

# Вероятностные и возможностные методы постановки медицинского диагноза. Компьютерные реализации, программный комплекс

В. А. Газарян, Ю. П. Пытьев, П. Б. Росницкий

Для решения ряда задач медицинской диагностики разработаны и исследованы возможностные методы моделирования медицинских объектов и процесса постановки диагноза, основанные на современной теории возможностей [1]. Статья посвящена применению вероятностных и возможностных алгоритмов и программ обучения и распознавания в условиях нечёткого описания медицинских объектов и изменчивости во времени их вероятностных характеристик. Программный комплекс диагностики заболеваний позволяет проводить процедуру обучения и постановки диагноза в диалоговом режиме с участием врача и применяется в данной работе для решения задачи диагностики острого аппендицита.

**Ключевые слова:** распознавание образов, программный комплекс, задача идентификации, вероятностная модель диагностики, возможностная модель диагностики, гранулирование, алгоритм классификации Кора, острый аппендицит.

## Введение

Математическое моделирование диагностической процедуры связано с решением задачи распознавания образов, в которой заболевания представлены как некоторые классы образов, охарактеризованные совокупностями признаков — симптомов, переменных в количественном и качественном отношении. Поэтому задача медицинской диагностики решается как задача идентификации (классификации), в которой необходимо отнести медицинский объект — пациента — к

одному из  $M$  определённых врачом классов заболеваний, среди которых может быть и класс «норма». Отсутствие выделенных заболеваний рассматривается как отдельный класс, в то же время у диагностируемого объекта могут наблюдаться заболевания, диагностика которых выходит за рамки данного исследования. Задача идентификации решается в два этапа — обучения и принятия решения о диагнозе. В результате обучения находятся характерные значения признаков (симптомов) каждого класса заболеваний на основании выборки объектов с верифицированным диагнозом, то есть обучающей выборки. На этапе классификации диагностируемый объект относится к одному из  $M$  заданных классов, либо к классу «отсутствие выделенных заболеваний».

Вероятностные методы широко применяются при моделировании многих аспектов случайности, неточности и нечёткости, которые свойственны медицинским объектам. Такие объекты часто характеризуются изменчивостью во времени, неформализованным и, в ряде случаев, субъективным характером их описания. В то же время, по многим причинам вероятностные методы на практике оказываются недостаточно эффективными при математическом моделировании медицинских объектов. Вероятностные методы в своей основе используют предположения о стохастической природе объектов и существовании вероятностной модели изучаемых явлений. Если моделируемый объект не является стохастическим, то неточность и нечёткость его описания нельзя охарактеризовать в вероятностных терминах. Однако судить о стохастичности таких объектов не всегда просто вследствие отсутствия критерия вероятностной природы изучаемых явлений в теории вероятностей.

При эмпирическом построении вероятностной модели заведомо стохастического медицинского объекта также возникают серьёзные проблемы: в процессе обследования и опроса пациента его состояние, самочувствие и, как следствие, вероятностные характеристики симптомов его заболевания непредсказуемо изменяются, а их оценки оказываются неадекватными. В этом случае эмпирическое построение вероятностной модели стохастического медицинского объекта невозможно, поскольку данные наблюдений не позволяют оценить эволюцию его вероятностных свойств. Если даже вероятностные характеристики симптомов не меняются во времени, эмпирически построить вероятностную модель объекта с необходимой точностью невозмож-

но вследствие такого большого объема необходимых данных наблюдений, который не удаётся реализовать в медицинской практике.

Для решения ряда задач медицинской диагностики разработаны, исследованы и применены возможностные методы моделирования медицинских объектов и процесса постановки диагноза [2–5], основанные на современной теории возможностей Ю.П. Пытьева [1], отличающейся от теории возможностей Л.А. Заде. Показано, что применение возможностных методов позволяет не только построить математическую модель феномена вероятностной случайности изменяющегося во времени медицинского объекта, но и восстановить её эмпирически, причём точно и на основании почти наверняка конечного числа наблюдений при определённых ограничениях на характер эволюции вероятностной модели.

## 1. Вероятностная модель медицинской диагностики. Вероятностный алгоритм классификации Кора

В вероятностной модели распознавания при существовании вероятностной модели объектов оптимальным является байесовский алгоритм распознавания, минимизирующий ожидаемую вероятность потерь, вызванных неблагоприятными для здоровья пациента последствиями постановки ему ошибочного диагноза. В [6] сформулировано правило постановки медицинского диагноза, минимизирующее математическое ожидание вероятности потерь. Предполагается, что объекты имеют стохастическую природу: каждый из них описывается  $n$ -мерным случайным вектором признаков  $\chi = (\chi^1, \chi^2, \dots, \chi^n)^T$ , принимающим значения  $\mathbf{x} \in \mathbf{X}$ , где

$$\mathbf{x} = (x^1, x^2, \dots, x^n)^T, \quad (1)$$

$x^j \in X^j$  — значение  $j$ -ого признака (симптома),  $n$  — число признаков,  $\mathbf{X} = X^1 \times X^2 \times \dots \times X^n$ .

