

***Математические
модели в биологии
и медицине***

Использован алгоритм на байесовских сетях по отношению к показателям модифицированной скорости оседания эритроцитов. Построены эффективные байесовские процедуры для этих показателей.

УДК 519.68

Н.Я. ГРИДИНА, А.А. ВАГИС,
А.Л. ТАРАСОВ

ПРОЦЕДУРЫ РАС- ПОЗНАВАНИЯ НА БАЙЕСОВСКИХ СЕТЯХ ДЛЯ ПОКАЗАТЕЛЕЙ СКОРОСТИ ОСЕДАНИЯ ЭРИТ- РОЦИТОВ

Введение. Данная работа посвящена исследованию процедур распознавания на байесовских сетях для показателей модифицированной скорости оседания эритроцитов (МСОЭ). Последняя – это модификация известной методики скорости оседания эритроцитов (СОЭ). Данные СОЭ снимаются на 60-й минуте при комнатной

температуре. Фиксируется показатель оседания красной части крови (эритроцитов). СОЭ находится в норме, если у мужчин оно 1–10 мм, а для женщин 2–15 мм.

МСОЭ снимается динамически (рис. 1) по времени каждые 5 мин. до 90-й включительно. Последний показатель МСОЭ снимается через сутки. Данные снимаются при температуре 37⁰С (внутренняя температура здорового человека) с искусственным гематокритом в 35 % (гематокрит – отношение объема красной части крови по отношению ко всему объему). В кровь также добавляются различные вещества-добавки, меняющие процесс оседания эритроцитов (рис. 1).

Данные МСОЭ снимались для пациентов с глиомами II, III, IV степеней злокачественности, пациентов с метастазами в головной мозг, с черепно-мозговыми травмами (ЧМТ), а также для контрольной группы, в которую входят здоровые люди.

В результате получена динамическая картина оседания эритроцитов. При анализе МСОЭ выделяются два вида графиков – *S* и *V* (рис. 2). Первые соответствуют естественному снятию показателей МСОЭ.

Вторые получены из показателей графиков S . Например, чтобы получить показатель МСОЭ для графика V на 20-й минуте от данного показателя графика S на 20-й минуте вычитается этот же показатель графика S на 15-й минуте.

