УДК 004.89 Вестник СПбГУ. Прикладная математика. Информатика... 2019. Т. 15. Вып. 4
MSC 68T27

Постановка предварительного медицинского диагноза на основе теории нечетких множеств с использованием меры Сугено

А. Б. Гончарова

Санкт-Петербургский государственный университет, Российская Федерация,
199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Гончарова А. Б.* Постановка предварительного медицинского диагноза на основе теории нечетких множеств с использованием меры Сугено // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2019. Т. 15. Вып. 4. С. 529–543. https://doi.org/10.21638/11702/spbu10.2019.409

При постановке диагноза пациенту врач должен всесторонне и систематически изучить больного, т. е. собрать анамнез, объективно исследовать состояние больного, провести анализ лабораторных, рентгенографических и других видов обследований, а также учесть наличие других заболеваний у пациента. Обработка большого объема информации о больном и соотнесение его с симптоматикой различных заболеваний, не обязательно профильных специализаций конкретного врача, является сложной задачей, помочь в решении которой с учетом имеющейся информации может система поддержки постановки диагноза. В статье предлагается метод постановки диагноза по заданным симптомам на основе теории нечетких множеств. По заданному набору выраженных симптомов производится вычисление обобщенного показателя диагноза с использованием меры Сугено. Применяя поиск по максимуму обобщенных показателей, осуществляется вывод наиболее вероятного диагноза.

Ключевые слова: нечеткое множество, мера Сугено, система поддержки принятия решения, медицинская диагностика, многокритериальный выбор, обобщенный показатель.

Введение. В настоящее время оказание медицинских услуг пациентам происходит при непрерывном внедрении технологий в различные сферы медицинской диагностики, что приводит к увеличению объема информации, которое должен обрабатывать медицинский работник. В связи с этим появляется необходимость применения специальных программных систем для решения самых разных задач в медицинских учреждениях. К основным особенностям систем поддержки принятия решения относится то, что они программируются для решения определенной задачи: выявления наличия заболевания, прогнозирования течения болезни и др. [1–3]. В современной медицине существует множество лечебных и диагностических методик, таким образом при постановке диагноза учитывается большое количество факторов. При этом объем знаний непрерывно увеличивается, но время, отведенное врачу на принятие решения, остается неизменным, что ведет к росту врачебных ошибок, который, в свою очередь, приводит как к снижению эффективности системы здравоохранения, так и к большему уровню смертности. Применение систем поддержки принятия решения при постановке предварительного диагноза пациенту на основе заданных симптомов, лабораторных показателей, ультразвукового исследования и других методов дифференциальной диагностики, а также с учетом биометрических показателей конкретного пациента поможет поставить диагноз наискорейшим образом и назначить необходимое лечение.

Статья построена следующим образом: сначала описывается аппарат нечетких множеств, вводится понятие нечеткой меры, введенной на нечеткой базе знаний Сугено, затем строится модель многокритериального поиска предварительного медицинского диагноза с применением λ-нечеткой меры Сугено, разработан алгоритм реализации данной модели и в конце приводится пример принятия решения для постановки диагноза заболеваний глаз на основе нечеткой логики.

Понятие нечетких множеств впервые было введено в 1965 г. Л. Заде [4]. Приведем это определение. Пусть задано U — некоторое множество элементов u и отображение $\mu_A : U \to [0, 1]$. Нечетким подмножеством A в U называется график отображения μ_A, т. е. множество вида $\{(u, \mu_A(u)) : u \in U\}$, при этом $\mu_A(u)$ называется степенью принадлежности u к A. Таким образом, задание нечеткого подмножества A в U эквивалентно заданию его функции принадлежности $\mu_A(u)$.

Связь между нечетким подмножеством универсального множества U и семейством обычных его подмножеств задается при помощи понятия подмножества α-уровня нечеткого множества. Пусть $\alpha \in [0, 1]$, подмножеством α-уровня нечеткого множества A называется множество

$$A_\alpha = \{u \in U : \mu_A(u) \geqslant \alpha\}.$$

В дальнейшем для построения системы поддержки принятия решения понадобится применение следующей теоремы.

Теорема о декомпозиции. Любое нечеткое множество A можно представить так:

$$\Phi = \max_\alpha \alpha \times A_\alpha.$$

Пусть U — универсальное множество, C — класс из подмножеств U, $A, B \in C$. Функция $g(\cdot)$, определяемая в виде $g : C \to [0, 1]$, называется нечеткой мерой, если она удовлетворяет следующим условиям:

1) $\forall A \subseteq C \Longrightarrow g(A) \geqslant 0$; $g(\emptyset) = 0, g(C) = 1$;

2) если $A, B \in C, A \subseteq B$, то $g(A) \leqslant g(B)$ (монотонность);

3) если $A_i \in C$ и $\{A_i, i = 1, 2, \ldots\}$ является монотонной последовательностью $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_i \supseteq \ldots$, то $\lim_{n\to\infty} g(A_n) = g(\lim_{n\to\infty} A_n)$ (непрерывность).

Сугено [5] ввел понятия нечетких интегралов и нечеткой меры. Кроме того, он предложил λ-нечеткую меру как частный случай нечеткой меры. По сравнению с другими нечеткими мерами λ-нечеткая мера допускает более естественную параметризацию. Сугено определил функцию F-распределения λ-нечеткой меры для бесконечного случая и ее функцию плотности для конечного случая. Пусть $X = \{X_1, X_2, \ldots, X_n\}$ — конечное множество, g_λ — нечеткая мера, удовлетворяющая соотношению

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B),$$

здесь $A \cup B = \emptyset$, $A, B \subset X$ с параметром нормировки $-1 < \lambda < \infty$. Пусть $0 \leqslant g_i \leqslant 1$, $1 \leqslant i \leqslant n$, где $g_i = g(\{x_i\})$. При условии, что величины g_i, $i = 1, 2, \ldots, n$, заданы для любого подмножества $X' \subset X$, можно получить удовлетворяющую λ-правилу меру [6, 7]

$$g_\lambda(X') = \left[\prod_{x_i \in X'} (1 + \lambda g_i) - 1\right] / \lambda.$$

Параметр нормировки λ находится из соотношения [8]

$$\left[\prod_{x_i \in X'} (1+\lambda g_i) - 1\right] / \lambda = 1, \ -1 < \lambda < +\infty.$$

Свойства меры g_λ:

1) $-1 < \lambda < +\infty$.

