

УДК 512.816

ОМАРОВ М.А., д.т.н., профессор (ХНУРЭ),
ЦЕХМИСТРО И.И., к.ф.-м., доцент (НТУ «ХПИ»),
ТРУБИЦИН А.А., инженер (ХНУРЭ)

Поиск способов построения информационных систем оценивания сложности ситуации в медицине и на транспорте на примере конкретных ситуаций

Проведенные экспертные оценивания с использованием средневзвешенных оценок и Байесовское оценивание четко показали необходимость учета приоритетности признаков заболевания при построении базы знаний и необходимости установления взаимосвязи между количеством баллов, отводимых на оценку каждого признака, с их приоритетностью.

Ключевые слова: *атопический дерматит, Байесовская оценка, база данных, экспертная система, нечеткое множество, подвижной состав, интенсивность движения.*

Введение

Процесс прогнозирования медицинской проблемы (например, atopического дерматита) является актуальной, ему посвящено немалое количество работ связанных как с диагностикой, так и с моделированием процесса заболевания [1]. Одна из сложностей построения эффективных диагностических (экспертных) систем состоит в необходимости учета более 80 данных пациента, необходимости принимать решения о стадии заболевания как на основании объективных, так и субъективных данных. Нет однозначного подхода к построению диагностической оценочной экспертной системы для прогнозирования дерматита (медицина) и оптимального времени занятости перегона (железнодорожный транспорт).

Экспертные системы, как известно, состоят из базы данных, базы знаний и решающего модуля. Последний, используя исходные данные из рабочей памяти и знания из базы знаний, формирует такую последовательность правил, которые, будучи примененными к исходным данным, приводят к решению задачи. Выбор и создание последовательности правил и составляет одну из главных проблем построения медицинских экспертных систем. Эта проблема непосредственно связана с созданием и эффективным использованием баз данных, из которых берутся исходные данные для выбора и запуска того или иного правила. Правила представляют собой один или несколько алгоритмов, моделирующих процесс диагностики заболевания.

Одним из способов построения экспертных систем является использование Байесовских сетей доверия в сравнении с использованием теории нечеткой логики [2 - 4]. Здесь возникает

необходимость эффективного поиска информации в базах данных, что в немалой степени используется в алгоритмах, моделирующих процессы диагностики и прогнозирования заболеваний [3 - 5].

Оценим какой метод (сети Байеса или нечетких множеств по Мамдани) предъявляет менее жесткие требования как к наполняемости баз данных, так и к их поиску без потери эффективности. Применение реляционных баз данных позволяет задействовать язык SQL-запросов, который позволяет практически мгновенно выводить данные из таблиц с количеством символов более 10 000 и взаимодействовать между таблицами. Системы, подобные PostgreSQL, позволяют еще и проводить вычисления в таблицах, используя встроенный язык программирования. Однако подобные средства разработки (Oracle, PostgreSQL и т.д.) созданы исключительно для работы с базами данных и непригодны для построения экспертных систем. Создание реальных экспертных систем, несмотря на имеющийся прогресс в области проектирования баз данных, по-прежнему сопряжен с трудностями, связанными как с визуализацией полученных результатов (придания наглядности), так и с необходимостью привлечения квалифицированных разработчиков и экспертов. Коммерческие средства разработки экспертных систем недостаточно адаптированы под конкретные отечественные условия и несколько дорогостоящи.

Следовательно, разработки конкретной медицинской экспертной системы под конкретное заболевание необходимо, с точки зрения программирования, начинать с методов моделирования течения заболевания.

© М.А. Омаров, И.И. Цехмистро, А.А. Трубицин, 2015

В оценке состояния больного, страдающего atopическим дерматитом (АД) и для принятия решений, связанных с выбором дальнейшей тактики лечения, важную информацию для лечащего врача дает анализ субъективных факторов заболевания. На железнодорожном транспорте признаки построения оптимальной системы безопасности и эффективности перевозок также могут быть противоречивыми.

На сегодняшний день для характеристики тяжести АД наиболее часто используется система SCORAD (scoring of atopic dermatitis). Индекс SCORAD представляет собой результат суммирования баллов, полученных при субъективной оценке выраженности и распространенности проявления заболевания.

Сложность объективной оценки перечисленных факторов является причиной использования методов, ориентированных на построение моделей, учитывающих неточность исходных данных. Именно в таких случаях применение математического аппарата нечетких множеств (НМ) является наиболее эффективным.

Цель работы

Целью исследований является выработка рекомендаций для разработки математической модели поддержки принятия решений, связанных с диагностикой больного и принятия решений относительно дальнейшей стратегии лечения. Работа также затрагивает вопрос о выборе рекомендаций по выработке эффективных алгоритмов, которые обеспечивают безопасность движения поездов (предотвращения заезда на занятую колею).

Выработка практических рекомендаций для построения экспертной системы оценки состояния больного (и оценки алгоритма пропуска поездов) на основе базы знаний и алгоритмов теории нечетких множеств путем анализа базы знаний и Байесовского оценивания.

Описание модели

Воспользуемся данными пациентов, полученные на базе кафедры пропедевтики № 2 ХНМУ.

В качестве лингвистических переменных воспользуемся такими качественными показателями АД как: оценка площади поражения кожи; оценка морфологических элементов сыпи; оценка субъективных симптомов.

