

УДК 51-76:57.045

## МЕТОДИКИ РАСПОЗНАВАНИЯ ОБРАЗОВ В ЗАДАЧЕ ВИЗУАЛИЗАЦИИ БИОТРОПНОЙ ФОРМЫ ВОЛН ТЕМПЕРАТУРЫ ВОЗДУХА И АТМОСФЕРНОГО ДАВЛЕНИЯ

© 2011 г. В.А. Ожередов

Институт космических исследований РАН, г. Москва, Россия

Известно, что длительные периоды с неблагоприятной погодой негативно сказываются на заболеваемости сердечно-сосудистыми болезнями. В ряде работ был установлен нижний предел длительности периодов неблагоприятной погоды – около пяти дней. Однако вопрос о том, как именно выглядит негативная динамика основных характеристик погоды – температуры воздуха и атмосферного давления, оставался открытым. В настоящей работе для ответа на указанный вопрос предлагается применить конъюнкционный метод теории распознавания образов. Показано, что данный метод приближает глобально-оптимальную в смысле ошибок распознавания неймановскую критическую область и его можно использовать для решения различных задач в гелиобиологии. В качестве иллюстрации эффективности метода показано, что определенным образом быстро релаксирующие короткие последовательности временных рядов температуры и давления (так называемые температурные волны и волны изменения атмосферного давления) повышают риск возникновения сердечно-сосудистых заболеваний и могут приводить к серьезным органическим поражениям, в частности инфарктам миокарда. Установлено, что температурные волны повышают средний уровень заболеваемости инфарктом миокарда на 90 %, а волны атмосферного давления на 110 %. Атмосферное давление оказалось более биотропным фактором, чем температура воздуха.

*Ключевые слова:* конъюнкция, распознавание образов, температурные волны, инфаркт миокарда.

### Введение

Традиционно решаемы в медицине и гелиобиологии так называемые прямые задачи – выявление биологического эффекта (отклика) при заданной конфигурации параметров (признаков) среды. Примером может служить обнаружение эффектов ухудшения самочувствия при заданных изменениях космической и земной погоды (см., например, [Гурфинкель, 2004; Ожередов и др., 2004; Ревич, Шапошников, 2006]). Однако обратная задача – выяснение, какие именно факторы космической и земной погоды по-

вливали на ухудшение самочувствия, – не менее важна. На формальном математическом языке она формулируется как поиск биотропных областей признаков космической и земной погоды, внутри которых наблюдается интересующий нас отклик – ухудшение самочувствия, прирост заболеваемости, обострение гипертонических кризов и т.д. – отклик интересующего нас I класса.

Данная проблема относится к классу задач распознавания образов. Ее решением служит критическая область, при попадании в которую вектора признаков погоды можно с некоторой уверенностью говорить, что отклик будет принадлежать к интересующему нас I классу. Если вектор признаков оказывается вне этой области, то соответствующий ему отклик принадлежит II классу. Измеренная пара признаки–отклик называется прецедентом. Прецедент принадлежит некоторому классу в соответствии с его откликом. Граница критической области называется сепаратрисой.

В соответствии с формой сепаратрисы методы распознавания образов делятся на линейные и нелинейные. Нелинейная методика более гибкая. В частности, определить многосвязную критическую область способны только нелинейные методы. Наиболее широко используемая нелинейная методика распознавания образов – нейронная сеть. Однако оптимизация алгоритма принятия решений у нейронной сети носит локальный характер. Иначе говоря, критическая область, приближаемая нейросетевым алгоритмом, не является глобально-оптимальной. Поэтому вместо нейронной сети мы обратились к конъюнкциональным методикам принятия решений [Карп, 2002]. При таком способе принятия решений каждая точка пространства признаков принадлежит либо не принадлежит критической области в соответствии с ситуацией, складывающейся в ее окрестности (конъюнкции), – точка пространства признаков попадает внутрь критической области, если внутрь ее конъюнкции попадает значительно больше прецедентов I класса, чем II класса; в противном случае точка лежит вне критической области.

Таким образом, конъюнкциональная методика принятия решений приближает так называемую неймановскую область, которая глобально минимизирует ошибку второго рода (ошибкой второго рода называется попадание в критическую область прецедента второго класса) при заданной верхней границе вероятности ошибки первого рода (т.е. попадания прецедента I класса за пределы критической области) [Леман, 1978].

## Материалы

База данных содержала 2755 измерений ежедневных уровней заболеваемости инфарктом миокарда, а также ежедневные данные по температуре и давлению за пять предшествующих дней в ЦКБ № 1 МПС РЖД за период с 1992 по 2005 гг., а также в Госпитале ветеранов войны (г. Москва) за 7 лет (1995–2001 гг.). Температуры и давления в Москве за анализируемый период брались с сайта <http://moscow-pogoda.ru/archive.html>.