Решение о принадлежности пациента  $\chi = \mathbf{x}$ ,  $\mathbf{x} \in \mathbf{X}$ , к классу  $k$  принимается при  $\mathbf{x} \in X_k$ , где  $X_1, X_2, \dots, X_M$  — некоторое упорядоченное разбиение множества значений признаков:  $\mathbf{X} = \bigcup_{j=1}^M X_j$ ,  $X_j \cap$

$X_i = \emptyset, i \neq j; i, j = 1, \dots, M$ . Минимум потерь достигается на любом упорядоченном разбиении  $X^*$ , удовлетворяющем условию  $X_j^* \subset \left\{ \mathbf{x} \in \mathbf{X}, S_j(\mathbf{x}) = \min_{1 \leq i \leq M} S_i(\mathbf{x}) \right\}$  [6], где  $S_j(\mathbf{x}) = \sum_{k=1}^M l_{kj} pr(\mathbf{x}|k)pr(k)$  — математическое ожидание вероятности потерь при отнесении пациента  $\mathbf{x}$  к  $j$ -му классу,  $pr(\mathbf{x}|k)$  — переходная вероятность вектора признаков  $\mathbf{x}$  в классе  $k$ ,  $pr(k)$  — априорные вероятности заболеваний,  $k = 1, \dots, M$ ,  $l_{kj}$  — задаваемая врачом вероятность потерь, сопутствующих постановке диагноза  $j$  пациенту, страдающему заболеванием  $k$ , в отличие от используемой ранее в теории статистических решений величины потерь.

Для оценки переходных вероятностей  $pr(\mathbf{x}|k)$  и решения задачи классификации заболеваний в рамках вероятностной модели, разработан вероятностный алгоритм классификации типа Кора [7–9]. В результате обучения алгоритма на выборке пациентов с верифицированным диагнозом определены представительные наборы классов — характерные сочетания значений признаков каждого класса заболеваний.

Для принятия решения о диагнозе требуется не полное описание объекта  $\mathbf{x}$  вектором значений признаков (1), а только его представительные наборы  $D_q^k, q \in \{1, \dots, Q_k\}, k \in \{1, \dots, M\}$ . Приведены оценки Хёфдинга ошибок приближения вероятностей  $pr(D_q^k|i), i \in \{1, \dots, M\}$  представительных наборов их частотами  $\nu(D_q^k|i)$  как для неизменных, так и для меняющихся вероятностных характеристик объектов [6]. На этапе распознавания осуществляется постановка диагноза согласно следующему решающему правилу:

Дан объект  $\mathbf{x} = (x^1, x^2, \dots, x^n)^T$ , который следует отнести к одному из  $M$  классов. Пусть  $D_{q_k}^k$  — представительный набор объекта  $\mathbf{x}$ , имеющий минимальное математическое ожидание вероятности потерь среди всех представительных наборов класса  $k$ :

$$S_k(D_{q_k}^k) = \min_q S_k(D_q^k), q = 1, \dots, Q_k, S_k(D_q^k) = \sum_{i=1}^M l_{ik} pr(D_q^k|i)pr(i). \quad (2)$$

Тогда, согласно оптимальному решающему правилу, объект  $\mathbf{x}$  относится к классу  $k^*$ , в котором математическое ожидание вероятности потерь (2) минимально:  $S_{k^*}(D_{q_{k^*}}^{k^*}) = \min_k S_k(D_{q_k}^k)$ .

Если  $l_{ik} = 1 - \delta_{ik}$ , то есть потери равны нулю при правильном решении ( $i = k$ ) и единице при любом ошибочном решении, решающее правило минимизирует ожидаемую долю неверных диагнозов и имеет следующий вид:

Дан объект  $\mathbf{x} = (x^1, x^2, \dots, x^n)^T$ , который следует отнести к одному из  $M$  классов. Пусть  $D_{q_k}^k$  — представительный набор объекта  $\mathbf{x}$ , имеющий максимальную апостериорную частоту  $\nu(k|D_{q_k}^k)$  на множестве всех представительных наборов  $D_q^k$ ,  $k = 1, \dots, M$ , класса  $k$ :  $\nu(k|D_{q_k}^k) = \max_q \nu(k|D_q^k)$ ,  $q = 1, \dots, Q_k$ . Тогда объект  $\mathbf{x}$  относится к классу  $k^*$ , которому соответствует максимальная частота  $\nu(k^*|D_{q_{k^*}}^{k^*}) = \max_k \nu(k^*|D_{q_k}^k)$ .

## 2. Возможностная модель медицинской диагностики. Возможностный алгоритм классификации Кора

В [2, 4] построена и исследована возможностная модель медицинского объекта, которая фактически является возможностной моделью симптомов его заболевания. По аналогии с вероятностным моделированием,  $i$ -ый объект (пациент) описывается  $n$ -мерным нечетким вектором признаков  $\chi_i = (\chi_i^1, \chi_i^2, \dots, \chi_i^n)^T$ , принимающим значения  $\mathbf{x}_i \in \mathbf{X}$ ,

$$\mathbf{x}_i = (x_i^1, x_i^2, \dots, x_i^n)^T. \quad (3)$$

Пусть  $pr_i = pr(\mathbf{x}_i)$  — значение вероятности того, что  $\chi_i = \mathbf{x}_i$ , а  $p_i = p(\mathbf{x}_i)$  — значение возможности этого события,  $i = 1, \dots, N$ ,  $j = 1, \dots, n$ . В [3, 4] разработана и исследована возможностная модель медицинской диагностики, в которой оптимальным является такое правило постановки диагноза, которое минимизирует возможность сопутствующих ему потерь.

Пусть  $\mathfrak{e}$  — нечеткий элемент, значениями которого являются номера классов заболеваний  $k \in \{1, \dots, M\}$ .  $\phi^{\chi, \mathfrak{e}}(\mathbf{x}, k)$  — возможность равенств  $\chi = \mathbf{x}$ ,  $\mathfrak{e} = k$ ,  $\mathbf{x} \in \mathbf{X}$ ,  $k \in \{1, \dots, M\}$ .  $l_{kd} \in [0, 1]$  — задаваемая врачом возможность потерь при отнесении объекта класса  $k$  к классу  $d$ ,  $d = 1, \dots, M$ . Нечеткий элемент  $\delta$ , принимающий значения на множестве  $\{1, \dots, M\}$ , — решение о состоянии больного (рекомендуемый

диагноз).  $\pi^{\delta|\chi}(d|\mathbf{x})$  — переходная возможность постановки диагноза  $\delta = d$ , когда  $\chi = \mathbf{x}$  — наблюдающиеся у больного симптомы.