Лингвистическая переменная «оценка площади поражения» была представлена как

$\langle b1, S, X1, G1, M1 \rangle$,

где $b1$ – площадь поражения (в %);

$S = \{ "S_{\min}$ – малая область поражения", " $S_{\text{сред}}$ – средняя область поражения", " $S_{\text{знач}}$ – значительная область поражения" };

$X1 = [0, 100]$ – количество термов (элементов множества);

$G1$ – процедура образования новых термов с помощью логических связок “И”, “ИЛИ” и модификаторов типа “очень”, “НЕ”, “слегка” и др.;

$M1$ – процедура семантические правила, задающие функции принадлежности нечетких термов, порожденных синтаксическими правилами $G1$.

Лингвистическая переменная «оценка морфологических проявлений сыпи» была представлена как

$\langle b2, K, X2, G2, M2 \rangle$,

где $b2$ – оценка морфологических проявления сыпи;

$K = \{ "0$ – отсутствует", " $K_{\text{лег}}$ – легкая", " $K_{\text{сред}}$ – средняя", " $K_{\text{тяж}}$ – тяжелая" };

$X2 = [0, 18]$ – количество термов (элементов множества);

$G2$ – процедура образования новых термов с помощью логических связок “И”, “ИЛИ” и модификаторов типа “очень”, “НЕ”, “слегка” и др.;

$M2$ – процедура семантические правила, задающие функции принадлежности нечетких термов, порожденных синтаксическими правилами $G2$.

Лингвистическая переменная «оценка субъективных симптомов», оценивающая зуд и нарушение сна в баллах по 10 бальной шкале для каждого из признаков, представлена как

$\langle b3, C, X3, G3, M3 \rangle$,

где $b3$ – оценка субъективных симптомов;

$T3 = \{ "C_{\text{лег}}$ – незначительное беспокойство", " $C_{\text{умер}}$ – умеренное беспокойство", " $C_{\text{сред}}$ – среднее беспокойство", " $C_{\text{знач}}$ – значительно беспокойство", " $C_{\text{умер}}$ – сильное беспокойство", " $C_{\text{очензнач}}$ – очень сильное беспокойство" };

$X3 = [0, 20]$ – количество термов (элементов множества);

$G3$ – процедура образования новых термов с помощью логических связок “И”, “ИЛИ” и модификаторов типа “очень”, “НЕ”, “слегка” и др.;

$M3$ – процедура семантические правила, задающие функции принадлежности нечетких термов, порожденных синтаксическими правилами $G3$.

Лингвистическая переменная «степень тяжести состояния больного» была представлена как

$\langle b4, T4, X4, G4, M4 \rangle$,

где $b4$ – степень тяжести состояния больного;

$T4 = \{ "легкая$ степень тяжести", "средняя степень тяжести", "тяжелое состояние", "очень тяжелое состояние" };

X4=[0, 10];

G4 – процедура образования новых термов с помощью логических связок “И”, “ИЛИ” и модификаторов типа “очень”, “НЕ”, “слегка” и др.;

M4 – процедура аналогичная M3 для правил M4.

Нечеткая база знаний, представляющая совокупности лингвистических высказываний имеет вид

$$\mu(b_1, b_2, b_3, S, K, C) = \left\{ \begin{array}{l} d_{\text{лег}} \text{ if } \left((0 < b_1 \leq S_{\text{min}}) \text{ and } (0 < b_2 \leq K_{\text{лег}}) \text{ and } (0 < b_3 \leq C_{\text{лег}}) \right) \\ \text{or } \left((S_{\text{min}} < b_1 \leq S_{\text{сред}}) \text{ and } (0 < b_2 \leq K_{\text{лег}}) \text{ and } (C_{\text{лег}} < b_3 \leq C_{\text{умер}}) \right) \\ d_{\text{сред}} \text{ if } \left((S_{\text{min}} < b_1 \leq S_{\text{сред}}) \text{ and } (K_{\text{лег}} < b_2 \leq K_{\text{сред}}) \text{ and } (C_{\text{умер}} < b_3 \leq C_{\text{сред}}) \right) \\ d_{\text{тяжел}} \text{ if } \left((0 < b_1 \leq S_{\text{min}}) \text{ and } (K_{\text{лег}} < b_2 \leq K_{\text{сред}}) \text{ and } (C_{\text{сред}} < b_3 \leq C_{\text{знач}}) \right) \\ d_{\text{оченьтяжел}} \text{ if } \left((S_{\text{сред}} < b_1 \leq S_{\text{знач}}) \text{ and } (K_{\text{лег}} < b_2 \leq K_{\text{сред}}) \right) \\ \text{and } (C_{\text{знач}} < b_3 \leq C_{\text{оченьзнач}}) \\ \text{or } \left((S_{\text{сред}} < b_1 \leq S_{\text{знач}}) \text{ and } (K_{\text{сред}} < b_2 \leq K_{\text{тяжел}}) \text{ and } (C_{\text{умер}} < b_3 \leq C_{\text{знач}}) \right) \\ \text{or } \left((S_{\text{сред}} < b_1 \leq S_{\text{знач}}) \text{ and } (K_{\text{лег}} < b_2 \leq K_{\text{сред}}) \text{ and } (C_{\text{знач}} < b_3 \leq C_{\text{оченьзнач}}) \right) \end{array} \right. \quad (1)$$

В данном соотношении присутствует пять переменных, не считая констант на которые раскладываются переменные S, K, C, что затрудняет построение функции принадлежности в явном виде.