## Методы

В этом разделе детально рассмотрен вариант конъюнкциональной методики распознавания образов, применяемый автором для выявления биотропных форм волн атмосферных температуры и давления. Прежде всего докажем, что существует и является единственной оптимальная критическая область с минимальной вероятностью ошибки второго рода при ограниченной сверху вероятности ошибки первого рода неймановская

область. Однако она явным образом зависит от плотностей векторов признаков прецедентов I класса  $p_\xi(x)$  и II класса  $p_\eta(x)$ . Последние плотности неизвестны, поэтому далее строится их конъюнкциональная точечная оценка. Доказывается, что эта оценка не смещена (ее математическое ожидание равно истинной плотности в данной точке) и выводится формула ее дисперсии, которая в дальнейшем используется для интервальных оценок плотностей  $p_\xi(x)$  и  $p_\eta(x)$ . Далее переходим к приближению неймановской области. Вместо точного критерия принадлежности точки пространства признаков неймановской области строится так называемый эффективный неймановский критерий принадлежности, исходя из интервальных оценок плотностей вероятности:  $p_\xi(x)$  и  $p_\eta(x)$ . Полученное приближение неймановской области параметрически зависит от размеров конъюнкции и порога принятия решений (см. ниже). Показано, что оптимальное значение этих параметров может быть найдено в результате кроссвалидации.

### Доказательство оптимальности неймановской области

Конъюнкциональная методика принятия решений приближает неймановскую область, которая глобально минимизирует ошибку второго рода при заданной верхней границе вероятности ошибки первого рода. Покажем это.

Предположим, что:

- 1) процесс генерации прецедентов стационарен с плотностями вероятностей  $p_\xi(x)$  для I класса и  $p_\eta(x)$  для II класса;
- 2) прецеденты генерируются независимо.

Тогда оптимальной критической областью для принятия решения о том, что ситуация относится к I классу, является *неймановская* область, т.е.

$$G = \left\{ x : \frac{p_\xi(x)}{p_\eta(x)} \geq c \right\}, \int_{G[c]} p_\eta(x) dx = \varepsilon, \quad (1)$$

где  $\varepsilon$  – экспертно определяемый предел ошибки второго рода. Неймановская область минимизирует вероятность ошибки второго рода при условии ограниченности снизу вероятности правильного ответа. В этом нетрудно убедиться, рассмотрев наряду с  $G$  новую область –  $G'$ , которая, по предположению, окажется лучше  $G$  (рис. 1).

По предположению, ошибка первого рода уменьшается при замене  $G$  на  $G'$ . Это означает, что

$$1 - \int_{G'} p_\xi(x) dx < 1 - \int_G p_\xi(x) dx$$

или, сократив на 1 и помножив обе части на  $(-1)$ , получим

$$\int_{G'} p_\xi(x) dx > \int_G p_\xi(x) dx.$$

Разделяя  $G$  и  $G'$  на общие и не общие между собой части, имеем

$$\int_{G \cap G'} p_\xi(x) dx + \int_{G \setminus (G \cap G')} p_\xi(x) dx > \int_{G \cap G'} p_\xi(x) dx + \int_{G \setminus (G \cap G')} p_\xi(x) dx.$$

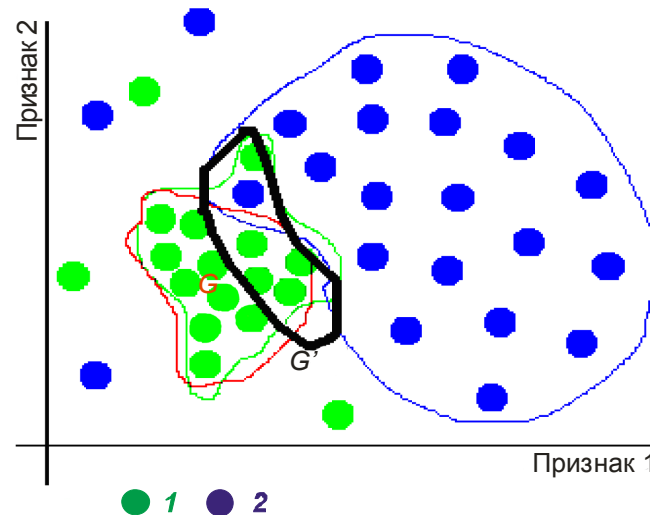


Рис. 1. Неймановское ( $G$ ) и альтернативное ( $G'$ ) критические множества: 1 – I класса; 2 – II класса

Поскольку интегралы по общим частям равны, их можно сократить. Тогда окажется, что

$$\int_{G' \setminus (G' \cap G)} p_{\xi}(x) dx > \int_{G \setminus (G' \cap G)} p_{\xi}(x) dx. \quad (2)$$

Точно так же, в виде суммы интегралов по общей и не общей для двух множеств частей, можно представить и ошибку второго рода:

$$\int_{G' \cap G} p_{\eta}(x) dx + \int_{G \setminus (G' \cap G)} p_{\eta}(x) dx = \varepsilon, \quad (3)$$

как это следует из (1) и

$$\int_{G' \cap G} p_{\eta}(x) dx + \int_{G' \setminus (G' \cap G)} p_{\eta}(x) dx \leq \varepsilon. \quad (4)$$

С другой стороны, исходя из (1), внутри области  $G$

$$\int_{G \setminus (G' \cap G)} p_{\eta}(x) dx \leq \frac{1}{c} \int_{G \setminus (G' \cap G)} p_{\xi}(x) dx, \quad (5)$$

тогда как вне области  $G$

$$\int_{G' \setminus (G' \cap G)} p_{\eta}(x) dx > \frac{1}{c} \int_{G' \setminus (G' \cap G)} p_{\xi}(x) dx. \quad (6)$$