Доказательство. Если

$$1 = g_\lambda(X) = g_\lambda(A \cup \bar{A}) = g_\lambda(A) + g_\lambda(\bar{A}) + \lambda g_\lambda(A) g_\lambda(\bar{A}),$$

то

$$g_\lambda(\bar{A}) = \frac{1 - g_\lambda(A)}{1 + \lambda g_\lambda(A)} \geqslant 0$$

для любого $A \subset X, 1 + \lambda g_\lambda(A) > 0$, тогда $\lambda > \sup_{A \subset X} \frac{-1}{g_\lambda(A)} = -1$;

2) если $X = \{X_1, X_2\}$, то

$$g_\lambda(\{X_1, X_2\}) = g_1 + g_2 + \lambda g_1 g_2 = \frac{1}{\lambda}\left((1+\lambda g_1)(1+\lambda g_2) - 1\right).$$

Рассмотрим функцию

$$f(\lambda) = \left[\prod_{i=1}^{n} (1+\lambda g_i) - 1\right] / \lambda - 1.$$

Ее свойства:

а) $\lim_{\lambda \to -1} f(\lambda) < 0$.

Доказательство:

$$\lim_{\lambda \to -1} f(\lambda) = -\prod_{i=1}^{n} (1 - g_i) < 0,$$

б) $\lim_{\lambda \to \infty} f(\lambda) > 0$.

Доказательство:

$$\lim_{\lambda \to \infty} = \lim_{\lambda \to \infty} \left(g_1 \prod_{i=2}^{n} (1+\lambda g_i) - 1 \right) = \infty > 0,$$

в) $f'(\lambda) > 0$, т. е. функция строго возрастающая.

Доказательство по индукции: $0 < g_i \leqslant 1$.

Пусть $n = 2$, тогда

$$f_2(\lambda) = \frac{1}{\lambda}\left((1+\lambda g_1)(1+\lambda g_2) - 1\right) - 1.$$

При

$$f_2'(\lambda) = \frac{1}{\lambda^2}\left(g_1(1+\lambda g_2) + g_2(1+\lambda g_1) - ((1+\lambda g_1)(1+\lambda g_2) - 1)\right) = g_1 g_2 > 0$$

условие выполняется.

Пусть оно выполняется при $n = k$, т. е.

$$f_k(\lambda) = \left[\prod_{i=1}^{k}(1+\lambda g_i) - 1\right]/\lambda - 1,$$

$$f'_k(\lambda) = \frac{1}{\lambda^2}\left(g_1(1+\lambda g_2)(1+\lambda g_3)\dots(1+\lambda g_k) + g_2(1+\lambda g_1)(1+\lambda g_3)\dots\right.$$
$$\dots(1+\lambda g_k) + g_k(1+\lambda g_1)(1+\lambda g_2)\dots(1+\lambda g_{k-1}) - ((1+\lambda g_1)(1+\lambda g_2)\dots$$
$$\left.\dots(1+\lambda g_k) - 1)\right) > 0.$$

Проверим выполнимость этого соотношения при $n = k+1$:

$$f_{k+1}(\lambda) = \left[\prod_{i=1}^{k+1}(1+\lambda g_i) - 1\right]/\lambda - 1 = \left[\prod_{i=1}^{k}(1+\lambda g_i) - 1\right]/\lambda - 1 + g_{k+1}\prod_{i=1}^{k}(1+\lambda g_i) =$$

$$= f_k(\lambda) + g_{k+1}\prod_{i=1}^{k}(1+\lambda g_i),$$

$$f'_{k+1}(\lambda) = f'_k(\lambda) + g_{k+1}\sum_{i=1}^{k} g_i \prod_{j=1, j\neq i}^{k}(1+\lambda g_i) > 0,$$

так как $0 < g_i \leqslant 1$, то $0 < g_{k+1}g_i \leqslant 1$ для любого i, при $-1 < \lambda < \infty$ $(1+\lambda g_i) \geqslant 0$ для любого i, таким образом,

$$g_{k+1}\sum_{i=1}^{k} g_i \prod_{j=1, j\neq i}^{k}(1+\lambda g_i) \geqslant 0,$$

что и требовалось доказать.

Функция $f(\lambda)$ строго возрастающая, $\lim_{\lambda\to -1} f(\lambda) < 0$ и $\lim_{\lambda\to\infty} f(\lambda) > 0$, значит, на интервале $-1 < \lambda < \infty$ у нее существует единственный корень.

Применение λ-нечеткой меры Сугено при поиске предварительного медицинского диагноза. Предположим, что общий показатель симптоматики разрабатываемых проектных решений выражается в виде иерархии n частных показателей симптоматики $S = \{S_i,\ i = 1, 2, \dots k\}$, которые оценивают варианты принятых решений из множества болезней $D = \{d\}$. На практике число диагностируемых болезней D конкретной системы поддержки принятия решений является конечным, так что их можно перечислить.