Аналитические результаты

Аналитические зависимости между термом d_i и переменными b_1, b_2, b_3 устанавливается в случае объединения множеств S, K, C, либо рассмотрения по отдельности каждого множества. Например, введя в рассмотрение $a=0, b=S_{\text{min}}, c=S_{\text{сред}}, b=S_{\text{знач}}$ получим трапецидальную функцию принадлежности

$$\mu(b_1) = \left\{ \begin{array}{l} d1 = \left(1 - \frac{S_{\text{min}} - b_1}{S_{\text{min}}} \right), \left(0 < b_1 \leq S_{\text{min}} \right) \\ d2 = 1, \left(S_{\text{min}} < b_1 \leq S_{\text{сред}} \right) \\ d3 = \left(1 - \frac{b_1 - S_{\text{сред}}}{S_{\text{тяжел}} - b_1} \right), \left(S_{\text{сред}} < b_1 \leq S_{\text{тяжел}} \right) \\ d4 = \text{"другие варианты"} \end{array} \right. \quad (2)$$

Аналогичные треугольные или трапецидальные принадлежности можно представить как множества K и C, если рассматривать эти факторы по отдельности, однако это не будет описывать реальную базу знаний для диагностики заболевания, к тому же этот вариант усложнит или сделает невозможным процесс автоматизации (возможность построения экспертной системы).

Однако выражение (1) можно упростить, используя функции принадлежности, введенные в рассмотрение Болдвином [5 - 9].

Применяя квантификаторы, получим функции принадлежности термов применительно к соотношению (1) в предположении наличия одного нечеткого множества (признака заболевания), например проявления сыпи:

$$\begin{aligned} -\mu_{\text{сред}}(X2) &= d_{\text{сред}}(b2) = d_{\text{лег}}(b2)^2; \\ -\mu_{\text{тяж}}(X2) &= d_{\text{тяж}}(b2) = 1 - d_{\text{лег}}(b2). \end{aligned} \quad (3)$$

Данные соотношения представлены для нечеткого множества К, для нечеткого множества S с константами (S_{мин}, S_{сред}, S_{знач}). Функции принадлежности имеют вид:

$$\begin{aligned} -\mu_{\text{сред}}(X1) &= d_{\text{сред}}(b1) = d_{\text{мин}}(b1)^2; \\ -\mu_{\text{тяж}}(X1) &= d_{\text{тяж}}(b1) = 1 - d_{\text{мин}}(b1). \end{aligned} \quad (4)$$

Для нечеткого множества С определяем следующие функции принадлежности, принимая во внимание большее число констант, характеризующих границы степеней заболевания:

$$\begin{aligned} -\mu_{\text{умер}}(X3) &= d_{\text{умер}}(b3) = d_{\text{лег}}(b3)^{1/2}; \\ -\mu_{\text{сред}}(X3) &= d_{\text{сред}}(b3) = d_{\text{знач}}(b3)^{1/2}; \\ -\mu_{\text{знач}}(X3) &= d_{\text{знач}}(b3) = 1 - d_{\text{лег}}(b3); \\ -\mu_{\text{оченьзнач}}(X3) &= d_{\text{оченьзнач}}(b3) = d_{\text{знач}}(b3)^2. \end{aligned} \quad (5)$$

Таким образом, мы имеем дело с тремя нечеткими множествами S, K, C с разным значением конечных элементов (термов) b1=X1={0...100}, b2=X2{0...18}, b3=X3{0..20} в каждом множестве и с разным количеством функций принадлежности, которые обуславливаются медицинскими потребностями в рассматриваемой ситуации.

Для эффективной оценки состояния больного мы не можем уменьшить количество функций принадлежности и безосновательно пренебрегать количеством функций принадлежности данных на рассмотрение нечетких множеств.

Рассмотрим нечеткие отношения, следующие из модели базы знаний (1) на основании импликации и t-нормам, предварительно установив, что множества S, K, C являются входными, а множество D – выходным. Отношение с нечеткими множествами строятся на определенных правилах, наиболее важными из которых являются объединение и пересечения (правило ассоциативности для трех множеств):

$$\begin{aligned} -\mu_{S \cup K \cup C} &= \max(\mu_S(x), \max(\mu_K(x), \mu_C(x))), \\ -\mu_{S \cap K \cap C} &= \max(\mu_S(x), \max(\mu_K(x), \mu_C(x))). \end{aligned} \quad (6)$$

Введем в рассмотрение множество KC путем объединения множеств K и C, для которого:

$$\begin{aligned} -\mu_{K \cup C} &= \max(\mu_K(x), \mu_C(x)), \\ -\mu_{K \cap C} &= \max(\mu_K(x), \mu_C(x)), \end{aligned}$$

и выражения для отрицаний и импликаций:

$$\begin{aligned} \mu_{\overline{KC}}(X) &= 1 - (\mu_{KC}(x)), \\ \mu_{S=KC}(X) &= \max(1 - \mu_S(x), \mu_{KC}(x)). \end{aligned}$$

Введенные выше операции над нечеткими множествами основаны на использовании операций max и min, что является удобным для их реализации в программных пакетах научного назначения Mathcad и Matlab. Эти соотношения реализуемы и в средах разработки, пригодных для создания экспертных систем и баз данных (они есть в языке SQL). В теории нечетких множеств разрабатываются вопросы построения обобщенных, параметризованных операторов пересечения, объединения и дополнения, позволяющих учесть разнообразные смысловые оттенки соответствующих им лингвистических связей естественного языка «и», «или», «не» [7]. Один из подходов к операторам пересечения и объединения заключается в их определении, при помощи нечетких операторов, т.н. треугольных норм. Следует обратить внимание на то, что представленные выше операции пересечения min(μ_A(x), μ_B(x)) и объединения max(μ_A(x), μ_B(x)), использующиеся как самостоятельно, так и при введении операций разности, симметрической разности и дизъюнктивной суммы – это только один из возможных вариантов определения данных операций, введенный основоположником теории нечетких множеств Л.Заде.