Теперь умножим (2) на  $\frac{1}{c}$ :

$$\frac{1}{c} \int_{G' \setminus (G' \cap G)} p_{\xi}(x) dx > \frac{1}{c} \int_{G \setminus (G' \cap G)} p_{\xi}(x) dx. \quad (7)$$

Компоненты (7) такие же, как правые части неравенств (5) и (6). Отсюда следует цепочка неравенств

$$\int_{G \setminus (G' \cap G)} p_\eta(x) dx > \frac{1}{c} \int_{G' \setminus (G' \cap G)} p_\xi(x) dx > \frac{1}{c} \int_{G' \setminus (G' \cap G)} p_\xi(x) dx \geq \int_{G \setminus (G' \cap G)} p_\eta(x) dx.$$

Таким образом,

$$\int_{G \setminus (G' \cap G)} p_\eta(x) dx > \int_{G \setminus (G' \cap G)} p_\eta(x) dx \quad (8)$$

и

$$\int_{(G' \cap G)} p_\eta(x) dx + \int_{G \setminus (G' \cap G)} p_\eta(x) dx > \varepsilon,$$

что противоречит формуле (4). Поэтому неймановская область, определяемая уравнением (1), единственно оптимальна в смысле минимизации ошибки первого рода при том, что ошибка второго рода в точности равна  $\varepsilon$ . Попытка уменьшить ошибку первого рода ведет к увеличению ошибки второго рода и превышению заданного порога  $\varepsilon$ .

Итак, неймановская область оптимальна, но для ее нахождения необходимы плотности вероятностей, которые неизвестны. Попробуем оценить искомые плотности вероятностей, используя имеющиеся прецеденты.

### **Конъюнкциональная оценка плотности вероятности. Условия несмещенности и малости дисперсии**

Конъюнкциональная оценка плотности вероятности перекликается с методом перебора конъюнкций [Карп, 2008]. Для каждого значения вектора признаков определим в пространстве признаков конъюнкцию – прямое произведение сегментов (рис. 2):

$$\bigcap_{j=1}^q \left[ x^j - \frac{1}{2} d^j; x^j + \frac{1}{2} d^j \right].$$

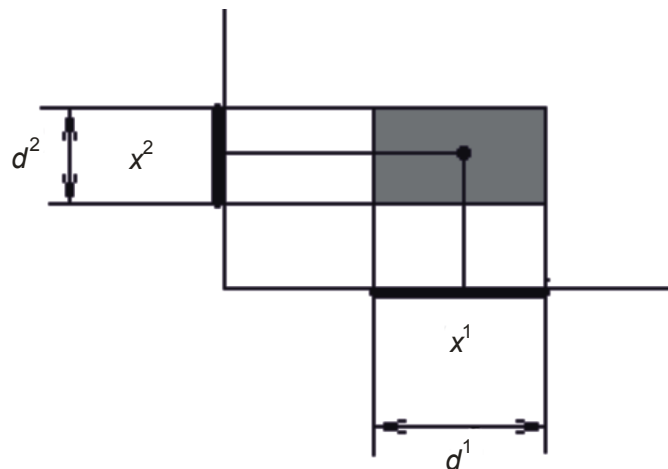


Рис. 2. Схема, иллюстрирующая понятие конъюнкции

Ядерная оценка плотности вероятности  $p(x)$

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N f_x(\chi_i), \quad (9)$$

где

$$f_x(\chi_i) = \begin{cases} \frac{1}{\prod_{j=1}^Q d_j}, & \chi_i \in \bigcap_{j=1}^Q \left[ x^j - \frac{1}{2}d^j; x^j + \frac{1}{2}d^j \right], \\ 0, & \chi_i \notin \bigcap_{j=1}^Q \left[ x^j - \frac{1}{2}d^j; x^j + \frac{1}{2}d^j \right], \end{cases} \quad (10)$$

здесь  $\chi_i$  –  $i$ -й прецедент;  $Q$  – размерность пространства признаков. Как любая точечная оценка, выражение (9) обладает двумя основными характеристиками – смещением и дисперсией. Предположим следующее:

1) размеры конъюнкции по каждому из направлений  $d^j$  настолько малы, что для разложения в ряд Тейлора реальной плотности вероятности для  $\chi_i$ , находящегося в пределах границ конъюнкции, вполне возможно ограничиться первым приближением:

$$p(\chi_i) = p(x) + \sum_{j=1}^Q a^j (x^j - \chi_i^j);$$

2) генерация отдельных прецедентов происходит стационарно;

3) признаки отдельных прецедентов не зависят от признаков других прецедентов.

Тогда соответственно среднее значение и дисперсия оценки (9) составят:

$$\begin{aligned} \langle \hat{p}(x) \rangle &= \frac{1}{N} N \langle f_x(\chi) \rangle = \langle f_x(\chi) \rangle, \\ D(\hat{p}(x)) &= \frac{1}{N^2} N D(f_x(\chi)) = \frac{1}{N} D(f_x(\chi)). \end{aligned} \quad (11)$$