Возможность потерь, определяющая качество правила  $\pi^{\delta|\chi}$  постановки диагноза, определена как

$$PL(\pi^{\delta|\chi}) = \sup_{x \in X, k \in \{1, \dots, M\}, d \in \{1, \dots, M\}} \min(l_{kd}, \pi^{\delta|\chi}(d|\mathbf{x}), \phi^{\chi, \mathfrak{A}}(\mathbf{x}, k)), \quad (4)$$

Оптимальным является такое правило  $\pi^{*\delta|\chi}$  постановки диагноза, которое минимизирует возможность потерь (4). Задача минимизации возможности потерь (4) сводится к следующей задаче на минимум для каждого вектора значений признаков  $\mathbf{x} \in \mathbf{X}$ :

$$\max_{1 \leq d \leq M} \min(\pi^{\delta|\chi}(d|\mathbf{x}), P_d(\mathbf{x})) \sim \min_{\pi^{\delta|\chi}(\cdot|\mathbf{x})},$$

где  $P_d(\mathbf{x}) = \max_{1 \leq k \leq M} \min(l_{kd}, \phi^{\chi|\mathfrak{A}}(\mathbf{x}|k))$  (5)

— возможность потерь, сопутствующая решению о постановке диагноза  $\delta = d$  больному при наличии у него симптомов  $\chi = \mathbf{x}$ ,  $\mathbf{x} \in \mathbf{X}$ ,  $d \in \{1, \dots, M\}$ . Для оценки переходных возможностей  $\phi^{\chi|\mathfrak{A}}(\mathbf{x}|k)$  (5),  $k = 1, \dots, M$ , применяется метод гранулирования пространства значений признаков заболеваний [2, 4].

В [6] разработан возможностный алгоритм классификации Кора. Нечёткими представительными наборами класса  $k$  называются гранулы — наборы значений признаков, имеющие в  $k$ -м классе возможность, равную единице, а в остальных классах — меньше единицы. В возможностном алгоритме на этапе обучения по обучающей выборке находятся все нечёткие представительные наборы классов. Для диагностики объекта  $\mathbf{x}$  информативными являются наиболее характерные признаки заболеваний, наблюдаемые у данного пациента, то есть нечёткие представительные наборы, которыми он обладает. Поэтому в качестве  $\phi^{\chi|\mathfrak{A}}(\mathbf{x}|k)$  в (5) используется значение переходной возможности  $p(w_s|k)$  нечёткого представительного набора  $w_s$  объекта  $\mathbf{x}$  при условии принадлежности его к классу  $k$ . На этапе распознавания осуществляется постановка диагноза согласно следующему решающему правилу:

Дан объект  $\mathbf{x} = (x^1, x^2, \dots, x^n)^T$ , который следует отнести к одному из  $M$  классов. Пусть  $w_{s_k}$  — нечёткий представительный набор объекта  $\mathbf{x}$ , имеющий минимальную возможность потерь на множестве всех представительных наборов класса  $k$ :  $P_k(w_{s_k}) = \min_s P_k(w_s)$ ,

$s = 1, \dots, S_k$ ,  $P_k(w_s) = \max_{1 \leq d \leq M} \min(l_{dk}, p(w_s|d))$ . Тогда объект  $\mathbf{x}$  относится к классу  $d^*$ , в котором возможность потерь при постановке диагноза  $d^*$ , минимальна (10):  $P_{d^*}(w_{s_{d^*}}) = \min_d P_d(w_{s_d})$ ,  $d = 1, \dots, M$ .

Если  $l_{dk} = 1 - \delta_{dk}$ , то есть потери равны нулю при правильном решении ( $d = k$ ) и единице при любом ошибочном решении, решающее правило имеет следующий вид:

Дан объект  $\mathbf{x} = (x^1, x^2, \dots, x^n)^T$ , который следует отнести к одному из  $M$  классов. Пусть  $w_{s_k}$  — нечёткий представительный набор, имеющий максимальную возможность в классе  $k$ :  $p(w_{s_k}|k) = \max_s p(w_s|k)$ ,  $s = 1, \dots, S_k$ . Тогда объект  $\mathbf{x}$  относится к классу  $q^*$ , которому соответствует максимальная возможность:  $p(w_{s_{q^*}}|q^*) = \max_q p(w_{s_q}|q)$ .

Проведён сравнительный анализ результатов вероятностного и возможностного моделирования процесса постановки медицинского диагноза, а также вероятностного и возможностного алгоритмов Кора [6]. Если вероятностные характеристики стохастического медицинского объекта непредсказуемо меняются во времени, то его вероятностную модель эмпирически построить невозможно при любом объёме обучающих выборок. Возможностная модель такого объекта может быть восстановлена точно на основании конечного числа наблюдений, если при изменчивости вероятностей векторов значений признаков выполняется условие максимальной согласованности возможностной модели с вероятностной, которое является достаточно слабым ограничением на характер эволюции вероятностной модели [1]. Для практического построения возможностной модели симптомов важна эмпирическая интерпретация возможности, согласно которой при достаточно большом объёме обучающих выборок упорядоченность возможностей векторов симптомов из  $\mathbf{X}$  (или любых подмножеств  $\mathbf{X}$ ) почти наверно точно прогнозирует такую же упорядоченность частот. Фактически, чем выше значимость (возможность) симптомов при определённом заболевании, тем чаще при этом заболевании наблюдаются такие симптомы.