Треугольной нормой (t-нормой) называется двуместная действительная функция T, отображающая две функции принадлежности нормальных нечетких множеств μ_A(x), μ_B(x) в одну функцию принадлежности нормального нечеткого множества и удовлетворяющая следующим условиям. Рассмотрим вначале дизъюнктивную сумму множеств S и K, отметим что данные множества являются не доминируемыми:

$$S \otimes K = (S - K) \cup (K - S) = (S \cap \bar{K}) \cup (\bar{K} \cap S), \quad (7)$$

с функцией принадлежности

$$\mu_{S-K}(X) = \max\{\{\min(\mu_S(x), 1 - \mu_K(x))\}, \{\min(1 - \mu_S(x), \mu_K(x))\}\}$$

$$\mu_{A-B}(X) = \max\{\{\min(\mu_A(x), 1 - \mu_B(x))\}, \{\min(1 - \mu_A(x), \mu_B(x))\}\}.$$

Рассмотренные соотношения хоть и не дают конкретной информации практического характера, но могут быть полезны для построения экспертных систем для отдельных ситуаций на транспорте и медицине.

Предшествующие рассуждения, несмотря на относительную несложность базы знаний (1), не привели к конкретной наглядной концепции получения методов оценок. Перейдем к численным методам, введя в рассмотрение Байесовское оценивание.

На основании результатов медицинского обследования пациента по трем признакам - оценка площади поражения кожи; оценка морфологических элементов сыпи; оценка субъективных симптомов - рассмотрим таблицу 1 диагностики 10 (десяти) пациентов, обследованных на атопический дерматит.

Представим их результаты в упрощенном виде, при котором количество значащих элементов одинаково, т.е. соответствует 10, однако, если одно и то же значение множества встречается более одного раза, количество переменных уменьшается.

Таблица 1

Данные 10 пациентов, больных дерматитом

Номер больного	SCORAD	Площадь поражения кожи (баллы) - S(100 макс.)	Субъективные признаки (зуд) - C(10 макс.)	Сумма баллов объективных признаков - К (эритема, отек, сухость) (18 макс.)
1	38	5	5	8
2	61	95	5	9
3	28	10	2	6
4	21	20	0	5
5	20	4	2	5
6	60	12	4	15
7	24	20	3	5
8	15	5	0	4
9	7	0	0	2
10	20	2	5	8

Например, в множестве К значение 5 встречается у 4-ого, 5 и 7 пациента, таким образом увеличивается их степень принадлежности значению 5 (составляет 0,3).

Таблица 2

Степень присутствия признака S в группе больных

Степень принадлежности элементов множества S	0,2	0,1	0,1	0,2	0,1	0,1	0,1
Значение элементов множества S	5	95	10	20	4	12	2

Таблица 3

Степень присутствия признака K в группе больных

Степень принадлежности элементов множества K	0,2	0,1	0,1	0,3	0,1	0,1	0,1
Значение элементов множества K	8	9	6	5	4	15	2

Таблица 4

Степень присутствия признака C в группе больных

Степень принадлежности элементов множества C	0,3	0,2	0,3	0,1	0,1
Значение элементов множества C	5	2	0	4	3

Рассмотрим матрицу оценок заболевания девяти пациентов конкретным врачом.

Матрица оценок из таблицы 1 имеет вид (8):

$$M = \begin{pmatrix} 5 & 5 & 8 \\ 95 & 5 & 9 \\ 10 & 2 & 6 \\ 20 & 0 & 5 \\ 4 & 2 & 5 \\ 12 & 4 & 15 \\ 20 & 3 & 5 \\ 5 & 0 & 4 \\ 0 & 0 & 2 \end{pmatrix} \quad (8)$$

Целесообразно представить все три качества в системе с одинаковым максимальным количеством баллов для каждой группы. Однако по количеству баллов, которыми пользуются медики для каждого фактора можно судить о степени важности каждого фактора. На рис. 1 представлена зависимость $Cr_i = f(i)$ – усредненное графическое отображение оценок больных врачом по каждому фактору, соотношение (9):

$$Cr_i = \frac{1}{m} \sum_{j=0}^{m-1} M_{j,i} \quad (9)$$

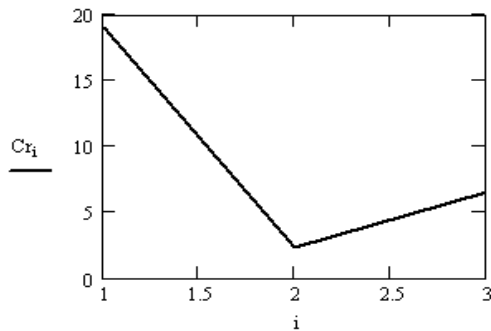


Рис. 1. Усредненное графическое отображение оценок больных по каждому фактору

На рис. 2 представлена зависимость $E_{\text{карт}_j} = f(j)$ усредненное графическое отображение общей оценки состояния конкретного больного по соотношению (10):

$$E_{\text{карт}_j} = \frac{1}{n} \sum_{j=0}^{n-1} (M^T)_j \quad (10)$$

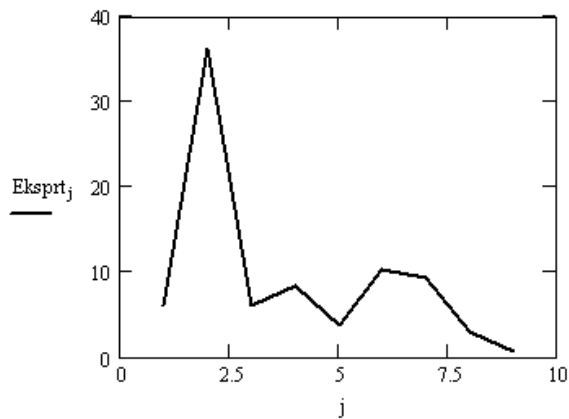


Рис. 2. Общая оценка состояния конкретного (j-) больного

Указанные оценки имеют как количественную, так и качественную составляющую.