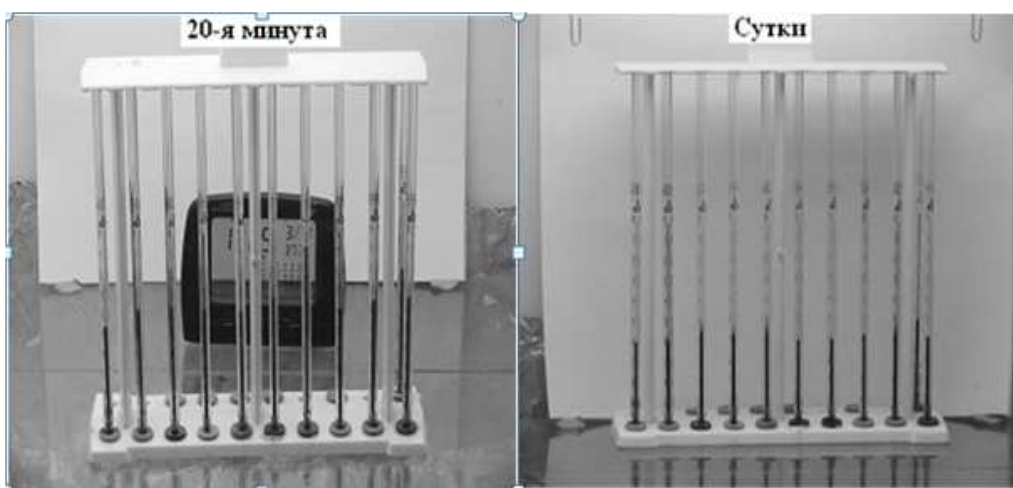


РИС. 1. Модифицированная скорость оседания эритроцитов

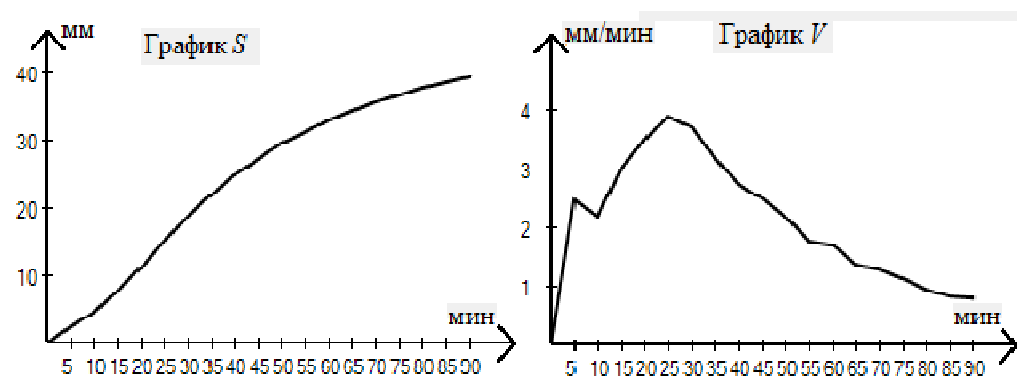


РИС. 2. Графики S и V для МСОЭ

Для анализа показателей МСОЭ предлагается процедура распознавания на байесовских сетях для сети-дерево. Эта процедура базируется на известной формуле Байеса, т.е. мы определяем состояние l ($l \in \{1, \dots, h\}$) для объекта (l в нашем случае – диагноз пациента), описывающийся вектором $\mathbf{x} = (x_1, \dots, x_n)$:

$$P(f = l | x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n | f = l) P(f = l)}{P(x_1, \dots, x_n)}, \quad (1)$$

где $P(x_1, \dots, x_n | f = l)$ – оценка переходных вероятностей; $P(f = l)$ – априорная вероятность класса (при расчетах заменяется частотой класса в обучающей выборке); $P(x_1, \dots, x_n)$ – вероятность появления вектора $\mathbf{x} = (x_1, \dots, x_n)$ во всей обучающей выборке. Отметим, что знаменатель в (1) не принимает участия в процессе распознавания из-за сокращения.

Оценка переходных вероятностей [1] или вероятностное распределение дерева $P^t(x)$ записывается в виде произведения $n - 1$ условных вероятностей (n – количество узлов дерева) и определяется на основе

$$P^t(x) = \prod_{i=1}^{n-1} P(x_i | x_{j(i)}). \quad (2)$$

Таким образом, байесовская процедура для сети-дерево строится аналогично процедуре распознавания для цепи Маркова [2]. Для ее построения используются оценки переходных вероятностей

$$\hat{P}(x_i | x_j, f = l) = \frac{k(x_j, x_i, l)}{k(x_j, l)}, \quad (3)$$

где $f = l$ – состояние объекта, $k(x_j, l)$, $k(x_j, x_i, l)$ – соответственно число наблюдений значений x_j и пар (x_j, x_i) в обучающей выборке.

Прежде чем приступить к подсчету переходных вероятностей, необходимо построить структуру сеть-дерево. Она выводится на основе формулы взаимной информации между парами переменных [1]

$$I(X_i, X_j) = \sum_{x_i, x_j} P(x_i, x_j) \ln \frac{P(x_i, x_j)}{P(x_i)P(x_j)} \geq 0. \quad (4)$$

В соответствии с (4) вычисляют веса всех $n(n - 1) / 2$ ребер и упорядочивают их по величине [1]. Алгоритм построения структуры-дерево выполняется на основе таких шагов.

Шаг 1. Выбирают два ребра, имеющих наибольший вес.

Шаг 2. Проверяют следующее по порядку ребро и добавляют его в дерево, если при этом не образуется петля. В противном случае ребро исключается и выбирается следующее по порядку величины веса ребро.

Шаг 3. Повторяют шаг 2 до тех пор, пока не будут выбраны $n - 1$ ребер.

