горбунов И.В., Ходашинский И.А. Методы построения трехкритериальных парето-оптимальных нечетких классификаторов // Искусственный интеллект и принятие решений. 2015. № 2. С. 75-87.
7. Scherer R. Multiple Fuzzy Classification Systems // Studies in Fuzziness and Soft Computing. 2012. Vol. 288. 132 p.
8. Ходашинский И.А., Горбунов И.В., Сарин К.С., Субханкулова С.Р. Алгоритмы структурной идентификации компактных и точных нечетких систем // Информационные и математические технологии в науке и управлении. 2016. № 1. С. 82-93.
9. Тюмрин И.И., Шписман М.Н., Стеценко А.И. Методика исследования и интегральной оценки реологических свойств крови (расшифровка и интерпретация графика изменений агрегатного состояния крови) / В кн.: Актуальные проблемы клинических исследований крови. – Томск: STT, 1997. С. 8-18.
10. Cordon O., Del Jesus M.J., Herrera F. A proposal on reasoning methods in fuzzy rule-based classification systems // International Journal of Approximate Reasoning. 1999. Vol. 20. PP. 21-45.
11. Rashedi E., Nezamabadi-pour H., Saryazdi S. GSA: A Gravitational Search Algorithm // Information Sciences. 2009. Vol. 179. PP. 2232-2248.
12. Yen J., Wang L. Application of Statistical Information Criteria for Optimal Fuzzy Model Construction // IEEE Trans. Fuzzy Systems. 1998. Vol. 6. № 3. PP. 362-372.

Илья Александрович Ходашинский,
д-р техн. наук, профессор,
кафедра комплексной информационной безопасности
электронно-вычислительных систем,
Томский государственный университет
систем управления и радиоэлектроники,
Ирина Борисовна Бардамова,
канд. мед. наук, зав. клинико-диагностической
лабораторией ОГАУЗ «Родильный дом № 1»,
Марина Борисовна Бардамова,
студентка,
кафедра комплексной информационной безопасности
электронно-вычислительных систем,
Томский государственный университет
систем управления и радиоэлектроники,
г. Томск,
e-mail: hodashn@rambler.ru