Из сопоставления зависимостей вытекает вывод о том, что разброс абсолютных значений по первому графику составил 10 раз, по второму - около 20 раз.

Расчет весовых коэффициентов по результатам обследования больных проводится по соотношению:

$$\alpha_i = \frac{1}{n} \sum_{j=1}^n \left(M_{j,i} / \sum_{i=1}^m M_{j,i} \right) \quad (11)$$

$$\alpha = \begin{pmatrix} 0.503 \\ 0.095 \\ 0.402 \end{pmatrix}$$

$$\alpha^T = (0.503 \quad 0.095 \quad 0.402)$$

Расчет средневзвешенной оценки результатов обследования больных в баллах

$$K_m = \frac{1}{n} \cdot \sum_{j=1}^n \sum_{i=1}^m \left(\alpha_i \cdot M_{j,i} \right) = 12,413. \quad (12)$$

Весовые коэффициенты I (строка) фактора заболевания (переменной) в общей оценке заболевания каждого пациента j (столбец):

$$\alpha_{i,j} = \frac{M_{j,i}}{\sum_{i=1}^m M_{j,i}} \quad (13)$$

$$\alpha = \begin{pmatrix} 0,276 & 0,872 & 0,556 & 0,8 & 0,364 & 0,387 & 0,714 & 0,556 & 0 \\ 0,278 & 0,046 & 0,111 & 0 & 0,182 & 0,129 & 0,107 & 0 & 0 \\ 0,444 & 0,083 & 0,333 & 0,2 & 0,455 & 0,484 & 0,179 & 0,444 & 1 \end{pmatrix}$$

Оценка результирующего вывода о заболевании пациентов при равной значимости весовых коэффициентов

$$K1_m = \frac{1}{n} \cdot \sum_{j=1}^n \sum_{i=1}^m \left(\frac{M_{j,i}}{n} \right) = 3,099. \quad (14)$$

Численные значения весовых коэффициентов для трех признаков.

В числителях находятся коэффициенты, величина которых зависит от приоритетности признаков. Признак S=0,95 для (α1), K=0,8 для (α2), C=0,5 для (α3). Поскольку приоритеты уже учтены в баллах, отводимых на каждый признак:

$$\alpha1 = \frac{1,15}{n}, \alpha2 = \frac{0,95}{n}, \alpha3 = \frac{0,95}{n}.$$

Оценка степени заболевания пациента при разном значении весовых коэффициентов, учитывающих приоритеты признаков, представлена как

$$K1_m = \frac{1}{n} \cdot \sum_{j=1}^n \sum_{i=1}^m \left(\frac{M_{j,i}}{n} \cdot (\alpha1_i) \right) \quad (15)$$

$$K1=9,462.$$

Данное оценивание нельзя назвать абсолютно четким, оно показало, что средневзвешенная оценка заболевания пациентов врачом составило 12,5 усл. баллов. При предположении равнозначности весовых коэффициентов оценка заболевания больных пациентом падает до 3,1 усл. баллов. При учете приоритета признаков оценка заболевания составляет 9,5 усл. баллов. Данный метод экспертного оценивания основывается на гипотезе равной значимости весовых коэффициентов, значения которых четко зависят от максимального количества баллов, которые определяются медиками по определению. К тому же его нельзя назвать методом четкого оценивания.

Для установления влияния приоритетности, ввиду важности признаков, проведем дополнительное Байесовское оценивание.

Рассмотрим определения важности одного из трех приведенных диагностических признаков. Разделим приведенные диагностические признаки в зависимости от их важности (табл. 5). Важность определим следующим образом: средняя важность – К, существенная важность – S, малосущественный признак – С. Таким образом, рассматривая конкретный признак, существует три варианта. Признак попадает в одну из трех групп. Примем, что 50 % признаков имеет среднюю важность, 20 % признаков имеют существенную важность и 30 % признаков не существенно важны. Априорные вероятности в данном случае $P(B_1)=0,5$, $P(B_2)=0,3$, $P(B_3)=0,2$. Одними из критериев важности того или иного признака является использование его врачами и указывание на него пациентами. Существенный признак используют 80 % процентов врачей, 50 % процентов врачей используют признак средней важности, и 20 % процентов врачей используют также малосущественные признаки. Укажем также на то, что на существенный признак отмечается 80 % больных $P(A_2/B_1)=0,8$, у которых данный признак явно присутствует, на несущественный указывает 40 % больных (среди тех у которых он явно указывает на дерматит), $P(A_2/B_3)=0,8$, на признак средней важности указывает 20 % больных, $P(A_2/B_1)=0,2$ (однако по обследованиям он полностью подтверждает дерматит).