Частные показатели симптомов $S_i(X_1, X_2, \dots, X_m)$ зависят от различных половых, возрастных, клинических, биохимических и других параметров (X_j) принятия решения на разных этапах жизненного цикла. Эта взаимосвязь параметров и показателей симптоматики принятия решений может задаваться не только в аналитическом, но и в алгоритмическом виде путем построения разнообразных моделей функционирования системы [9, 10]. При этом сведения о параметрах носят неточный, неопределенный характер. Так, симптомы подразделяются на специфические, присущие одному заболеванию, и неспецифические, сопровождающие целый ряд болезней. Имеется значительная неопределенность в возможностях достижения характеристик системы тех или иных своих значений, т. е. интенсивность данного симптома. В связи с этим предполагается, что возможности параметров реализации заданы в виде нечетких множеств $X_j = (x_j, \mu_j(x_j))$, здесь $\mu_j(x_j)$ — возможности того, что параметры X_j могут принимать соответственно значения x_j.

Учитывая вышесказанное, считаем, что частные показатели симптоматики принятия решения могут быть представлены как нечеткие события $S_i : D \to \mathrm{Re}\,(Y_i)$, где для каждого варианта-диагноза $d \in D \quad S_i(d) \subseteq Y_i$ является нечетким множеством $S_i(d) = (s_i, \mu_{S_i(d)}(s_i)), s_i \in Y_i$, показатель S_i, $\mu_{S_i(d)}$ — функция принадлежности. При этом $\mu_{S_i(d)}$ интерпретируется как возможность того, что показатель симптоматики S_i примет значение s_i для диагноза $d \in D$.

Представление понятия диагноза d в виде нечеткого множества есть

$$d = \frac{\mu_{S_1(d)}}{S_1} + \frac{\mu_{S_2(d)}}{S_2} + \ldots + \frac{\mu_{S_k(d)}}{S_k},$$

здесь знак «+» не является обозначением операции сложения, а имеет смысл объединения.

В общем случае задача диагностирования заболевания относится к большим данным, которые можно представить в виде табл. 1 [3].

Таблица 1. **База данных принадлежности симптомов конкретному заболеванию**

S	d_1	d_2	$\ldots$	d_m
S_1	$\mu_{S_1(d_1)}$	$\mu_{S_1(d_2)}$	$\ldots$	$\mu_{S_1(d_m)}$
S_2	$\mu_{S_2(d_1)}$	$\mu_{S_2(d_2)}$	$\ldots$	$\mu_{S_2(d_m)}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
S_k	$\mu_{S_k(d_1)}$	$\mu_{S_k(d_2)}$	$\ldots$	$\mu_{S_k(d_m)}$

Таким образом, задача заключается в выработке обобщенного показателя симптоматики принятия решения (S), представляющего собой некоторую операцию над нечеткими событиями $S(d) = H\,(S_1(d), S_2(d), \ldots, S_k(d))$ для любого диагноза $d \in D$, объединяющего частные показатели симптоматики и учитывающего их влияние на оценку вариантов принятия.

Частные показатели симптоматики в задачах диагностики заболеваний, как правило, носят неравнозначный по важности характер. Распространенным методом выражения различий критериев по важности является назначение каждому из них некоторого веса с последующим суммированием этих весов в рамках операции свертки. Данный подход, как показывает анализ [10], приводит к потерям в эффективности применения принятия решения. Указанные потери вызваны тем, что коэффициенты в свертке частных показателей эффективности не учитывают нелинейный характер влияния показателей S_i друг на друга и в целом на обобщенный показатель симптоматики. Чтобы избежать указанные недостатки и учесть нечетковозможностное представление частных показателей симптоматики, в настоящей работе предлагается при построении обобщенного показателя симптоматики принятия решения нечетковозможностная свертка, основанная на нечеткой мере и нечетком интеграле [8].

Пусть g_i, где $i = 1, 2, \ldots, k$, $0 \leqslant g_i < 1$, — коэффициенты выраженности отдельно взятых частных показателей симптоматики при построении обобщенного показателя. Данная информация может быть получена от экспертов-врачей с применением методов экспертного оценивания [11].

Учет влияния совокупности различных показателей S_i, $i \in M$ $(M = \{1, 2, \ldots, n\})$, на оценку вариантов из множества D предлагается осуществлять путем построения конструктивной нечеткой меры Сугено на конечном множестве частных показателей $S_i, i \in M$, где g_i — плотность распределения этой нечеткой меры. Например,

g_i — степень выраженности симптома: 0.1 — незначительно выражен (незначительное увеличение числа лейкоцитов в крови), 0.5 — умеренно выражен (небольшая болезненность при пальпации), 0.9 — значительно выражен (сильная боль в области). В случае отсутствия информации g_i по данному симптому S_i у эксперта ставим 0.

В исходной табл. 1 добавляем столбец g (табл. 2).

Таблица 2. **База данных принадлежности симптомов конкретному заболеванию со степенью их выраженности у обследуемого пациента**

S	g	d_1	d_2	$\cdots$	d_m
S_1	g_1	$\mu_{S_1(d_1)}$	$\mu_{S_1(d_2)}$	$\cdots$	$\mu_{S_1(d_m)}$
S_2	g_2	$\mu_{S_2(d_1)}$	$\mu_{S_2(d_2)}$	$\cdots$	$\mu_{S_2(d_m)}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
S_k	g_k	$\mu_{S_k(d_1)}$	$\mu_{S_k(d_2)}$	$\cdots$	$\mu_{S_k(d_m)}$

Для исключения из анализа данных, о которых нет информации, вычеркиваем из табл. 1 те строки, где в столбце g стоят нули, тем самым уменьшаем количество данных в расчетах (табл. 3).