### 3. Результаты применения вероятностного алгоритма Кора к решению задачи диагностики острого аппендицита

Качество врачебной диагностики острого аппендицита (ОА) и хирургические возможности достигли в настоящее время высокого уровня. Однако и в современной клинической практике наблюдаются случаи ошибочной постановки диагноза [10, 11]. Своевременное обнаружение острого аппендицита чрезвычайно важно для здоровья и жизни больного. Поэтому если у врачей возникают сомнения относительно диагноза ОА, предпочтение отдаётся хирургическому вмешательству. Острый аппендицит, как класс заболеваний, в свою очередь делится на три группы: гангренозный, флегмонозный и катаральный. Врачи отмечают, что в этих группах наблюдается ряд общих симптомов, то есть признаки классов пересекаются и необходимо дальнейшее развитие методов вычислительной математики в данной области медицины.

В обучающую выборку входят результаты обследования 103 человек из 4 классов — это 3 группы острого аппендицита: гангренозный (28 чел.), флегмонозный (25 чел.), катаральный (26) и группа больных с неподтвержденным диагнозом (НД) (24 чел.). В качестве признаков врачами выбрано 8 различных симптомов. Каждый признак может принимать от 2 до 4 значений в порядковой шкале в зависимости от степени тяжести симптома. Программная реализация вероятностного и возможностного алгоритмов Кора позволяет проводить обучение и распознавание объектов в диалоговом режиме.

Рассмотрим результаты работы вероятностного алгоритма Кора (рис. 1а). Введём матрицу потерь с элементами  $l_{kd}$ . В данной задаче  $k = 1, \dots, 4$ ,  $d = 1, \dots, 4$ . При диагностике острого аппендицита тяжесть последствий для здоровья больного неодинакова при разных ошибках классификации. Самая опасная форма аппендицита — гангренозный (1 класс). Следующий по тяжести — флегмонозный (2 класс), затем — катаральный (3 класс), то есть тяжесть заболевания убывает с ростом номера класса. Поэтому целесообразно считать, что  $l_{1i} > l_{i1}$  при  $i \neq 1$  и  $l_{12} < l_{13} < l_{14}$ . Гораздо опаснее отнести пациента, страдающего острым аппендицитом, к классу НД, чем совершить противоположную ошибку — поставить ошибочный диагноз ОА здо-



ровому человеку. Следовательно,  $l_{4i} < l_{i4}$  при  $i \neq 4$ . Пример матрицы потерь приведён на рис. 1б.

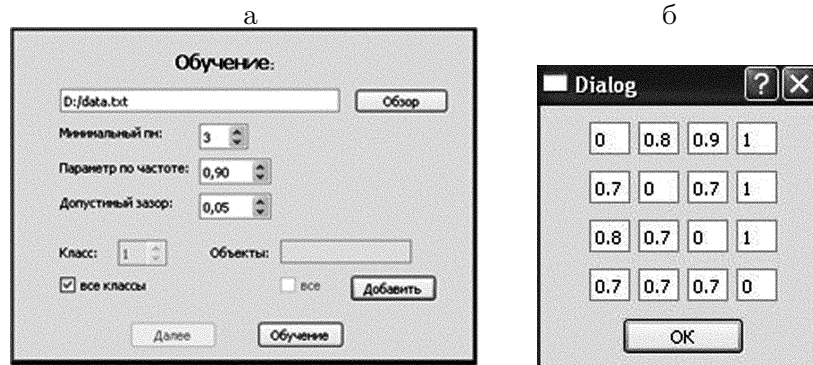


Рис. 1. Выбор объектов и параметров обучения (а). Пример матрицы потерь (б).

В диалоговом окне «Обучение» следует указать минимальную длину представительного набора  $r_{min}$  — минимальное число признаков, входящих в представительный набор. Поскольку одиночные значения признаков не являются представительными наборами, по умолчанию  $r_{min} = 2$ . Чтобы незначительная разница частот не влияла на результат классификации, можно ввести «допустимый зазор» (рис. 2), в пределах которого частоты будут считаться неразличимыми. Зададим порог по частоте встречаемости представительного набора  $\nu_k = 0,5$  для всех  $k = 1, \dots, 4$ .

Выберем для обучения 103 объекта: 28 — из первого класса, 25 — из второго, 26 — из третьего и 24 — из четвертого. Данная выборка сформирована на основании клинических наблюдений больных острым аппендицитом. У части из этих больных наблюдаются симптомы, типичные для данного заболевания с точки зрения врачей. Другая часть обучающей выборки характеризуется «нетипичным» течением болезни. Поскольку для постановки правильного диагноза очень важно как можно точнее и подробнее разобраться в особенностях реакции организма на воспаление аппендикса, и те и другие объекты были выбраны для обучения.

Выберем в соответствующем диалоговом окне файл для классификации, состоящий из 150 объектов: 39 объектов первого класса,

37 — второго, 38 — третьего и 36 — четвёртого. Выборка для классификации включает как объекты обучающей выборки, так и те, которые не были использованы при обучении: это 11 новых объектов первого класса и по 12 новых объектов остальных классов (рис. 2).

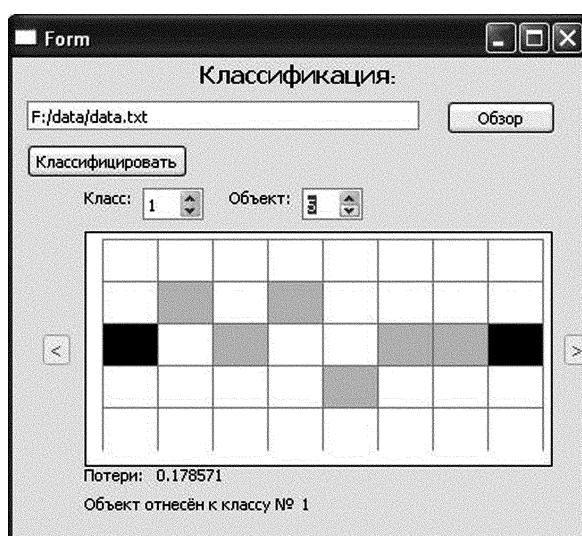


Рис. 2. Выбор данных для классификации. Результат диагностики 5-го объекта первого класса. По горизонтали — 8 признаков (симптомов) объекта, по вертикали — значения каждого признака у данного объекта. Чёрным отмечены значения признаков, входящих в представительный набор, по которому принято решение о диагнозе.