Вероятность того, что признак S считают для себя приоритетным, указывает 80 % врачей: $P(A_1/B_1)=0,8$. Признак “К” приоритетным для себя считают 50 % врачей. И 20 % врачей считают необходимым учитывать признак С.

$$P\left(\frac{B_i}{A_i}\right) = \frac{P\left(\frac{A_i}{B_i}\right) \cdot P(B_i)}{\sum_{K=1}^3 P\left(\frac{A_i}{B_i}\right) \cdot P(B_i)}, \quad (16)$$

$$P\left(\frac{B_1}{A_1}\right) = \frac{0,8 \cdot 0,5}{0,8 \cdot 0,5 + 0,5 \cdot 0,2 + 0,25 \cdot 0,2} = \frac{0,4}{0,6} = 0,66,$$

$$P\left(\frac{B_2}{A_2}\right) = \frac{0,5 \cdot 0,3}{0,8 \cdot 0,5 + 0,5 \cdot 0,3 + 0,25 \cdot 0,2} = \frac{0,15}{0,6} = 0,25,$$

$$P\left(\frac{B_3}{A_3}\right) = \frac{0,25 \cdot 0,2}{0,8 \cdot 0,5 + 0,5 \cdot 0,3 + 0,25 \cdot 0,2} = \frac{0,05}{0,6} = 0,083.$$

Таблица 5

Распределение вероятностей по признакам важности

Вероятность	1 высокая важность признака	2 высокая важность признака	3 высокая важность признака
$P(B_i)$	0,50	0,30	0,20
$P(A_1/B_i)$	0,80	0,50	0,25
$P(A_2/B_i)$	0,80	0,20	0,40

После того, как жалобы больных на возможное присутствие деоматита полностью подтвердились, доверия к предположениям B_2 и B_3 потеряло важность, т.е. признаки средней и меньшей значимости подтверждаются как мнением врачей, так и жалобами пациентов с меньшей вероятностью. Событие A_1 – использование врачом признака как главного, A_2 – жалоба пациента на признак заболевания. B_1 – диагностическая признак имеет существенную важность, B_2 – диагностический признак имеет среднюю важность, B_3 – признак имеет малую важность.

Если независимость событий A_2 и A_3 объективна, то при верности гипотез B_1, B_2, B_3 проведем оценки по формуле Байеса в следующем виде [8 - 10]:

$$P\left(\frac{B_i}{A_1 A_2}\right) = \frac{P\left(\frac{A_1}{B_i}\right) \cdot P\left(\frac{A_2}{B_i}\right) \cdot P(B_i)}{\sum_{K=1}^3 P\left(\frac{A_1}{B_i}\right) \cdot P\left(\frac{A_2}{B_i}\right) \cdot P(B_i)}, \quad (17)$$

$$P\left(\frac{B_1}{A_1 A_2}\right) = \frac{0,8 \cdot 0,5 \cdot 0,8}{0,8 \cdot 0,8 \cdot 0,5 + 0,5 \cdot 0,3 \cdot 0,2 + 0,25 \cdot 0,2 \cdot 0,4} = 0,86,$$

$$P\left(\frac{B_2}{A_1 A_2}\right) = \frac{0,3 \cdot 0,5 \cdot 0,2}{0,8 \cdot 0,8 \cdot 0,5 + 0,5 \cdot 0,3 \cdot 0,2 + 0,25 \cdot 0,2 \cdot 0,4} = 0,08,$$

$$P\left(\frac{B_3}{A_1 A_2}\right) = \frac{0,2 \cdot 0,25 \cdot 0,4}{0,8 \cdot 0,8 \cdot 0,5 + 0,5 \cdot 0,3 \cdot 0,2 + 0,25 \cdot 0,2 \cdot 0,4} = 0,05.$$

Данные оценки подтверждают оценки по предыдущему соотношению с использованием экспертного оценивания.

Численный анализ оценивания вероятности столкновения подвижных составов.

В рекомендациях по построению системы управления безопасности перевозок большое внимание уделяется использованию результатов аудитов, инспекторских проверок. Важным элементом системы является информация о недостатках в области обеспечения безопасности, полученная от сотрудников.

Возникает необходимость разработки дополнительного механизма учета этой важной информации. Предлагается разработать его на основе широко применяемой для решения подобных задач формулы Байеса. Будем рассматривать вероятность фактора опасности, полученную с помощью одной из процедур, кратко описанных выше, как априорную оценку. Полученная дополнительная информация используется для ее уточнения и получения апостериорной оценки случайного события и его устойчивости [11 - 12].

Рассмотрим решение задачу на примере фактора опасности «Попадание товарного состава на занятую колею».

В основе метода – анализ статистики этих событий и экспертное оценивание железнодорожной станции (перегона) по двум показателям: интенсивность движения и качество управления железнодорожным движением по трехбалльной шкале. По интенсивности движения участки (станции, перегоны) - малой, средней и высокой интенсивности.

По качеству управления железнодорожным движением – на участки (перегоны) с хорошим, удовлетворительным и неудовлетворительным качеством управления (диспетчеризирования). Модель рассчитана на товарный состав 20 вагонов, его масса более 8000 т.

Таблица 6

Оценка вероятности попадания состава на занятую колею

Показатель интенсивности движения	Показатель качества управления движением состава		
	Хорошее	Удовлетв.	Неудовлетв.
Малая	0,000006	0,00009	0,00049
Средняя	0,00037	0,00584	0,025
Высокая	0,00158	0,0213	0,0777

Предположим, что для участка движения X в предстоящем сеансе (однократной) перевозки получена оценка вероятности P. Именно эта оценка и будет использована в дереве для прогнозирования вероятности события «Небезопасное касание». В то же

время, стало известно, что во время предыдущего рейса поступило сообщение от бригады машинистов, доставившего состав на ту же станцию, о возможном попадании состава на занятую колею. Очевидно, что оценка P должна быть скорректирована.