Таблица 3. **Уменьшение количества симптомов базы данных в зависимости от степеней выраженности симптома у изучаемого пациента**

S	g	d_1	d_2	$\cdots$	d_m
S_1	g_1	$\mu_{S_1(d_1)}$	$\mu_{S_1(d_2)}$	$\cdots$	$\mu_{S_1(d_m)}$
$\cancel{S}_2$	$\cancel{0}$	$\cancel{\mu}_{S_2(d_1)}$	$\cancel{\mu}_{S_2(d_2)}$	$\cancel{\cdots}$	$\cancel{\mu}_{S_2(d_m)}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cancel{S}_f$	$\cancel{0}$	$\cancel{\mu}_{S_f(d_1)}$	$\cancel{\mu}_{S_f(d_2)}$	$\cancel{\cdots}$	$\cancel{\mu}_{S_f(d_m)}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
S_k	g_k	$\mu_{S_k(d_1)}$	$\mu_{S_k(d_2)}$	$\cdots$	$\mu_{S_k(d_m)}$

Мера Сугено для рассматриваемого случая имеет следующий вид [6]:

$$G_\lambda(\{S_i, i \in M^1\}) = \left[\prod_{i \in m^1}(1 + \lambda g_i) - 1\right]/\lambda, \quad M^1 \subseteq M.$$

Перенумеруем симптомы и опять получим таблицу исходного вида с ненулевыми данными в столбце g (табл. 4).

Таблица 4. **База данных симптомов-болезней для обследуемого пациента**

S	g	d_1	d_2	$\cdots$	d_m
S_1	g_1	$\mu_{S_1(d_1)}$	$\mu_{S_1(d_2)}$	$\cdots$	$\mu_{S_1(d_m)}$
S_2	g_2	$\mu_{S_2(d_1)}$	$\mu_{S_2(d_2)}$	$\cdots$	$\mu_{S_2(d_m)}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
S_n	g_n	$\mu_{S_n(d_1)}$	$\mu_{S_n(d_2)}$	$\cdots$	$\mu_{S_n(d_m)}$

Следует указать, что значение λ находится из условия нормировки — нечеткой меры Сугено:

$$\left[\prod_{i=1}^{n}(1+\lambda g_i)-1\right]-\lambda=1, \quad -1<\lambda<+\infty.$$

Обобщенный показатель симптоматики принятия решения предлагается конструировать в виде нечеткой свертки, позволяющей гибко учитывать нелинейный характер влияния частных показателей. Для этого используем понятие нечеткого интеграла по λ-нечеткой мере Сугено [5, 8]:

$$e(d_i)=\int h^{\circ} G_\lambda=\sup_{\alpha\in[0;1]}\min\{\alpha, G_\lambda(S_\alpha(d_i))\},$$

где $S_\alpha(d_i)=\{S_j|h(S_j,d_i)\geqslant\alpha\}$ — множество показателей, степень влияния которых на оценку варианта $d_i\in D$ превышает порог α; $h: S\times D\to[0,1]$ — оценочная функция;

$$G_\lambda(S_\alpha(d_i))=\left[\prod_{S_i\in S_\alpha(d_i)}(1+\lambda g_i)-1\right]/\lambda.$$

В качестве оценочной функции h будем считать значения частных показателей симптоматики, приведенных к безразмерному виду с носителем нечеткого множества $S_i(d)$ $(i=1,2,\ldots,n, d\in D)$ в интервале $[0,1]$.

Алгоритм построения системы поддержки принятия решения о предварительном медицинском диагнозе. Он включает следующие этапы:

1) построение базы данных заболеваний-симптомов;

2) задание симптомов пациента в виде отклонений: g_i — степень выраженности симптома;

3) исключаем из анализа данные, о которых нет информации, т. е. выбираем из таблицы заболеваний-симптомов те строки, где в столбце g стоят ненулевые элементы, тем самым уменьшаем количество данных в расчетах; получаем новую базу заболеваний-симптомов с меньшим числом симптомов;

4) ищем значение $\tilde{\lambda}$ из условия нормировки λ-нечеткой меры Сугено:

$$\left[\prod_{i=1}^{n}(1+\lambda g_i)-1\right]/\lambda=1, \quad -1<\lambda<+\infty,$$

например по методу хорд;

5) вычисляем результирующий показатель симптоматики на основе нечеткой свертки частных показателей [8]:

$$e(d_i)=\sup_{\alpha\in[0;1]}\min\{\alpha, G_\lambda(S_\alpha(d_i))\},$$

где $S_\alpha(d_i)=\{S_j|S_j(d_i)\geqslant\alpha\}$, $d_i\in D$; $G_\lambda(S_\alpha(d_i))=[\prod_{S_i\in S_\alpha}(a+\tilde{\lambda}g_i)-1]/\tilde{\lambda}$;

6) находим вариант диагноза $d_{i0}\in D$, при котором обобщенный показатель имеет максимальное значение:

$$e(d_{i0})=\max_{d_i\in D}e(d_i).$$

Нечеткая модель принятия решения для постановки диагноза заболеваний глаз на основе нечеткой логики. Рассмотрим пример поиска класса заболевания глаз по возрастным характеристикам, симптомам заболевания и причинам возникновения заболевания. В работе Л. А. Коробовой и Т. В. Гладких [12] приведены статистические данные о заболевших по возрастным группам, о заболевших по первичным симптомам и о причинах возникновения заболеваний для пяти классов заболеваний зрительного анализатора человека. На основании этих данных строим базу данных диагнозов-симптомов (табл. 5).

Таблица 5. **База данных функций принадлежности симптомов заболеваний глаз по возрастным группам, по первичным симптомам и причинам возникновения**

Группы данных о заболеваниях глаз		Гордеолум и халазион, d_1	Воспаление век, d_2	Болезни век, d_3	Болезни слезного аппарата, d_4	Болезни глазницы, d_5
По возрастным группам	Возраст 0–1 года	0.0620	0.1117	0	0.0124	0.0099
	Возраст 1–3 года	0.0124	0.1117	0	0.0124	0.0261
	Возраст 3–14 лет	0.0124	0.0124	0	0.0620	0.0310
	Возраст 14–25 лет	0.0124	0.0124	0	0.0620	0.0323
	Возраст 25–40 лет	0.0124	0.0012	0	0.0620	0.0347
	Возраст 40–60 лет	0.0186	0	0.0223	0.0620	0.0385
	Возраст 60+ лет	0.0620	0.1117	0	0.0124	0.0099
По симптомам	Температура, лихорадка	0	0.5567	0	0	0
	Отек в области глаза	1	0.5808	0	0.1688	0.5000
	Снижение остроты зрения	0	0	0.7222	0.1688	0.2500
	Слезотечение	0.3478	0.5945	0.6944	0.8292	0
	Сухость в глазу	0	0.4055	0.3056	0.1688	0
	Косметический дефект	0.5304	0.4192	0.7778	0.8292	0.7500
По причинам возникновения	Инфекционное заболевание	1	0.7423	0.5833	0.6667	0.1972
	Травма глаза	0	0	0.4167	0	0
	Снижение остроты зрения	0	0	0.7222	0.1688	0.2500
	Врожденное заболевание	0	0	0	0	0.2042
	Заболевания других органов	0	0.2577	0	0.3333	0.5845