Поскольку представительные наборы, состоящие из двух признаков, имеют частоты, превышающие частоты представительных наборов большей длины, именно они играют решающую роль в распознавании при выборе  $r_{min} = 2$ , а наличие у объектов представительных наборов большей длины не влияет на постановку диагноза. В этом случае каждый объект характеризуется значениями только двух наиболее часто встречающихся признаков. Такое упрощённое описание объекта, имеющего 8 признаков, приводит к низкой точности диагностики (табл. 1). Зададим  $r_{min} = 3$  и получим следующие результаты диагностики: все объекты первого класса имеют верную классификацию, во втором классе 29 объектов ошибочно отнесены к первому классу, 8 — к четвёртому. 21 объект третьего класса отнесен к четвёр-

тому (рис. 3), остальные 17 — к первому. Четвёртый класс распознан безошибочно. При  $r_{min} = 4$  качество диагностики ухудшается (табл. 1), а увеличивать  $r_{min}$  далее не имеет смысла, поскольку ни в одном классе, кроме четвёртого, нет таких длинных сочетаний значений признаков с частотами не менее 0,5, позволяющих верно распознать хотя бы 30% объектов. Применение вероятностного алгоритма показало, что наилучшие результаты диагностики (98% объектов) наблюдаются в первом и четвёртом классах, то есть при самой тяжёлой гангренезной форме аппендицита и при неподтверждённом диагнозе. Классификация менее тяжёлых форм аппендицита большей частью ошибочна: такие больные относятся либо к первому, либо к четвёртому классам, причём, чем «легче» форма аппендицита, тем больше ставится ошибочных диагнозов НД, что является недопустимой ошибкой при диагностике острого аппендицита.

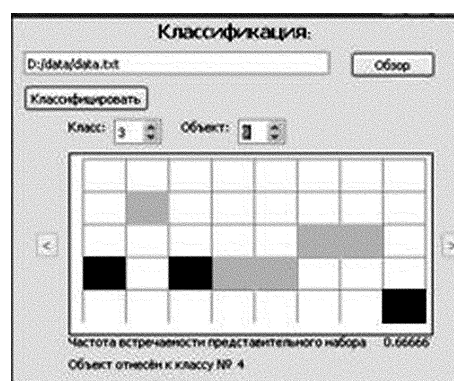


Рис. 3. Ошибочная классификация объекта третьего класса при  $r_{min} = 3$  по представительному набору четвёртого класса вероятностным алгоритмом.

Чтобы уменьшить число ошибочных отнесений больных к четвёртому классу, было увеличено пороговое значение  $\nu_k$  при  $k = 4$ : теперь объект можно отнести к четвёртому классу только при наличии у него представительного набора этого класса, имеющего частоту, превышающую  $\nu_4$ . Анализ результатов показал, что при  $r_{min} = 2$  целесообразно выбрать  $\nu_4 = 0,95$ , при  $r_{min} = 3$   $\nu_4 = 0,9$  и при  $r_{min} = 4$   $\nu_4 = 0,7$ . Получим при  $r_{min} = 3$  безошибочную диагностику первого класса; во втором классе теперь нет ошибочных отнесений к четвёр-

тому классу, однако большинство объектов (33) отнесено к первому классу, 3 объекта — и к первому и ко второму, а один не классифицирован. В третьем классе теперь лишь 2 объекта (вместо 21 при  $\nu_4 = 0,5$ ) отнесены к четвёртому классу, однако остальные 29 — снова к первому, а 8 объектов не классифицированы из-за низких частот собственных им сочетаний признаков. Этот результат значительно лучше предыдущего (при  $\nu_k = 0,5$  для всех  $k$ ), поскольку такие ошибки диагноза, как отнесение больного одной формой ОА к другой его форме или отказ от диагностики, менее опасны для здоровья пациента, чем зачисление больного аппендицитом в класс НД. В четвёртом классе один объект не классифицирован, остальные классифицированы безошибочно. На рис. 4 в качестве примера показан результат диагностики одного из объектов третьего класса, который при  $\nu_4 = 0,5$  (рис. 3) был отнесён к четвёртому классу. При  $r_{min} = 4$  результаты диагностики снова ухудшаются, поскольку сильно возрастает число неклассифицированных объектов (табл. 1).

Таким образом, наилучшие результаты классификации на группы ОА и НД достигаются вероятностным алгоритмом при  $r_{min} = 3$  и  $\nu_4 = 0,9$ : класс ОА распознаётся с точностью 91%, НД — 97%, однако даже при таком подборе параметров два объекта третьего класса ОА отнесены к классу НД.

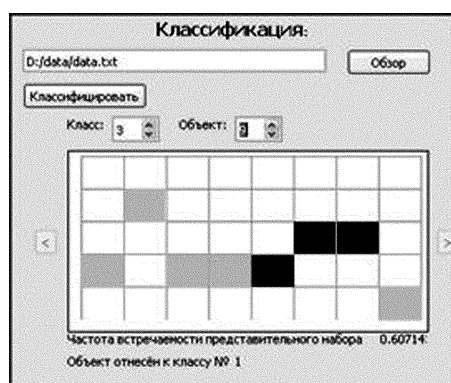


Рис. 4. Результат диагностики объекта третьего класса при  $\nu_4 = 0,9$ ,  $r_{min} = 3$ .