Предлагается подход к решению задачи с использованием формулы Байеса (16), в которой теперь A – гипотезы относительно условий, в которых может произойти событие Ai; P(B | Ai) – вероятность события B при условии, что наступит событие Ai (при справедливости гипотезы Ai).

Формула позволяет вычислить условные вероятности P(Ai | B) событий Ai (или вероятность справедливости гипотез Ai) на основании того факта, что событие B произошло (или не произошло). В рассматриваемой задаче событие B – это попадание на заданную колею. Условимся для исключения путаницы в дальнейшем оценку вероятности этого события называть «частотой».

Для упрощения дальнейших расчетов диапазон возможных (по таблице 1) вероятностей от 10⁻¹ до 10⁻⁶ разбивается на интервалы с дискретностью 10⁻¹.

Событие Ai – это попадание частоты события B в один из интервалов. Все Ai образуют полную группу событий.

Принимается «пессимистический» вариант оценки: каждое значение вероятностей в таблице 6 округляется до левой (большей) границы соответствующего интервала логарифмической шкалы (см. таблицу 7).

Таблица 7

«Пессимистический» вариант оценки вероятностей столкновения составов

	Качество управления движением		
	Хорошее	Удовлетв.	Неудовлетв.
Малая	10 ⁻⁵	10 ⁻⁴	10 ⁻³
Средняя	10 ⁻³	10 ⁻²	10 ⁻³
Высокая	10 ⁻²	10 ⁻¹	10 ⁻¹

Для получения распределения вероятностей для конкретной станции проводится экспертное оценивание с участием 10 экспертов, результаты сведены в таблице 8.

Итоги опроса рассматриваем как два нечетких множества соответствующих значений параметров и записываем в виде функций принадлежности A1 и A2. Значения элементов нечетких множеств – первые буквы соответствующих характеристик: A1=M/0,2+C/0,8+V/0, A2=X/0,5+Y/0,4+N/0,1.

Произведение A=A1A2 записываем в виде матрицы, представляющей собой набор значений функции принадлежности нечеткого множества «Интенсивность движения/Качество управления движением» для участка движения X (таблица 9).

Таблица 8

Экспертный опрос по станции (участку движения) X

	Интенсивность перевозок			Качество управления движением поезда			
	Малая	Средняя	Высокая		Хор.	Удовл.	Неуд.
Э-1	1			Э-1	1		
Э-2		1		Э-2	1		
Э-3	1			Э-3	1		
Э-4		1		Э-4	1		
Э-5		1		Э-5		1	
Э-6		1		Э-6			1
Э-7		1		Э-7	1		
Э-8		1		Э-8		1	
Э-9		1		Э-9		1	
Э-10		1		Э-10		1	
Итог	0,2	0,8	0	Итог	0,5	0,4	0,1

На основании таблиц 7 и 9 строим априорное распределение вероятностей гипотез. Метод построения поясним на примере.

По таблице 7 для интенсивности «Малая» и показателя качества управления движением «Хорошее» левая граница интервала частоты события В «попадание состава на занятую колею» составляет 10^{-5} .

Таблица 9

Функция принадлежности нечеткого множества

Показатель интенсивности движения	Показатель качества управления движением			
		Хорошее	Удовлетв.	Неудовлетв.
Малая	0,1	0,08	0,02	
Средняя	0,4	0,32	0,08	
Высокая	0	0	0	

По таблице 9 имеем: степень принадлежности элемента «Малая/Хорошее» для станции (участка движения) X равна 0,1. Принимаем, что степень принадлежности равна вероятности попадания частоты события В в интервал частоты (10^{-5} , 10^{-6}), который соответствует малой интенсивности и хорошему качеству управления движением в таблице 7.

В интервал (10^{-3} , 10^{-4}) попадают два элемента нечеткого множества: «Малая/Неудовлетв» и «Средняя/Хорошее». Соответственно, вероятность попадания события в данный интервал принимается равной сумме степеней принадлежности $0,4+0,02=0,42$.

Выводы

Созданная модель базы знаний позволяет проводить экспертную оценку степени заболевания (или его отсутствия) пациента atopическим дерматитом применительно к медицине и оценку попадания поезда на занятый путь (железнодорожный транспорт). Численный анализ функций принадлежности термов от их значений при разном

количестве термов показал реальные результаты, качественно сопоставимые с результатами медицинского обследования. Применение разного количества термов, т. е. разные размеры исходных множеств, в предложенном методе не существенно повлияли на усложнение алгоритмов численного анализа базы знаний благодаря использованию матричного представления введенных в рассмотрение нечетких множеств. Данный подход является менее наглядным, но практически более универсальным, нежели использование функций одной и более переменных для определения функций принадлежности, с помощью которых строится база знаний. К тому же он более функционален для построения экспертных систем, поскольку легче реализуется в средствах разработки, связанных с базами данных.

Проведенное экспертное оценивание с использованием средневзвешенных оценок и Байесовское оценивание четко показали необходимость учета приоритетности признаков заболевания при построении базы знаний и необходимости установления взаимосвязи между количеством баллов, отводимых на оценку каждого признака, с их приоритетностью. Приоритетность признаков заболевания также может варьироваться при разных стадиях заболевания.

Проведенные оценки степени безопасности движения железнодорожного транспорта показали необходимость учета факторов интенсивности движения, оценок безопасности движения на прогнозирование оптимальной загрузки станций (перегонов) на предмет использование свободных путей (путей), что важно для дальнейшего планирования места пропуска скоростных составов.