Как и в [12], обследуем пациента, обозначив его пациент 1: возраст — 5 лет;

первичный симптом — лихорадка, сухость в глазу; возможная причина — заболевания других органов.

Преобразуем исходную базу данных (табл. 6).

Таблица 6. **База данных симптомов-болезней для обследуемого пациента 1**

Данные о заболевании глаз	Выраженность симптомов данного пациента, g	Гордеолум и халазион, d_1	Воспаление век, d_2	Болезни век, d_3	Болезни слезного аппарата, d_4	Болезни глазницы, d_5
Возраст 3–14 лет (S_1)	0.9 — значительно выражен	0.0124	0.0124	0	0.0620	0.0310
Температура, лихорадка (S_2)	0.9 — значительно выражен	0	0.5567	0	0	0
Сухость в глазу (S_3)	0.9 — значительно выражен	0	0.4055	0.3056	0.1688	0
Заболевания других органов (S_4)	0.3 — незначительно выражен	0	0.2577	0	0.3333	0.5845

Находим $\tilde{\lambda}$ из условия

$$\frac{1}{\lambda}\left((1+\lambda 0.9)(1+\lambda 0.9)(1+\lambda 0.9)(1+\lambda 0.3)-1\right)=1.$$

Графическое решение данного уравнения представлено на рис. 1. Получаем, что $\tilde{\lambda} \approx -0.999$.

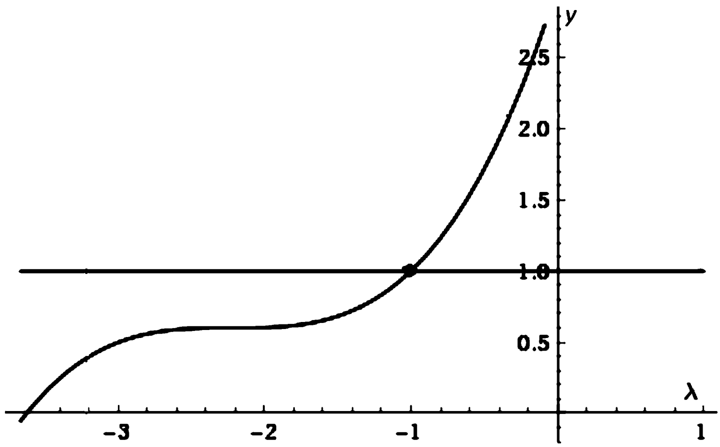

Рис. 1. Графическое решение уравнения поиска значения $\tilde{\lambda}$ из условия нормировки λ-нечеткой меры Сугено для пациента 1

Рассмотрим диагноз d_1 и разложим его по α-уровням:

• $a = 0$ нечеткого множества соответствует набор симптомов S_1, S_2, S_3, S_4, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

• $a = 0.0124$ нечеткого множества соответствует набор симптомов $\{S_1\}$, это

$$G_\lambda = \frac{(1 + 0{,}9 \cdot (-0.999)) - 1}{-0.999} = 0.9,$$

выбираем $\min\{0.0124; 0.9\} = 0.0124$.

Обобщенный показатель симптоматики болезни d_1 равен $e(d_1) = \max\{0; 0.0124\} = 0.0124$.

Рассмотрим диагноз d_2 и разложим его по α-уровням:

• $\alpha = 0.0124$ нечеткого множества соответствует набор симптомов $\{S_1, S_2, S_3, S_4\}$, это $G_\lambda = 1$, выбираем $\min\{0.0124; 1\} = 0.0124$;

• $\alpha = 0.2577$ нечеткого множества соответствует набор симптомов $\{S_2, S_3, S_4\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.999))(1 + 0.9 \cdot (-0.999))(1 + 0.3 \cdot (-0.999)) - 1}{-0.999} = 0.994,$$

выбираем $\min\{0.2577; 0.994\} = 0.2577$;

• $\alpha = 0.4055$ нечеткого множества соответствует набор симптомов $\{S_2, S_3\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.999))(1 + 0.9 \cdot (-0.999)) - 1}{-0.999} = 0.99,$$

выбираем $\min\{0.4055; 0.99\} = 0.4055$;

• $\alpha = 0.5567$ нечеткого множества соответствует набор симптомов $\{S_2\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.999)) - 1}{-0.999} = 0.9,$$

выбираем $\min\{0.5567; 0.9\} = 0.5567$.

Обобщенный показатель симптоматики болезни d_2 равен $e(d_2) = \max\{0.0124; 0.2577; 0.4055; 0.5567\} = 0.5567$.

Рассмотрим диагноз d_3 и разложим его по α-уровням:

• $\alpha = 0$ нечеткого множества соответствует набор симптомов $\{S_1, S_2, S_3, S_4\}$, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

• $\alpha = 0.3056$ нечеткого множества соответствует набор симптомов $\{S_3\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99)) - 1}{-0.99} = 0.9,$$

выбираем $\min\{0.3056; 0.9\} = 0.3056$.

Обобщенный показатель симптоматики болезни d_3 равен $e(d_3) = \max\{0; 0.3056\} = 0.3056$.