В работах [3–5] по отношению к показателям МСОЭ применялась дискретизация. Рассмотрим, как можно ее применить для структуры сети-дерево (рис. 3). На рис. 3 рассмотрено только одно ребро. Переходы для остальных ребер строятся аналогичным образом.

Поясним рис. 3. Пусть, для простоты, вся выборка составляет 20 анализов. Показатели МСОЭ находятся в диапазоне 0–60 мм. Разбиваем этот диапазон на равные интервалы, например, на 4 интервала с шагом 15 мм (рис. 3). Выбираем ребро в котором стартовой минутой, т. е. стартовым узлом является 15 минута МСОЭ, а вторым ребром пусть будет 25 минута. В соответствии с выборкой

определяем для стартового узла (15 мин., рис. 3), что в первый интервал (0–15 мм) попало 3 показателя МСОЭ, т. е. 3 анализа, а во второй интервал (15–30 мм) попало 8 показателей, в третий (30–45 мм) попало 7 и в четвертый попало 2 показателя МСОЭ. После этого определяем, сколько показателей МСОЭ из каждого интервала стартового узла переходит во второй узел. Для того чтобы получить вес ребра согласно (4) имеем следующее:

$$P(x_{15,i}, x_{25,j}) = \frac{m_{[15,i][25,j]}}{m_r}, \quad P(x_{15,i}) = \frac{m_{15,i}}{m_r}, \quad P(x_{25,j}) = \frac{m_{25,j}}{m_r}, \quad (5)$$

где $m_{[15,i][25,j]}$ – количество показателей МСОЭ перешедших из состояния i в j , т. е. из стартового узла (15 мин. рис. 3) во второй узел (25 мин. рис. 3). Так, например, если i равно 1, первый интервал, а j равно 2, второй интервал, то согласно рис. 3 $m_{[15,i][25,j]}$ равно 1; $m_{15,i}$ – количество показателей МСОЭ в интервале i стартового узла (если i является первым интервалом, то $m_{15,i}$ равно 3); $m_{25,j}$ – количество показателей МСОЭ в интервале j второго узла (если j является вторым интервалом $m_{25,j}$ равно 6); m_r – количество анализов в выборке, т. е. в нашем случае 20 анализов (рис. 3).

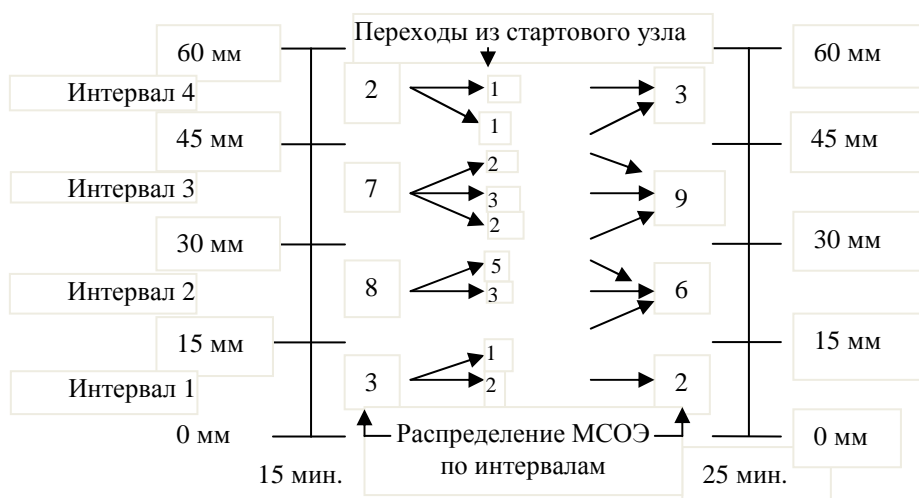


РИС. 3. Переходы значений МСОЭ из 15-й в 20-ю минуты

Расчеты согласно (4) и (5) проводятся для всех возможных комбинаций ребер и согласно описанному выше алгоритму строится структура сеть-дерево. Следует отметить, что после построения дерева каждый из узлов, поочередно, становится корневым узлом дерева, т. е. меняется направление переходов для подсчета переходных вероятностей (рис. 4). На рис. 4, слева, корневой узел дерева – это первый узел, а справа дерево перестраивается, меняя направления переходов таким образом, что корневой узел дерева – это третий узел.