Поскольку  $f_x(\chi)$  не равно нулю и, исходя из наших предположений, исключительно в пределах конъюнкции:

$$\begin{aligned} \langle f_x(\chi) \rangle &= \int_{x^1 - \frac{d^1}{2}}^{x^1 + \frac{d^1}{2}} \int_{x^2 - \frac{d^2}{2}}^{x^2 + \frac{d^2}{2}} \dots \int_{x^Q - \frac{d^Q}{2}}^{x^Q + \frac{d^Q}{2}} \left( p(x) + \sum_{j=1}^Q a^j (x^j - \chi_i^j) \right) \frac{1}{\prod_{j=1}^Q d_j} d\chi^1 d\chi^2 \dots d\chi^Q = \\ &= p(x) + \sum_{j=1}^Q a^j \int_{x^1 - \frac{d^1}{2}}^{x^1 + \frac{d^1}{2}} \int_{x^2 - \frac{d^2}{2}}^{x^2 + \frac{d^2}{2}} \dots \int_{x^Q - \frac{d^Q}{2}}^{x^Q + \frac{d^Q}{2}} (x^j - \chi_i^j) \frac{1}{\prod_{j=1}^Q d_j} d\chi^1 d\chi^2 \dots d\chi^Q = p(x) \end{aligned} \quad (12)$$

в силу симметрии пределов и антисимметрии подынтегральной функции в последнем интеграле формулы.

Дисперсия  $f_x(\chi)$

$$\begin{aligned}
 D(f_x(\chi)) &= \langle (f_x(\chi))^2 \rangle = \\
 &= \int_{x^1 - \frac{d^1}{2}}^{x^1 + \frac{d^1}{2}} \int_{x^2 - \frac{d^2}{2}}^{x^2 + \frac{d^2}{2}} \dots \int_{x^Q - \frac{d^Q}{2}}^{x^Q + \frac{d^Q}{2}} \left( p(x) + \sum_{j=1}^Q a^j (x^j - \chi_i^j) \right) \frac{1}{\prod_{j=1}^Q d_j} d\chi^1 d\chi^2 \dots d\chi^Q - p^2(x) = \quad (13) \\
 &= \frac{p(x)}{\prod_{j=1}^Q d_j} - p^2(x).
 \end{aligned}$$

Таким образом, в силу выражения (12) ядерная оценка плотности вероятности является несмещенной. Более того, в силу (11)–(13) ее дисперсия убывает с ростом  $N$ :

$$D(\hat{p}(x)) = \frac{1}{N} \left( \frac{p(x)}{\prod_{j=1}^Q d_j} - p^2(x) \right). \quad (14)$$

Слагаемые в уравнении (14) неравноценны. Для того чтобы убедиться в этом, разделим ее на квадрат среднего и найдем отношение шум/сигнал:

$$\frac{D(\hat{p}(x))}{p^2(x)} = \frac{1}{N \prod_{j=1}^Q (d_j) p(x)} - \frac{1}{N} \approx \frac{1}{n} - \frac{1}{N}. \quad (15)$$

Исходя из (15), где  $n$  обозначает среднее число прецедентов в конъюнкции,  $N$  – общее число прецедентов заданного класса, второй член в уравнении (14) гораздо меньше первого и им можно пренебречь. Тогда отношение шум/сигнал равно  $1/\sqrt{n}$ .

### **Интервальные оценки плотности вероятности и эффективный неймановский критерий**

Неймановский критерий (1) предполагает, что мы знаем истинные значения плотностей вероятности генерации прецедентов I и II классов. В реальной ситуации нам известна лишь точечная оценка (9) и ее характеристики, такие как несмещенность (12) и состоятельность (15). К счастью, приближенная пропорциональность дисперсии среднему значению (как уже говорилось, вторым слагаемым в скобках уравнения (14) можно пренебречь) позволяет оценить сегмент, с заданной вероятностью покрывающий истинное значение плотности вероятности. В самом деле, если обозначить

$$\langle \hat{p}(x) \rangle = p(x) \equiv \mu, \quad (16)$$

$$D(\hat{p}(x)) = \frac{1}{N \prod_{j=1}^Q (d_j)} (p(x) = k\mu) \quad (17)$$

и учесть, что ядерная оценка (9), будучи суммой большого числа независимых одинаково распределенных слагаемых, имеет гауссовское распределение, получим, что 95-

процентный доверительный интервал покрытия точечной оценки плотности вероятности приходится на сегмент  $S = \mu - 2\sqrt{k\mu}; \mu + 2\sqrt{k\mu}$  или, в терминах неравенств,

$$\mu - 2\sqrt{k\mu} \leq D(\hat{p}(x)) = \frac{1}{N \prod_{j=1}^o (d_j)} (p(x) = k\mu(x) \leq \mu + 2\sqrt{k\mu}). \quad (18)$$

Разрешая неравенство (18) относительно  $\mu$  (истинной плотности вероятности), получим интервал покрытия истинного значения – он лежит между верхней границей сегмента  $S$  при наименьшем значении из возможных оценок  $\hat{p}(x)$  и нижней границей сегмента  $S$  при наибольшем значении из возможных оценок  $\hat{p}(x)$  (рис. 3):

$$\begin{cases} \sqrt{\mu} \equiv t; \\ \begin{cases} t_1^2 + 2\sqrt{k}t_1 - (\hat{p}(x)) = 0, \\ t_2^2 - 2\sqrt{k}t_2 - (\hat{p}(x)) = 0, \end{cases} \end{cases} \quad (19)$$