Рассмотрим диагноз d_4 и разложим его по α-уровням:

• $\alpha = 0$ нечеткого множества соответствует набор симптомов $\{S_1, S_2, S_3, S_4\}$, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

• $\alpha = 0.0620$ нечеткого множества соответствует набор симптомов $\{S_1, S_3, S_4\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.999))(1 + 0.9 \cdot (-0.999))(1 + 0.3 \cdot (-0.999)) - 1}{-0.999} = 0.994,$$

выбираем $\min\{0.0620; 0.994\} = 0.0620$;

• $\alpha = 0.1688$ нечеткого множества соответствует набор симптомов $\{S_3, S_4\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.999))(1 + 0.3 \cdot (-0.999)) - 1}{-0.999} = 0.93,$$

выбираем $\min\{0.1688; 0.93\} = 0.1688$;

• $\alpha = 0.3333$ нечеткого множества соответствует набор симптомов $\{S_4\}$, это

$$G_\lambda = \frac{(1 + 0.3 \cdot (-0.999)) - 1}{-0.999} = 0.3,$$

выбираем $\min\{0.3333; 0.3\} = 0.3$.

Обобщенный показатель симптоматики болезни d_4 равен $e(d_4) = \max\{0; 0.0620; 0.1688; 0.3\} = 0.3$.

Рассмотрим диагноз d_5 и разложим его по α-уровням:

• $\alpha = 0$ нечеткого множества соответствует набор симптомов $\{S_1, S_2, S_3, S_4\}$, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

• $\alpha = 0.0310$ нечеткого множества соответствует набор симптомов $\{S_1, S_4\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.999))(1 + 0.3 \cdot (-0.999)) - 1}{-0.999} = 0.93,$$

выбираем $\min\{0.0310; 0.93\} = 0.0310$;

• $\alpha = 0.5845$ нечеткого множества соответствует набор симптомов $\{S_4\}$, это

$$G_\lambda = \frac{(1 + 0.3 \cdot (-0.999)) - 1}{-0.999} = 0.3,$$

выбираем $\min\{0.5845; 0.3\} = 0.3$.

Обобщенный показатель симптоматики болезни d_5 равен $e(d_5) = \max\{0; 0.0310; 0.3\} = 0.3$.

Ищем вариант диагноза, для этого выбираем максимум по обобщенной симптоматике

$$e(d_0) = \max\{e(d_1), e(d_2), e(d_3), e(d_4), e(d_5)\} =$$

$$= \max\{0.0124; 0.5567; 0.3056; 0.3000; 0.3000\} = 0.5567,$$

эта величина обобщенного показателя соответствует диагнозу d_2.

Таким образом, предварительный класс диагнозов по заданным симптомам — это воспаление век.

Как и в [12], рассмотрим еще одного пациента, обозначим его пациент 2: возраст — 14 лет; первичный симптом — снижение зрения; возможная причина — травма глаза.

Преобразуем исходную базу данных (табл. 7).

Находим $\tilde{\lambda}$ из условия

$$\frac{1}{\lambda}\left((1 + \lambda 0.9)(1 + \lambda 0.9)(1 + \lambda 0.9) - 1\right) = 1.$$

Графическое решение данного уравнения представлено на рис. 2. Получаем, что $\tilde{\lambda} \approx -0.99897$.

Рассмотрим диагноз d_1 и разложим его по α-уровням:

• $a = 0$ нечеткого множества соответствует набор симптомов S_1, S_2, S_3, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

Таблица 7. **База данных симптомов-болезней для обследуемого пациента 2**

Данные о заболевании глаз	Выраженность симптомов данного пациента, g	Гордеолум и халазион, d_1	Воспаление век, d_2	Болезни век, d_3	Болезни слезного аппарата, d_4	Болезни глазницы, d_5
Возраст 3–14 лет (S_1)	0.9 — значительно выражен	0.0124	0.0124	0	0.0620	0.0310
Снижение остроты зрения (S_2)	0.9 — значительно выражен	0	0	0.7222	0.1688	0.2500
Травма глаза (S_3)	0.9 — значительно выражен	0	0	0.4167	0	0

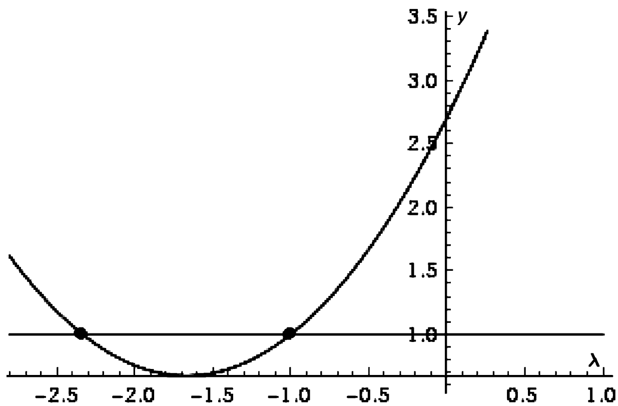

Рис. 2. Графическое решение уравнения поиска значения $\tilde{\lambda}$ из условия нормировки λ-нечеткой меры Сугено для пациента 2

• $a = 0.0124$ нечеткого множества соответствует набор симптомов $\{S_1\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99897)) - 1}{-0.99897} = 0.9,$$

выбираем $\min\{0.0124; 0.9\} = 0.0124$.

Обобщенный показатель симптоматики болезни d_1 равен $e(d_1) = \max\{0; 0.0124\} = 0.0124$.

Рассмотрим диагноз d_2 и разложим его по α-уровням. Аналогично диагнозу d_1 получим обобщенный показатель симптоматики болезни d_2: $e(d_2) = \max\{0;\ 0.0124\} = 0.0124$.