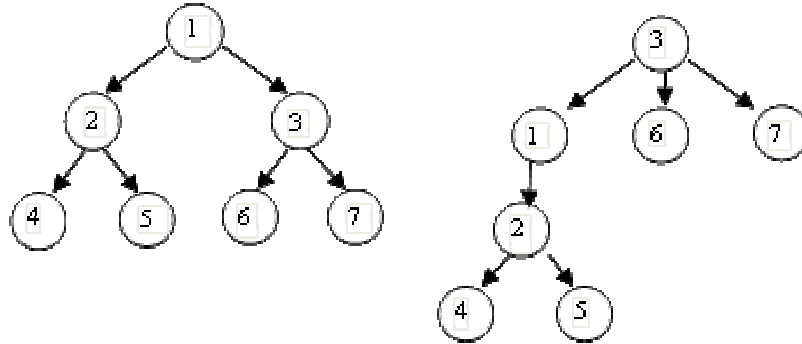


РИС. 4. Структура сеть-дерево

После того, как дерево построено и выбран корневой узел дерева, для каждого из классов, возможных состояний объектов (диагнозов) $l \in \{1, \dots, h\}$, строятся дискретные распределения подобные рис. 3 для всех ребер дерева. После подсчета этих дискретных распределений, по каждому анализу в выборке подсчитывается произведение переходных вероятностей согласно (2) и (3) и на их основе согласно (1) строится процедура распознавания на байесовских сетях.

Отметим, что ширина интервала на рис. 3, равная 15 мм, подбирается перебором таким образом, чтобы получить наилучшую эффективность распознавания. Эта ширина меняется итеративно и находится в диапазоне 0,1–30 мм. Интервалы получаются одинаковыми по ширине. На рис. 5 показано дерево для $MgSO_4$ на основе которого удается распознавать состояние больной или здоровый с точностью 100 %. В класс больные вошли пациенты с глиомами II, III, IV степеней злокачественности (15 человек), а класс здоровые составляет 12 человек. Ширина интервала, аналогичная рис. 3, составляет 15 мм (рис. 5). Показатели анализировались для графика V.

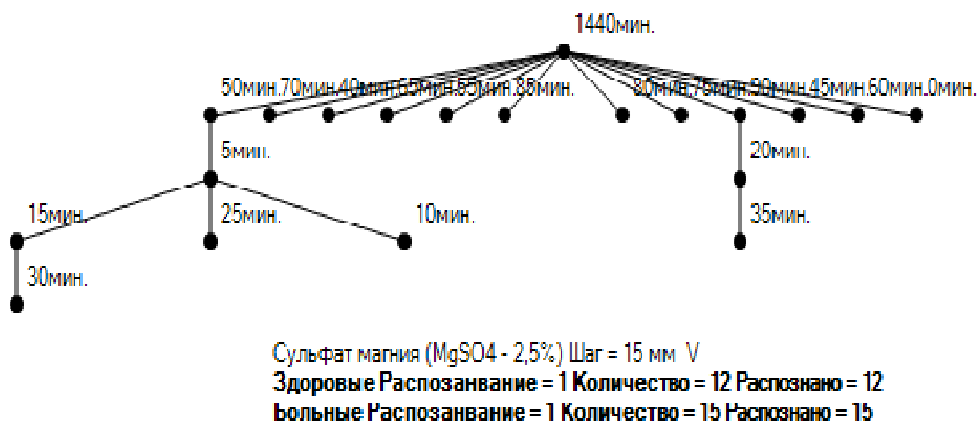
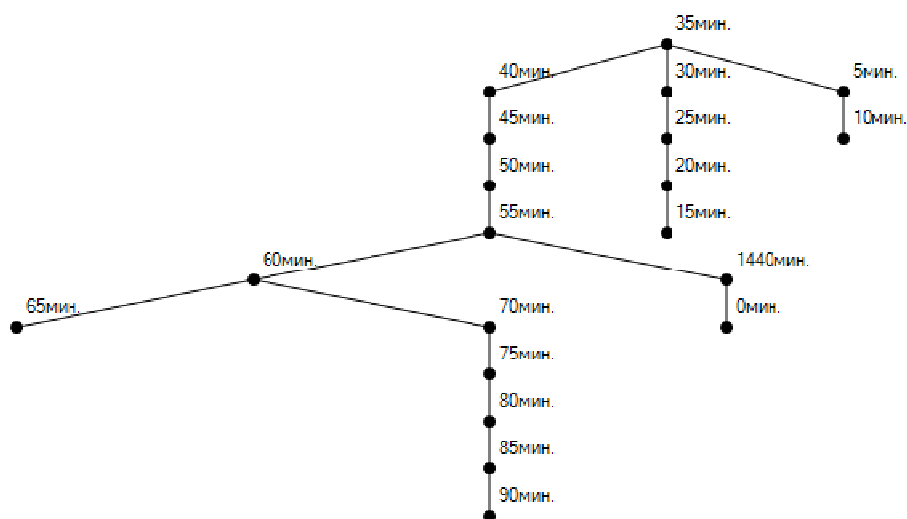