где  $[t_1, t_2]$  – искомый сегмент. Каждое из уравнений в формуле (19) имеет по два корня, из которых по одному корню – положительные:

$$\begin{aligned} t_1 &= \sqrt{\hat{p}(x) + k} - \sqrt{k}, \\ t_2 &= \sqrt{\hat{p}(x) + k} + \sqrt{k}, \end{aligned} \quad (20)$$

$$(\sqrt{\hat{p}(x) + k} - \sqrt{k})^2 \leq p(x) \leq (\sqrt{\hat{p}(x) + k} + \sqrt{k})^2. \quad (21)$$

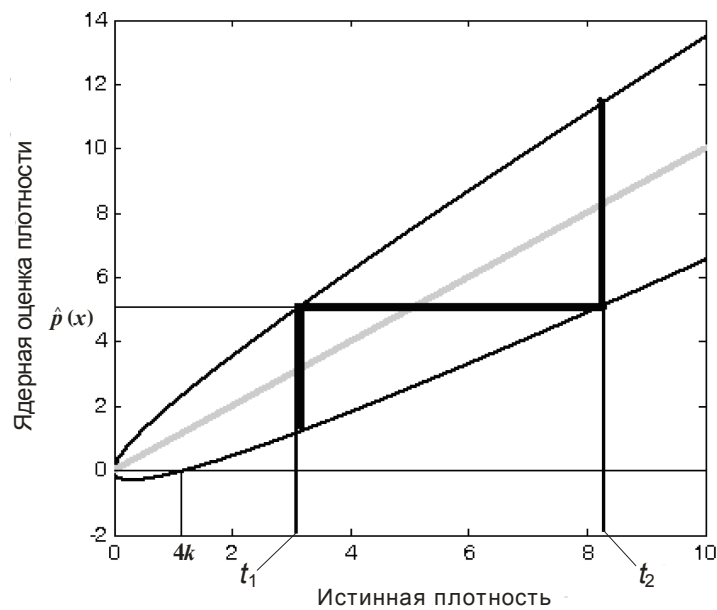


Рис. 3. Пояснения к получению интервальной оценки истинной плотности вероятности по ее измерениям



Стоит отметить тот факт, что при нулевой оценке плотности (внутри конъюнкции ничего не попадает) верхняя граница сегмента  $[t_1, t_2]$  :

$$4k = \frac{4}{N \prod_{j=1}^Q d^j} .$$

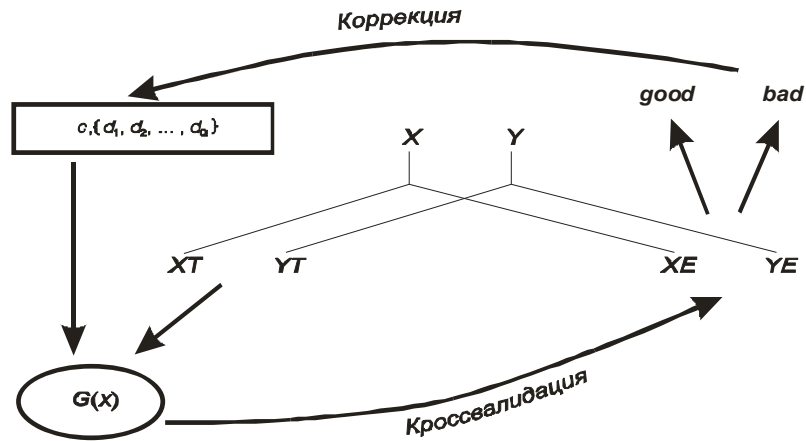
Как уже было отмечено, истинные плотности в критерии (1) нам недоступны, но с вероятностью  $0.95^2=0.9025$  для каждой из двух плотностей верно неравенство (21), и теоретически истинные плотности могут лежать в любом месте соответствующих неравенству (21) сегментов. Отношение (1) гарантированно (разумеется, все с той же вероятностью 0.9025) больше константы  $c$ , если нижняя граница интервала  $p_\xi(x)$  в  $c$  раз больше верхней границы интервала  $p_\eta(x)$ . Таким образом, приходим к *эффективному* неймановскому критерию:

$$G = \left\{ x : \frac{\left( \sqrt{\hat{p}_\xi(x) + k} - \sqrt{k} \right)^2}{\left( \sqrt{\hat{p}_\eta(x) + k} + \sqrt{k} \right)^2} \geq c \right\}. \quad (22)$$