Рассмотрим диагноз d_3 и разложим его по α-уровням:

• $\alpha = 0$ нечеткого множества соответствует набор симптомов $\{S_1, S_2, S_3\}$, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

• $\alpha = 0.4167$ нечеткого множества соответствует набор симптомов $\{S_2, S_3\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99897))(1 + 0.9 \cdot (-0.99897)) - 1}{-0.99897} = 0.99,$$

выбираем $\min\{0.4167; 0.99\} = 0.4167$;

• $\alpha = 0.7222$ нечеткого множества соответствует набор симптомов $\{S_2\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99897)) - 1}{-0.99897} = 0.9,$$

выбираем $\min\{0.7222; 0.9\} = 0.7222$.

Обобщенный показатель симптоматики болезни d_3 равен $e(d_3) = \max\{0; 0.4167; 0.7222\} = 0.7222$.

Рассмотрим диагноз d_4 и разложим его по α-уровням:

• $\alpha = 0$ нечеткого множества соответствует набор симптомов $\{S_1, S_2, S_3\}$, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

$\alpha = 0.0620$ нечеткого множества соответствует набор симптомов $\{S_1, S_2\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99897))(1 + 0.9 \cdot (-0.99897)) - 1}{-0.99897} = 0.99,$$

выбираем $\min\{0.0620; 0.99\} = 0.0620$;

• $\alpha = 0.1688$ нечеткого множества соответствует набор симптомов $\{S_2\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99897)) - 1}{-0.99897} = 0.9,$$

выбираем $\min\{0.1688; 0.9\} = 0.1688$.

Обобщенный показатель симптоматики болезни d_4 равен $e(d_4) = \max\{0; 0.0620; 0.1688\} = 0.1688$.

Рассмотрим диагноз d_5 и разложим его по α-уровням:

• $\alpha = 0$ нечеткого множества соответствует набор симптомов $\{S_1, S_2, S_3\}$, это $G_\lambda = 1$, выбираем $\min\{0; 1\} = 0$;

• $\alpha = 0.0310$ нечеткого множества соответствует набор симптомов $\{S_1, S_2\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99897))(1 + 0.9 \cdot (-0.99897)) - 1}{-0.99897} = 0.99,$$

выбираем $\min\{0.0310; 0.99\} = 0.0310$;

• $\alpha = 0.2500$ нечеткого множества соответствует набор симптомов $\{S_2\}$, это

$$G_\lambda = \frac{(1 + 0.9 \cdot (-0.99897)) - 1}{-0.99897} = 0.9,$$

выбираем $\min\{0.2500; 0.9\} = 0.25$.

Обобщенный показатель симптоматики болезни d_5 равен $e(d_5) = \max\{0; 0.0310; 0.25\} = 0.25$.

Ищем вариант диагноза, для этого выбираем максимум по обобщенной симптоматике $e(d_0) = \max\{e(d_1),\ e(d_2),\ e(d_3),\ e(d_4),\ e(d_5)\} = \max\{0.0124; 0.0124; 0.7222; 0.16880; 0.25\} = 0.7222$, эта величина обобщенного показателя соответствует диагнозу d_3.

Таким образом, предварительный класс диагнозов по заданным симптомам — это болезни век.

Заключение. В статье предложено использовать нечеткие базы знаний Сугено для постановки предварительного диагноза пациенту с заданным набором симптомов. Приведены примеры поиска диагноза заболеваний глаз для двух пациентов на основе первичной информации и заданных симптомов, а также пример реализации такой базы данных. Точность постановки диагноза зависит от построения исходной базы данных симптомов-болезней с учетом частоты встречаемости этого симптома у диагностированной конкретной болезни.

Предложенный подход позволяет провести обработку большого объема нечеткой информации. При этом экспертная составляющая, полученная статистической обработкой частоты встречаемости данного симптома при конкретном заболевании, а также встречаемости такого заболевания у определенной половозрастной группы, обеспечит высокую точность постановки диагноза при заданном наборе симптомов. Аналитический подход, основанный на исключении из поиска диагноза тех симптомов, информации о которых нет, дают расчет варианта предварительного диагноза за небольшое число математических операций.

Литература

1. *Богомолов А. И., Невежин В. П., Жданов Г. А.* Искусственный интеллект и экспертные системы в мобильной медицине // Хроноэкономика. 2018. № 3 (11). С. 17–28.

2. *Rudenko T. A., Vlasenko M. A.* Fuzzy logic systems in the diagnosis of myocardial dissynchronization // Scientific Journal "Science Rise". 2015. N 5. P. 52–61. https://doi.org/10.15587/2313-8416.2015.43286

3. *Тонеева Д. В., Гончарова А. Б.* Экспертная система диагностики заболеваний // EUROPEAN RESEARCH: сб. статей победителей VI Междунар. науч.-практич. конференции. Пенза: Наука и просвещение, 2016. С. 34–38.

4. *Zadeh L. A.* Fuzzy sets // Information and Control. 1965. Vol. 8. N 3. P. 338–353.

5. *Sugeno M.* Theory of fuzzy integrals and its application: Doct. Thesis. Tokyo: Tokyo Institute of Technology, 1974. 50 p.

6. *Keller J. M., Osborn J.* Training the fuzzy integral // International Journal of Approximate Reasoning. 1996. Vol. 15. Iss. 1. P. 1–24. https://doi.org/10.1016/0888-613X(95)00132-Z

7. *Tahani H., Keller J. M.* Information fusion in computer vision using the fuzzy integral // IEEE Transactions on Systems, Man and Cybernetics. 1990. Vol. 20 (3). P. 733–741. https://doi.org/10.1109/21.57289

8. *Павлов А. Н., Соколов Б. В.* Принятие решения в условиях нечеткой информации: учеб. пособие. СПб.: Изд-во СПбГУАП, 2006. 72 с.

9. *Ротштейн А. П., Штовба С. Д.* Моделирование надежности человека-оператора с помощью нечеткой базы знаний Сугено // Автоматика и телемеханика. 2009. С. 180–187.