РИС. 5. Структура сеть-дерево для $MgSO_4$

Для сравнения распознавание на однородных цепях Маркова у NaATФ составляет для класса больные 72,3 %, а для класса здоровые – 70,5 %. Оптимальные параметры полученные ранее результаты в работах [3–7] составляют для NaATФ 72,3 % и 76,4 % для классов больные и здоровые соответственно. Оптимальный параметр для верапамила, давший ранее наилучший результат [3–7], обладает эффективностью распознавания 88,5 %, т. е. процедура распознавания на байесовских сетях дает наилучший результат который только можно получить. Более того, ранее нам не удавалось распознавать более двух состояний объектов [3–7]. Распознавались только состояния больной и здоровый. В настоящее время процедуры распознавания на байесовских сетях позволяют различать 5 и 6 классов одновременно: глиомы II ст. зл.; глиомы III ст.; глиомы IV ст.; здоровые; черепно-мозговая травма; метастазы в головной мозг. На рис. 6 показано дерево для верапамила, позволяющее осуществлять такое распознавание.



Верапамил (вер) (1:1) Шаг = 30 мм S
 Глиомы II ст. злок Распознавание = 0,862 Количество = 29 Распознано = 25
 Глиомы III ст. злок Распознавание = 0,909 Количество = 33 Распознано = 30
 Глиомы IV ст. злок Распознавание = 0,925 Количество = 40 Распознано = 37
 Здоровые Распознавание = 0,891 Количество = 46 Распознано = 41
 Черепно-мозговая травма средней тяжести (ЧМТ) Распознавание = 0,947 Количество = 38 Распознано = 36

РИС. 6. Структура сеть-дерево для верапамила

Из рис. 6 видно, что процедура распознавания на байесовских сетях позволяет распознавать:

- глиомы II степени – 86,2 %;
- глиомы III степени – 90,9 %;
- глиомы IV степени – 92,5 %;
- здоровые – 89,1 %;
- черепно-мозговая травма – 94,7 %.

В таблице приведены некоторые результаты работы по распознаванию на байесовских сетях для некоторых веществ-добавок. В класс больные вошли пациенты с глиомами. Отметим также, что для случая, когда в класс больные входят пациенты только со II степенью злокачественности по показателям верапамила можно получить 100 % распознавания воспалительных процессов.

ТАБЛИЦА

Вещество-добавка	Распознавание класса больные, %	Распознавание класса здоровые, %	Шаг разбиения, мм	График
NaATФ	89,4	88,2	10,0	V
Мовалис	92,4	90,3	8,57	V
MgSO ₄	100,0	100,0	15,0	V
Вода	82,5	78,3	20,0	S
Контроль	90,1	87,0	5,45	V
CaCl ₂ (10%)	84,4	82,7	5,45	V
NaCl	71,4	77,4	17,6	S
CaCl ₂ (0,1%)	75,6	75	5,0	V
Верапамил	93,1	100,0	6,66	V
KCl (3%)	89,8	93,5	10	V
NaCl(0,9%)	71,4	80,6	30	S
Верапамил*	100,0	100,0	15	V

Примечание: в класс больные входят глиомы II, III, IV степеней злокачественности; * – в класс больные вошли пациенты только с глиомами II степени злокачественности.