### **Кроссвалидационная оценка параметров алгоритма распознавания**

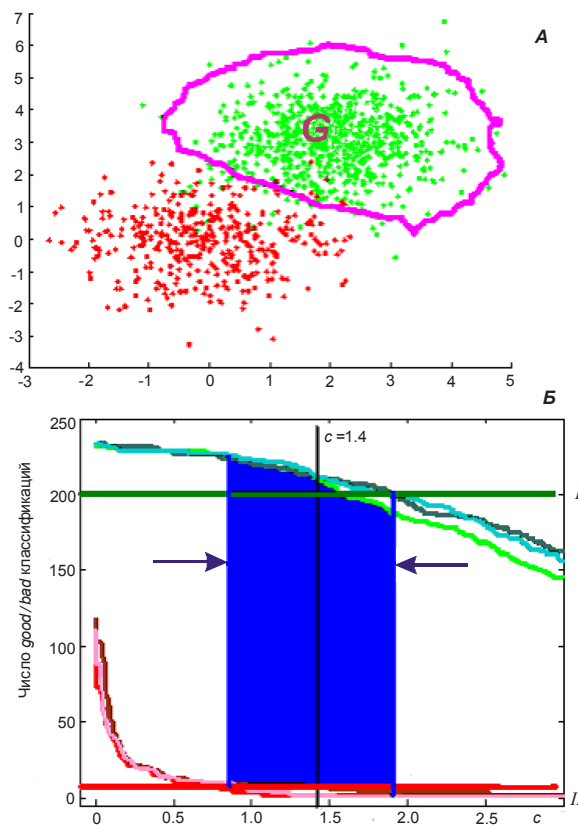
Алгоритм распознавания имеет  $Q+1$  параметр:  $Q$  размеров конъюнкций вдоль каждой из осей признаков и порог принятия решений  $c$ . Размер конъюнкции  $d_i$  вдоль оси признаков  $i$  пропорционален дисперсии вдоль этой оси с коэффициентом пропорциональности  $p$  – масштабным параметром конъюнкции. Параметры  $c$  и  $p$  должны быть оптимизированы в процессе обучения алгоритма. Значения параметров считаются удовлетворительными, если классификация вновь поступающих на вход системы распознавания ситуаций имеет хорошие характеристики. Единственный способ «смоделировать» поведение алгоритма при поступлении на вход прецедентов, не принимавших участия в формировании оптимальной критической области  $G$ , – кроссвалидация. Сущность кроссвалидационного метода состоит в том, что вся выборка делится на две части – обучающую и экзаменационную. При заданных априори масштабном параметре конъюнкции  $p$  и пороге принятия решений  $c$  обучающая выборка задает критическую область  $G = G(p, c)$ , характеристики которой (частоты ошибок первого и второго рода) измеряются при помощи экзаменационной выборки. Полезные свойства алгоритма распознавания характеризуют две величины: количество правильно классифицированных ситуаций I класса (так называемый параметр *good*) и число попавших в критическую область ситуаций II класса (параметр *bad*). Эксперимент с разбиением повторяется несколько раз, чтобы выявить репрезентативность базы данных прецедентов в целом (рис. 4).

Выбор размеров конъюнкции и порога принятия решений  $c$  зависит, прежде всего, от двух обстоятельств: 1) экспертного выбора нижнего порога для *good* и верхнего порога для *bad* (этот выбор обычно тривиально следует из разумного ограничения рисков, связанных с принятием решений); 2) поведения функциональных зависимостей *good* и *bad* от порога  $c$ .



**Рис. 4.** Схема одного цикла кроссвалидационной коррекции параметров алгоритма распознавания  
 $X, Y$  – матрицы из строк-прецедентов I и II классов;  $XT, XE$  – соответственно обучающая и экзаменационная выборки прецедентов I класса;  $YT, YE$  – то же II класса

Рассмотрим типичные зависимости  $good(c)$  и  $bad(c)$  на примере тестовой задачи разделения пространства признаков для двух нормальных совокупностей (рис. 5).



**Рис. 5.** Оптимальная критическая область принятия решений  $G$  (A) и зависимости  $good(c)$  и  $bad(c)$  (B). Пучки из трех кривых соответствуют трем разным циклам кроссвалидации

Горизонтальные линии на B: I – минимальное количество правильно классифицированных ситуаций I класса; II – максимальное число неправильно классифицированных ситуаций II класса,  $p = 0.33$ . Затемненная область соответствует значениям  $c$ , отвечающим экспертно заданным нижнему порогу для  $good$  и верхнему порогу для  $bad$

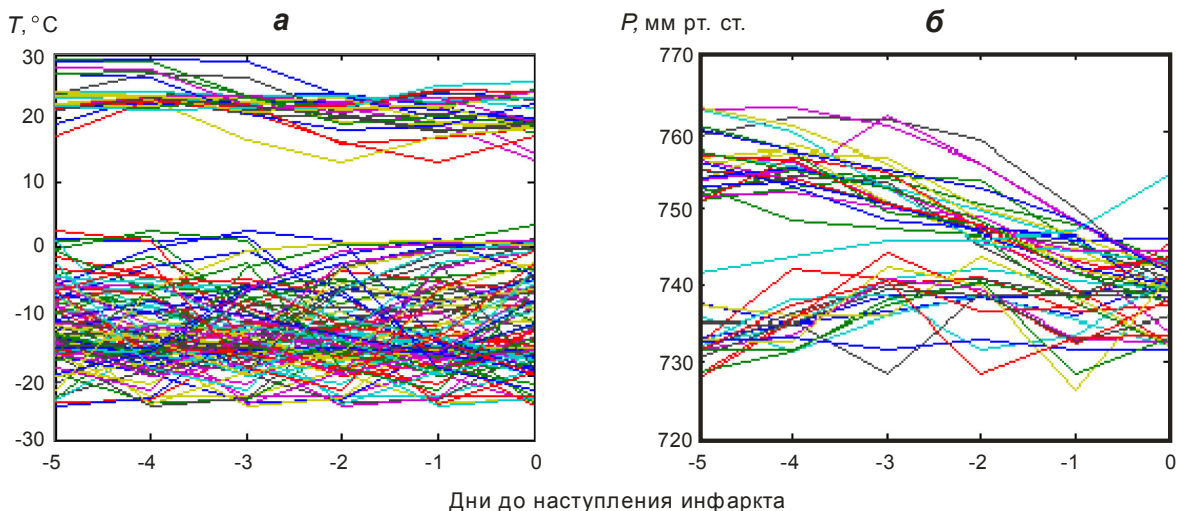
Из-за монотонной вложенности областей  $G(c)$  друг в друга (это означает, что для  $c' > c$ :  $G(c') \subset G(c)$ ) зависимость  $c$  не убывает, а  $bad(c)$  не возрастает). Поэтому значения  $c$ , где величина  $good$  больше минимального порога для  $good$ , находятся слева от точки пересечения кривой  $good(c)$  с линией уровня порога для  $good$ , а значения  $c$ , где величина  $bad$  меньше максимального порога для  $bad$ , находятся справа от точки пересечения кривой  $bad(c)$  с линией уровня порога для  $bad$ . Оба условия, наложенные экспертом на число правильно классифицированных ситуаций I класса и число неправильно классифицированных ситуаций II класса, выполняются в затемненной области на рис. 5. Ее ширина (в единицах  $c$ ) называется *шириной правдоподобия* и характеризует устойчивость алгоритма распознавания локальным нестационарным возмущениям плотности вероятности появления прецедентов обоих классов. За оптимальное значение  $c$  следует принять середину означенной области, как показано на рис. 5. Верхняя часть этого же рисунка иллюстрирует границы критической области  $G(c)$ , соответствующей оптимальному значению порога  $c$ .