10. *Солдатова О. П., Шепелев Ю. М.* Алгоритм минимизации базы правил нечеткой нейронной сети Такаги—Сугено—Канга // EUROPEAN RESEARCH: сб. статей победителей X Междунар. науч.-практич. конференции. Пенза: Наука и просвещение, 2017. С. 46–49.

11. *Павлов А. Н., Соколов Б. В.* Методы обработки экспертной информации: учеб. пособие. СПб.: Изд-во СПбГУАП, 2005. 34 с.

12. *Коробова Л. А., Гладких Т. В.* Разработка модели принятия решения для постановки диагноза заболеваний на основе нечеткой логики // Вестн. ВГУИТ. 2018. Т. 80. № 4. С. 80–89. https://doi.org/10.20914/2310-1202-2018-4-80-89

Статья поступила в редакцию 15 сентября 2019 г.
Статья принята к печати 7 ноября 2019 г.

Контактная информация:

Гончарова Анастасия Борисовна — канд. физ.-мат. наук, ст. преподаватель, a.goncharova@spbu.ru

Preliminary medical diagnostics based on the fuzzy sets theory using the Sugeno measure

A. B. Goncharova

St. Petersburg State University, 7–9, Universitetskaya nab.,
St. Petersburg, 199034, Russian Federation

For citation: Goncharova A. B. Preliminary medical diagnostics based on the fuzzy sets theory using the Sugeno measure. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2019, vol. 15, iss. 4, pp. 529–543. https://doi.org/10.21638/11702/spbu10.2019.409 (In Russian)

When diagnoses a patient, it is necessary for a doctor to conduct a comprehensive and systematic study of the patient in order to collect anamnesis and objectively examine the patient's condition, to conduct the laboratory, radiographic and other types of analyses research, as well as to take into account the presence of other diseases inpatient. Processing a large amount of information about the patient and correlating this information with the symptoms of various diseases, not always relevant to the particular physician specialization is a difficult task, which can be solved by the diagnostic support system, basing on the available information. The article proposes a method of diagnostics for given symptoms based on the theory of fuzzy sets. For a given set of symptoms expressed, a generalized diagnosis index is calculated using the Sugeno measure. Using the search for the maximum of generalized indicators, the most probable diagnosis is made.

Keywords: fuzzy sets, Sugeno measure, fuzzy set, Sugeno measure, decision support system, medical diagnostics, multicriteria choice, composite index.

References

1. Bogomolov A. I., Nevezhin V. P., Zhdanov G. A. Iskusstvennyj intellekt i ekspertnyie sistemy v mobil'noi meditsine [Artificial intelligence and expert systems in mobile medicine]. *Chronoeconomics*, 2018, no. 3(11), pp. 17–28. (In Russian)

2. Rudenko T. A., Vlasenko M. A. Fuzzy logic systems in the diagnosis of myocardial dissynchronization. *Scientific Journal "Science Rise"*, 2015, no. 5, pp. 52–61. https://doi.org/10.15587/2313-8416.2015.43286

3. Toneeva D. V., Goncharova A. B. Ekspertnaia sistema diagnostiki zabolevanyj [Expert system for diagnosing diseases]. *EUROPEAN RESEARCH. Collection of articles by the winners of the VI International scientific and practical conference.* Penza, Science and Education Publ., 2016, pp. 34–38. (In Russian)

4. Zadeh L. A. Fuzzy sets. *Information and Control*, 1965, vol. 8, no. 3, pp. 338–353.

5. Sugeno M. *Theory of fuzzy integrals and its application.* Doct. Thesis. Tokyo, Tokyo Institute of Technology Publ., 1974, 50 p.

6. Keller J. M., Osborn J. Training the fuzzy integral. *International Journal of Approximate Reasoning*, 1996, vol. 15, iss. 1, pp. 1–24. https://doi.org/10.1016/0888-613X(95)00132-Z

7. Tahani H., Keller J. M. Information fusion in computer vision using the fuzzy integral. *IEEE Transactions on Systems, Man and Cybernetics*, 1990, vol. 20(3), pp. 733–741. https://doi.org/10.1109/21.57289

8. Pavlov A. N., Sokolov B. V. *Priniatie reshenia v usloviyah nechenkoy informatsii* [*Decision making in the conditions of fuzzy information*]. Textbook. St. Petersburg, SPbGUAP Publ., 2006, 72 p. (In Russian)

9. Ropshtein A. P., Shtovba S. D. Modelirovanie nadezhnosti cheloveka-operatora s pomoschiyu nechetkoy bazy znaniy Sugeno [The reliability of a human operator using a fuzzy Sugeno knowledge base]. *Automation and telemechanics*, 2009, no. 1, pp. 180–187. (In Russian)

10. Soldatova O. P., Shepelev Yu. M. Algoritm minimizatsii bazy pravil nechetkoy neironnoy seti Takagi—Sugeno—Kanga [Algorithm for minimizing the rule base of the fuzzy neural network Takagi—Sugeno—Kanga]. *EUROPEAN RESEARCH. Collection of articles by the winners of the X International Scientific and Practical Conference.* Penza, Science and Education Publ., 2017, p. 46–49. (In Russian)

11. Pavlov A. N., Sokolov B. V. *Metody obrabotki ekspertnoy informatsii* [*Methods for processing expert information*]. Textbook. St. Petersburg. SPbGUAP Publ., 2005, 34 p. (In Russian)

12. Korobova L. A., Gladkikh T. V. Razrabotka modeli priniatia reshenia dlya postanovki diagnoza na osnove nechetkoy logiki [Development of a decision making model for diagnosing diseases based on fuzzy logic]. *Vestnik VGUIT* [*Proceedings of VSUET*], 2018, vol. 80, no. 4, pp. 80–89. (In Russian) https://doi.org/10.20914/2310-1202-2018-4-80-89

Received: September 15, 2019.
Accepted: November 07, 2019.

Author's information:

Anastasiya B. Goncharova — PhD in Physics and Mathematics, Senior Lecturer; a.goncharova@spbu.ru