Вычислительный эксперимент также показал, что дополнительное включение в обучающую выборку около десяти «типичных»

представителей трёх групп ОА незначительно увеличивает число правильных решений. Выделить существенные свойства различных форм ОА и провести их классификацию с помощью вероятностного алгоритма не представляется возможным ни при каких значениях параметров обучения.

#### 4. Результаты применения возможностного алгоритма Кора к решению задачи диагностики острого аппендицита

На этапе обучения возможностного алгоритма Кора методом гранулирования [2, 4] находятся все нечёткие представительные наборы классов. Обучающая выборка составлена из тех же 103-х объектов, на которых обучался вероятностный алгоритм. На рис. 5а, 5б, 5в представлены результаты обучения алгоритма — нечёткие представительные наборы каждого из четырёх классов — гранулы значений признаков, имеющие единичную возможность в данном классе и меньшие возможности в других классах. Здесь снова по горизонтали — номера  $j = 1, \dots, n$  признаков,  $n = 8$ , по вертикали — их значения в грануле.

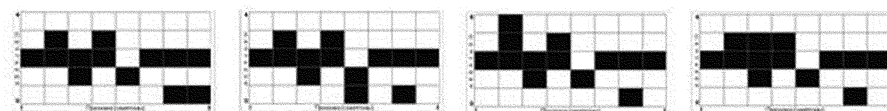


Рис. 5а. Четыре представительных набора первого класса.

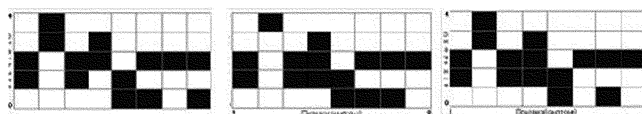


Рис. 5б. Три представительных набора второго класса.

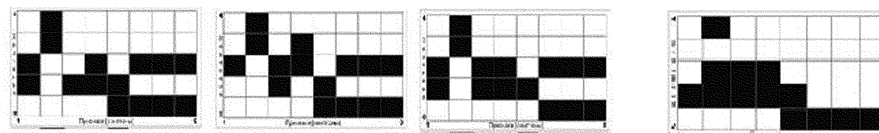


Рис. 5в. Три представительных набора третьего класса (слева) и один представительный набор четвертого класса (справа).

Если в обучающую выборку добавить 47 «типичных», с точки зрения врачей, объектов из четырёх классов, представительные наборы изменятся (рис. 6).

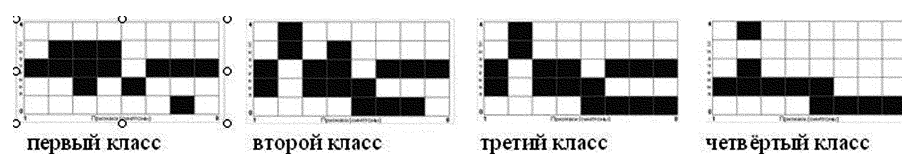


Рис. 6. Представительные наборы классов в результате обучения алгоритма гранулирования на обучающей выборке из 150 объектов.

Результаты классификации выборки, состоящей из 150 объектов, демонстрируются в соответствующем диалоговом окне после выбора в исходном окне входных данных и гранул для классификации. На рис. 7 приведены результаты классификации одного из объектов первого класса. Поскольку вариация значений признаков обучающих объектов каждого класса велика, каждый признак гранулы, имеющей максимальную возможность в классе, может принимать несколько значений (рис. 5, 6). Чем меньше значений в грануле принимает признак, тем более информативным он является для описания симптомов класса. Чем больше у объекта наблюдается признаков, принимающих одно значение в грануле, тем больше значимость данной гранулы при классификации. Отношение числа признаков, принимающих одно значение в грануле, по которой произведена классификация, к общему числу признаков, названо коэффициентом безальтернативности. Если классифицируемый объект имеет значения признаков, входящие в гранулы единичной возможности различных классов, он относится к классу, в котором значение коэффициента безальтернативности максимально. На рис. 7, 8 показано, что для 14 объекта первого класса коэффициент безальтернативности равен  $4/8$  в первом классе и  $1/8$  — во втором, следовательно, данный объект относится алгоритмом к первому классу.

На рис. 9 представлены результаты скользящего контроля (СК) по отдельным объектам [12] вероятностного и возможностного алгоритма в первом классе. Здесь ошибка 1 рода — ошибочное отнесение объектов контрольной подвыборки данного класса к другим классам,

ошибка 2 рода — ошибочное отнесение к данному классу объектов контрольных подвыборок других классов.



Рис. 7. Результат классификации 14 объекта первого класса. Все выделенные значения признаков принадлежат третьей грануле максимальной возможности первого класса (рис. 5). Серым отмечены значения признаков объекта, чёрным — остальные значения признаков, входящие в гранулу.

Точность классификации первого класса составляет 85% (табл. 1). Следует отметить снижение точности классификации при добавлении в обучающую выборку 11 пациентов с «типичными» симптомами гангренозного аппендицита. Поскольку частота встречаемости векторов признаков с типичными значениями симптомов возрастает, объекты, содержащие такие признаки, входят в гранулы единичной возможности (рис. 6) и больные с нетипичным проявлением заболевания ошибочно не относятся к данному классу. В клинической практике встречаются пациенты с замаскированным течением острого аппендицита, особенно в гангренозной форме (1 класс). Из-за пареза нервных окончаний боль исчезает, остаётся общая интоксикация организма. При этом перфорация может наступить в любое время от начала заболевания. Поэтому, чтобы не ошибиться с постановкой такого опасного диагноза, следует формировать обучающую выборку пациентами, у которых гангренозный аппендицит протекает по-разному.



Рис. 8. Гранула второго класса, свойственная 14 объекту первого класса. Значение коэффициента безальтернативности ( $1/8$ ) меньше, чем в первом классе ( $4/8$ ).