### **Практические результаты распознавания погодных условий, неблагоприятных в отношении возникновения гипертонического криза, и их обсуждение**

Рассмотрим обнаружение воздействия волн изменения атмосферных температуры и давления на организм человека с использованием описанного выше алгоритма распознавания образов.

Будем считать, что каждый прецедент имеет отклик – разность заболеваемости в текущий день и среднего уровня заболеваемости инфарктом миокарда (ИМ), и 6-мерный вектор признаков – температура/давление в предшествующие 5 дней и в текущий день. Таким образом, критическое множество представляет собой такие варианты изменения температуры/давления за 6 дней, чтобы средняя заболеваемость на шестой день была наибольшей. Данная проблема решалась описанным выше конъюнкциональным методом распознавания образов. Были получены оптимальные с точки зрения параметров  $good$  и  $bad$  значения порога принятия решений  $c$  и масштабного параметра  $p$ .

На рис. 6 представлен явный вид волн изменения температуры и давления, которые принадлежат критической области. Видно ярко выраженное разделение температурных волн на тепловые и холодные (рис. 6,а).



**Рис. 6.** Биотропные волны изменения атмосферных температуры (а) и давления (б), попавшие внутрь критической области с наилучшими параметрами  $good$

Тепловыми волнами являются достаточно быстрые (в течение одного-двух дней) переходы через значения  $+20\text{ }^{\circ}\text{C}$ . Этот уровень температуры оптимален для существования человеческого организма [Ревич, Шапошников, 2006], поэтому при температурных уровнях выше и ниже этого значения включаются разные компенсаторные механизмы.

Достаточно быстрые переходы через этот уровень разлаживают адаптационную систему человеческого организма. Холодовых волн за анализируемый период оказалось гораздо больше, и они также представляют достаточно быстрые и значительные колебания температуры ниже  $0\text{ }^{\circ}\text{C}$ .

В ряде работ было введено понятие тепловой волны в виде ситуации, когда температура держится выше отметки  $+22.7\text{ }^{\circ}\text{C}$  на протяжении 5 дней и в течение трех последующих дней она должна превышать  $+25\text{ }^{\circ}\text{C}$ . Холодовой волной считалась ситуация, когда температура держится ниже  $-14.4\text{ }^{\circ}\text{C}$  в течение не менее 9 последовательных дней, из которых в течение как минимум 6 последовательных дней она должна быть ниже  $-19.3\text{ }^{\circ}\text{C}$ . Имеются и другие определения температурных волн [Ревич, Шапошников, 2006]. Подход автора статьи отличается тем, что выявляются биотропные формы температурных волн, а не задается форма волны априорно. Биотропными волнами изменения давления являются релаксации (быстрые подъемы и спады) к нормальному давлению (940 мм рт.ст.).

Далее ставился вопрос – насколько велико превышение  $\Delta$  средней заболеваемости по соответствующим критическим областям, зависящим от  $p$  и  $c$ , над средним ее значением по всей выборке. Была построена диаграмма  $\Delta = \Delta(p, c)$ , показанная на рис. 7. Как и следовало ожидать, максимумы  $\Delta$  совпали с оптимальными значениями  $p, c$ , полученными выше кроссвалидационным способом.

Для температурных волн (рис. 7,а) максимум  $\Delta$  составил 0.9, что соответствует 90%-ному превышению заболеваемости инфарктом миокарда, связанному с температурными волнами.

Для волн атмосферного давления (рис. 7,б) максимум  $\Delta$  составил 1.1, что соответствует 110%-ному превышению заболеваемости, связанному с волнами давления. Таким образом, вариации атмосферного давления – более биотропный фактор относительно заболеваемости инфарктом миокарда, чем вариации температуры.