Литература

1. А.И. Бых. Метод анализа иерархии при построении мультифакторных моделей дифференциальной диагностики atopического дерматита /А.И. Бых, Е.В. Высоцкая, И.И. Ключник, А.А. Трубицин, А.П. Порван. // Восточно-Европейский журнал передовых технологий. с. 5-8.
2. В.А. Клименко. Информационная система диагностики и прогнозирования atopического дерматита у детей / В. А. Клименко, А. П. Порван, А. А.Трубицин. // Системы обработки информации. – Х.:ХВУ, 2004. – Вып. 8 (47). – С. 27-33.
3. Кораблев Н.М. Сравнительный анализ методов определения абсолютных приоритетов признаков при нечеткой исходной информации / Н.М. Кораблев, А.С. Непокупный, Алзин Ферас // Системы обработки информации. – Х.: ХВУ, 2005. – Вып. 9 (49). – С. 75-83.
4. Бих А.І. Розробка моделі бази даних інформаційної системи зберігання та обробки інформації про

- пацієнтів із захворюваннями різними дерматозами / Бих А.І., Панфьорова І.Ю., Висоцька О.В., Жуков В.І., Кириченко Ю.В. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2010. – № 31. – С. 23 – 29.
5. А.И. Поворознюк. Модель нечеткой экспертной системы прогноза риска развития профессионально обусловленных заболеваний / А.И. Поворознюк, Н.А. Чикина, И.В. Антонова // Системы обработки информации, 2010, вып. 5(86).- С.34-39.
 6. Baldwin J.F. Fuzzy logic and fuzzy reasoning. – London, Academic Press, 1981.
 7. Н.М. Кораблёв. Дифференциальная диагностика аллергодерматозов с использованием адаптивной модели нечеткого вывода / Н.М. Кораблёв, И.В. Сорокина, А.Э. Макогон // Системы обработки информации, 2009, вып. 3(77).
 8. Штовба С.Д. Проектирование нечетких систем средствами MATLAB / С.Д. Штовба. – М.: Горячая линия-Телеком, 2007. – 288 с.
 9. Е.В. Брежнев. Модель прогнозирования риска с использованием нечеткого вывода / Е.В. Брежнев, А.А. Адаменко // Системы обработки информации, 2009, вып. 2(76). – 23-29 с.
 10. Леоненков А. Нечеткое моделирование в среде MATLAB и fuzzyTECH / Леоненков А. – СПб.: БХВ-Петербург, 2005. – 736 с.
 11. R.I. Tsekhmistro. Comparative Analysis of Pseudorandom Number Generation in the Up-To-Date Wireless Data Communication / L.O. Kirichenko, O.Y. Krug, and A.W. Storozhenko/ Telecommunication and Radio Engeneering 70(4) (2011) p.325-333
 12. Мелихов А.Н. Ситуационные советующие системы с нечеткой логикой/ А.Н. Мелихов, Л.С. Бернштейн, С.Я. Коровин – М.: Наука, 1990.- 334 с.

М.А. Омаров, І.І. Цехмістро, А.А. Трубіцин. Пошук способів побудови інформаційних систем оцінювання складності ситуації в медицині і на транспорті на прикладі конкретних ситуацій. Проведені експертні оцінювання з використанням середньозважених оцінок і Байєсове оцінювання, вони чітко показали необхідність врахування пріоритетності ознак захворювання при побудові бази знань і необхідності встановлення взаємозв'язку між кількістю балів, які відводяться на оцінку кожної ознаки, з їх пріоритетністю.

Ключові слова: atopічний дерматит, Байєсова оцінка, база даних, експертна система, нечітка множина, рухомий склад, інтенсивність руху.

M. A. Omarov, I. I. Tsekhmistro, A. A. Trubitsin. The search for the means to construct information systems of estimating situation complexity in medicine and on transport having specific situations as examples. The approach to the creation of expert evaluations to determine the severity of atypical dermatitis by means of probabilistic approach based on Bayesian process in the comparison with the approach based on the theory of fuzzy sets is considered in the article. The created model of knowledge base allows conducting expert evaluation of patient's disease severity.

The conducted expert evaluation using average weighted estimations and Bayesian evaluation have clearly shown the necessity of the account of illness sign priority while creating knowledge base and the necessity to establish interconnection between the quantity of balls given to estimate each sign and their priority.

The proposed in the article method is more functional for the creation of expert systems as it is realized when data base is being developed.

The conducted evaluations of railway traffic safety level have shown the necessity to take into account factors of traffic intensity, estimations of traffic safety and their effect on the forecasting optimum work load of stations for the purpose of the usage of free tracks which is very important for further planning the passing of high-speed make-ups.

Key words: atypical dermatitis, Bayesian evaluation, database, expert systems, fuzzy set, rolling stock, traffic.

Рецензент д.т.н., професор Кириченко Л.О. (ХНУРЭ)

Поступила 17.09.2015г.

Омаров М.А., д.т.н., профессор, Харьковский национальный университет радиоэлектроники, декан факультета подготовки иностранных граждан.

Цехмистро И.И., к.фил.н. доцент кафедры высшей математики, Харьковский национальный политехнический университет.

Трубицин А.А., инженер кафедры проектирования и эксплуатации электронных аппаратов, Харьковский национальный университет радиоэлектроники.

Omarov M.A., Dr. of Tech. Sc., Professor, Pro-rector of international cooperation, Kharkov National University of Radio Electronics.

Tsekhmistro I.I., PhD. Department of Mathematics, Kharkov National Polytechnic University.

Trubitsin A.A., Engineer Department of design and operation of electronic devices, Kharkov National University of Radio Electronics.