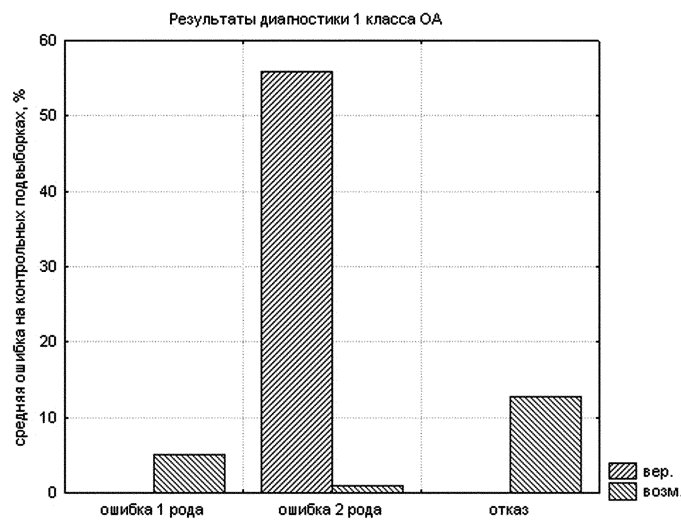


Рис. 9. Результаты СК вероятностного и возможностного алгоритмов Кора при диагностике первого класса.

В разделе 3 было показано, что второй и третий классы (флегмонозный и катаральный аппендицит соответственно) не удалось клас-



сифицировать с помощью вероятностного алгоритма Кора. Эти классы имеют много общих симптомов и в результате обучения возможностного алгоритма получено, что нечёткие представительные наборы этих классов похожи, но отличаются от нечётких представительных наборов остальных классов (рис. 5, 6). Поэтому было принято решение объединить 2 и 3 классы при диагностике ОА и проведении скользящего контроля алгоритмов. На рис. 10 представлены результаты диагностики 2 и 3 классов ОА вероятностным алгоритмом при различных параметрах обучения и возможностным алгоритмом.

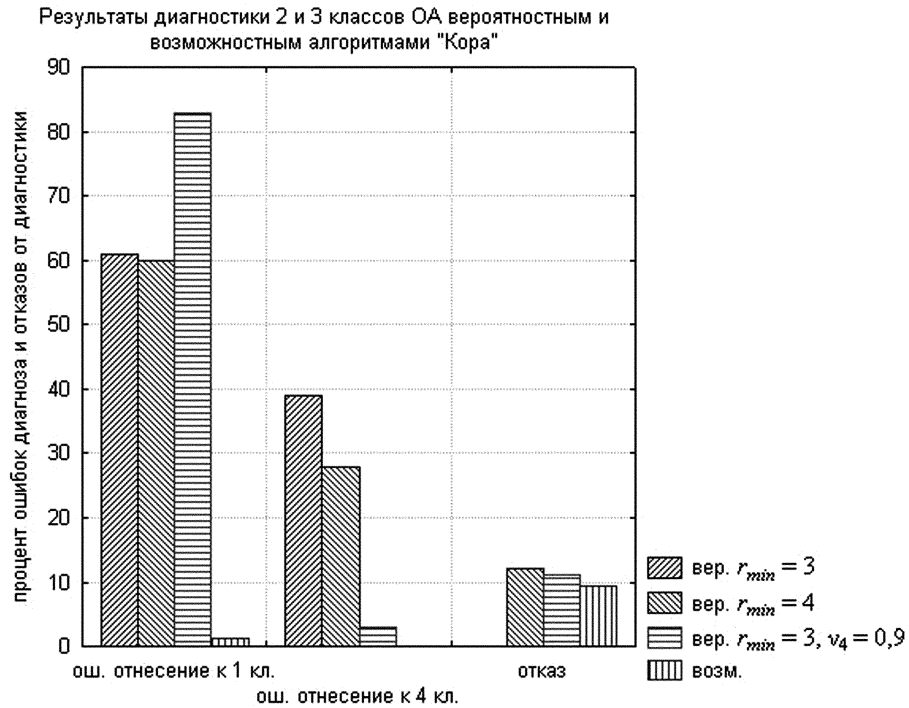


Рис. 10. Диаграмма результатов диагностики. По вертикали — процент ошибок при отнесении объектов 2 и 3 классов: к первому классу, к четвёртому классу, процент отказов от классификации с помощью вероятностного алгоритма (при  $r_{min} = 3$ ,  $r_{min} = 4$ ,  $r_{min} = 3$  и  $\nu_4 = 0,9$ ) и возможностного алгоритма (возм.)

Важно отметить, что у возможностного алгоритма нет ошибочных отнесений объектов второго и третьего классов к первому, кото-

рый, как уже было отмечено, является самым опасным проявлением ОА, и, что самое главное, ни один больной из трёх групп ОА не отнесён вероятностным алгоритмом к группе НД (4 класс). В четвёртом классе, как и в первом, результаты классификации оказались лучше при обучении алгоритма на обучающей выборке, составленной в равной степени из «типичных» и «не типичных» представителей данного класса.

Не имеет смысла включать в обучающую выборку много векторов признаков, характеризующих состояние совершенно здоровых людей. Разумнее формировать её пациентами, обратившимися с жалобами на боли в животе и сопутствующие нарушения, у которых диагноз ОА не подтвердился. Точность классификации четвёртого класса составляет 86%, причём 5 объектов не классифицированы, поскольку у них не наблюдается представительных наборов этого класса. Нечёткий представительный набор любого класса всегда состоит из восьми признаков и вывод о диагнозе в вероятностном алгоритме делается на основании значений всех признаков объекта, в отличие от вероятностного, где длину представительного набора можно задавать от двух до восьми признаков. Объект относится к классу НД только в случае, когда все восемь признаков имеют значения, входящие в гранулу этого класса. Если хотя бы один признак имеет значение, отличное от значения соответствующего признака представительного набора класса, алгоритм не отнесёт объект к этому классу. Вывод о том, что такие пациенты действительно здоровы, может сделать только врач.