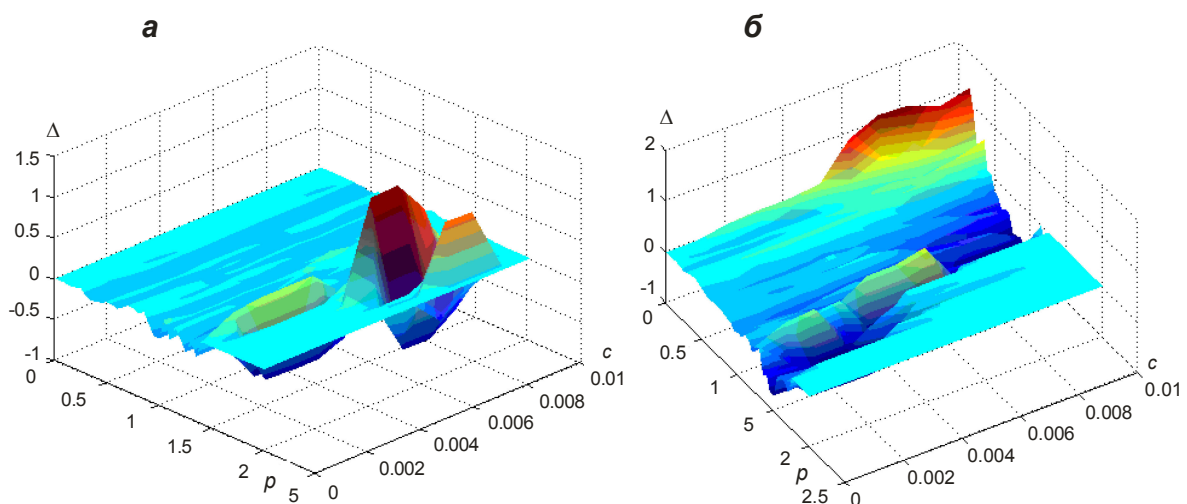


Рис. 7. Зависимость  $\Delta = \Delta(p, c)$  для температурных волн (а) и волн атмосферного давления (б)

### Выводы

Предложенный метод позволяет прогнозировать дни с повышенной вероятностью возникновения случаев инфаркта миокарда по вариациям температуры воздуха и атмосферного давления. Для практических целей важно не только предсказание неблагоприятных дней, но и представление о том, какого количества превышений средней заболеваемости можно ожидать в это время. Получены биотропные формы волн изменений температуры воздуха и атмосферного давления, эффективные с точки зрения возникновения инфаркта миокарда. Показано, что в неблагоприятные периоды изменений температуры воздуха и атмосферного давления наблюдается возрастание количества случаев возникновения инфаркта миокарда вплоть до двух раз.

### Благодарности

Автор благодарит Ю.И. Гурфинкеля за предоставленные данные по заболеваемости инфарктами миокарда, Т.К. Бреус за ценную критику и советы при проведении исследований.

Работа выполнена при поддержке гранта Программы Президиума РАН «Фундаментальные науки – медицине».

### Литература

- Гурфинкель Ю.И.* Ишемическая болезнь сердца и солнечная активность. М.: ИИКЦ «Эльф-3», 2004. 170 с.
- Карп В.П.* Вычислительные методы анализа временных рядов в биологии и медицине. М., 2002. 43 с.
- Карп В.П.* Методы и средства контроля и диагностики сложноорганизованных объектов. М., 2008. 62 с.
- Леман Э.Л.* Проверка статистических гипотез. М.: Наука, 1978. 408 с.
- Ожередов В.А., Бреус Т.К., Гурфинкель Ю.И., Ревич Б.А., Митрофанова Т.А.* Влияние отдельных погодных факторов на развитие острых кардиологических патологий // Биофизика. 2010. Т.55, вып. 1. С. 133–144.
- Ревич Б.А., Шапошников Д.А.* Климатические условия, качество атмосферного воздуха и смертность населения г. Москвы в 2000–2006 гг. // Климат, качество атмосферного воздуха и здоровье москвичей / Под ред. Б.А. Ревича. М.: Изд-во «АдамантЪ», 2006. С. 102–140.

### Сведения об авторе

**ОЖЕРЕДОВ Вадим Андреевич** – кандидат физико-математических наук, ведущий математик, Институт космических исследований РАН. 117997, г. Москва, ул. Профсоюзная, д. 84/34. Тел.: (8) 916-551-85-42. E-mail: ojymail@mail.ru.

## PATTERN RECOGNITION TECHNIQUE FOR THE QUEST OF VISUALISATION OF BIOTROPIC FORM OF TEMPERATURE AND PRESSURE WAVES

V.A. Ozheredov

Space Research Institute, Russian Academy of Sciences, Moscow, Russia

**Abstract.** Undoubtedly the long duration of adverse weather negatively affect a human cardiovascular system. In a number of works the lower limit of adverse weather periods – about five days – has been

established. However the question on negative dynamics of the basic characteristics of weather – atmospheric temperatures and pressure – remained open. In the present work a conjunctive method of the theory of pattern recognition will be applied to the heliobiology. It is shown that the given method approaches a global-optimum of Neuman's critical area in sense of errors of recognition. As an illustration of this method efficiency it is shown that quickly relaxing short sequences of time series of atmospheric temperatures and pressure (so-called temperature waves and atmospheric pressure waves), cause growth of cardiovascular diseases, which can lead to serious organic defeats, such as myocardial infarction. It is shown that temperature waves can increase daily number of myocardial infarctions above the average value up to 90 %, and pressure waves – up to 110 %. Pressure has appeared to be more biotropic factor, than atmosphere temperature.

*Keywords:* conjunction, pattern recognition, temperature waves, myocardial infarction.