Заключение. Удалось построить новые эффективные процедуры распознавания на байесовских сетях по отношению к показателям МСОЭ. Получены такие структуры сети-дерево, позволяющие распознавать воспалительные процессы, вызванные глиомами с точностью 100 %. Впервые удалось распознавать воспалительные процессы, вызванные различными степенями злокачественности. Удалось получить 100 % распознавание воспалительных процессов вызванных глиомами II степени злокачественности.

Н.Я. Грідіна, О.А. Вагіс, А.Л. Тарасов

ПРОЦЕДУРИ РОЗПІЗНАВАННЯ НА БАЙЄСІВСЬКИХ МЕРЕЖАХ ЗА ПОКАЗНИКАМИ МОДИФІКОВАНОЇ ШВИДКОСТІ ОСІДАННЯ ЭРИТРОЦИТІВ

Побудовані нові ефективні процедури розпізнавання на основі байєсівських мереж за відношенням до показників модифікованої швидкості осідання еритроцитів при гліомах. Уперше вдалося розрізнати запальні процеси обумовлені гліомами різних ступенів злоякісності поміж собою. Підвищилась ефективність розпізнавання запальних процесів при гліомах II ступені злоякісності.

N.Ya. Gridina, O.A. Vagis, A.L. Tarasov

PROCEDURES OF RECOGNITION BY BAYESIAN NETWORKS
FOR THE MODIFICATED ERYTHROCYTE SEDIMENTATION RATE

New effective procedures of recognition with Bayesian networks for modified erythrocyte sedimentation rate at gliomas were built. It is first succeeded to distinguish inflammatory processes that are caused with gliomas between each other. It is improved effectiveness of recognition the inflammatory at the II brain gliomas. Efficiency of recognition of inflammatory processes has risen at gliomas of the second degrees of malignancy.

1. *Вагис А.А.* Байесовские процедуры распознавания на сетях // Проблемы управления и информатики. – 2010. – № 6. – С. 85 – 90.
2. *Гупал А.М., Сергиенко И.В.* Оптимальные процедуры распознавания. – Киев: Наук. думка, 2008. – 231 с.
3. *Сергиенко И.В., Гридина Н.Я., Гупал А.М., Тарасов А.Л.* Применение байесовских процедур распознавания по показателям скорости оседания эритроцитов при глиомах головного мозга // Проблемы управления и информатики. – 2009. – № 1. – С. 35 – 39.
4. *Гридин Н.Я., Тарасов А.Л.* Дифференциальная диагностика воспалительных и онкологических заболеваний // Комп'ютерні засоби, мережі та системи. – 2008. – № 7. – С. 59–65.
5. *Палагин А.В., Гридина Н.Я., Гупал А.М., Тарасов А.Л.* Автоматизированная система анализа показателей скорости оседания эритроцитов при глиомах головного мозга // Проблемы управления и информатики. – 2009. – № 3. – С. 136–143.
6. *Сергиенко И.В., Гридина Н.Я., Гупал А.М., Тарасов А.Л.* Байесовская процедура распознавания глиом головного мозга // Там же. – 2009. – № 5. – С. 150–155.

Получено 05.09.2013

Об авторах:

Гридина Нина Яковлевна,

кандидат медицинских наук, старший научный сотрудник
Института нейрохирургии имени академика А.П. Ромоданова АМН Украины,
E-mail: gridinanina@ukr.net

Вагис Александра Анатольевна,

кандидат физико-математических наук, старший научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины,
E-mail: valex_ic@mail.ru

Тарасов Андрей Леонтьевич,

кандидат технических наук, научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины.
E-mail: freearcher@ukr.net