## Заключение

В данной работе представлены результаты работы программного комплекса диагностики заболеваний, включающего компьютерную реализацию вероятностного и возможностного алгоритмов Кора, а также позволяющего осуществлять постановку диагноза в диалоговом режиме с участием врача.

Применение вероятностного алгоритма Кора к решению задачи диагностики ОА приводит к следующим результатам:

1. Классификация больных на классы ОА (с точностью 91%) и НД (с точностью 97%) осуществляется только при определённых значе-

Таблица 1. Сводная таблица сравнения результатов классификации острого аппендицита вероятностным и возможностным алгоритмами Кора.

Диагностика объектов	Отнесено к классу	Вероятностный алгоритм				Возможностный алгоритм
		$\nu_i = 0,5$			$r_{min} = 3, \nu_4 = 0,9$	
		$r_{min} = 2$	$r_{min} = 3$	$r_{min} = 4$		
1 класс (гангренозный аппенд.) 39 чел.	1	36	39	38	39	33
	2	0	0	0	0	2
	3	0	0	0	0	0
	4	3	0	0	0	0
	отказ	0	0	1	0	5
	точн., %	92,3	100	97,4	100	85
2 класс (флегмонозный аппенд.) 37 чел.	1	24	29	30	33	0
	2	0	0	0	3	29
	3	0	0	0	0	5
	4	13	8	5	0	0
	отказ	0	0	2	1	3
	точн., %	0	0	0	9	78
3 класс (катаральный аппенд.) 38 чел.	1	7	17	15	29	1
	2	0	0	0	0	10
	3	0	0	0	0	23
	4	31	21	16	2	0
	отказ	0	0	7	7	4
	точн., %	0	0	0	0	60,5
4 класс НД (не аппенд.) 36 чел.	1	0	0	0	0	0
	2	0	0	0	0	0
	3	0	0	0	0	0
	4	36	36	36	35	31
	отказ	0	0	0	1	5
	точн., %	100	100	100	97,2	86

ниях входных параметров. Даже при выборе оптимальных параметров наблюдаются опасные для здоровья пациентов случаи ошибочного отнесения больных ОА к классу НД.

2. Выделить существенные свойства различных форм ОА и провести их классификацию с помощью вероятностного алгоритма Кора не представляется возможным ни при каких значениях эмпирических параметров.

Применение возможностного алгоритма Кора к решению задачи диагностики ОА приводит к следующим результатам:

1. На этапе обучения получены нечёткие представительные наборы 3-х классов ОА и класса НД.

2. На этапе диагностики пациенты с симптомами ОА разделены на два класса ОА и класс НД с точностью диагностики 85% в первом классе ОА, 89% — в объединённом втором классе ОА и 86% — в классе НД без использования эмпирических параметров. Ни один больной ОА не отнесён к классу НД.

3. Исследовано влияние размера и структуры обучающей выборки на результаты обучения и диагностики, что позволило установить улучшение качества диагностики при включении в обучающую выборку «нетипичных» представителей классов. Выполнен скользящий контроль вероятностного и возможностного алгоритмов Кора, свидетельствующий об устойчивости процедур обучения и диагностики.

На основании полученных результатов можно сделать вывод о преимуществах возможностного моделирования медицинских объектов и процесса компьютерной постановки медицинского диагноза.

## Список литературы

- [1] Пытьев Ю. П. Возможность как альтернатива вероятности. Математические и эмпирические основы и приложения. — М.: Физматлит, 2014.
- [2] Газарян В. А., Илюшин В. Л., Пытьев Ю. П., Шаховская А. К. // Вестник Моск. ун-та. Сер. 3. Физика. Астрономия. — 2005. № 4. — С. 3–6.
- [3] Газарян В. А., Иваницкая Н. В., Пытьев Ю. П., Шаховская А. К. // Вестник Моск. ун-та. Сер. 3. Физ. Астрон. — 2006. № 6. — С. 15–18.

- [4] Газарян В. А., Нагорный Ю. М., Пытьев Ю. П., Шаховская А. К. О теоретико-возможностных методах решения задач медицинской диагностики // Интеллектуальные системы. Т. 12. Вып. 1–4. — 2008. — С. 65–82.
- [5] Газарян В. А., Матвеева Т. В., Чехонина Ю. Г., Шаховская А. К. О теоретико-возможностных методах анализа эффективности лечения // Интеллектуальные системы. Т. 14. Вып. 1–4. — 2010. — С. 107–122.
- [6] Газарян В. А. Эмпирическое восстановление математических моделей медицинской диагностики: дисс. . . . канд. физ.-мат. наук. — Москва, 2013.
- [7] Газарян В. А., Иваницкая Н. В., Пытьев Ю. П., Шаховская А. К. // Вестник Моск. ун-та. Сер. 3. Физика. Астрономия. — 2003. № 2. — С. 12–15.
- [8] Kirnos E. A., Pyt'ev Yu. P., Djukova E. V. Training the «Kora» Type Algorithms // Pattern Recognition and Image Analyses. — 2002. V. 12, N 1. — P. 19–24.
- [9] Журавлёв Ю. И. // Журн. вычисл. математики и мат. физики. — 2002. Т. 42, № 9. — С. 1425–1435.
- [10] Doria A. S. // *Pediatr Radiol.* — 2009. Vol. 39. N 2. — P. 144–148.
- [11] Williams R. F., Blakely M. L., Fischer P. E. et al. Diagnosing ruptured appendicitis preoperatively in pediatric patients // *J Am Coll Surg.* — 2009. Vol. 208. N 5. — P. 819–825.
- [12] Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов // Математические вопросы кибернетики. — М.: Физматлит, 2004. Т. 13. — С. 5–36.