

*Александр Анатольевич ЗАХАРОВ —
зав. кафедрой информационной безопасности, доктор
технических наук, профессор*

*Евгений Александрович ОЛЕННИКОВ —
доцент кафедры информационной безопасности,
кандидат технических наук*

*Алексей Сергеевич ПЕТУХОВ — аспирант кафедры
информационной безопасности
Тюменский государственный университет*

УДК 004.9:61

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОЦЕНКИ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ В ДИАГНОСТИЧЕСКОЙ ПОДСИСТЕМЕ В СОСТАВЕ ЭЛЕКТРОННОЙ ИСТОРИИ БОЛЕЗНИ*

АННОТАЦИЯ. В статье приводится обзор математических методов для расчета диагностической ценности результатов медицинских исследований и качественных оценок значений медицинских показателей, а также математических методов оценки динамики изменения значений медицинских показателей, полученных в разные моменты времени.

This article gives review of mathematical methods for calculation of diagnostic cost of medical tests' results and for calculation of qualitative marks for medical proofs' values and mathematical methods for evaluation of dynamics of changes of medical proofs' values obtained at different moments.

Одним из аспектов современного этапа интеграции информационных технологий в сферу медицины является разработка информационных систем «Электронная история болезни». Внедрение таких систем в работу медицинских учреждений привело к быстрому накоплению большого объема информации в электронном виде, удобном для обработки. Статистическая обработка больших объемов медицинской информации позволяет выявлять в ней различные закономерности, что зачастую является основой научных исследований в медицине, а обработка накопленных результатов исследований конкретного пациента позволяет облегчить процесс диагностики его заболеваний. Для помощи лечащему врачу в диагностике заболеваний пациента создаются различные медицинские диагностические экспертные системы.

В настоящее время создано большое количество медицинских диагностических экспертных систем с различными возможностями, обусловленными различным генезисом диагностируемых ими заболеваний. В качестве наиболее крупных медицинских экспертных систем можно назвать систему для диагностики

* Работа выполнена при поддержке гранта на научные и экспедиционные исследования для аспирантов Тюменского государственного университета.

инфекционных заболеваний «MYCIN»[1], терапевтическую экспертную систему «Internist»[2], а также отечественную разработку — систему «Диаген»[3], предназначенную для комплексной диагностики заболеваний пациента. Однако большинство медицинских диагностических экспертных систем ориентировано на работу в диалоговом режиме с пользователем, что не позволяет эксплуатировать подобные системы для обработки информации напрямую из базы данных информационной системы типа «Электронная история болезни». В связи с этим актуальной задачей является разработка диагностической подсистемы в составе системы типа «Электронная история болезни», обрабатывающей данные непосредственно из базы данных этой системы. Данная разработка предназначена для помощи лечащему врачу в диагностике заболеваний пациента на основе большого объема накопленных результатов его исследований.

Создание диагностической подсистемы в составе системы типа «Электронная история болезни» требует привлечения или разработки математических методов для решения ряда задач. При этом, на наш взгляд, наиболее актуальными задачами являются: 1) задача оценки динамики изменения значений числовых показателей результатов исследований во времени для диагностики сложных и хронических заболеваний; 2) задача качественных оценок значений числовых показателей, которые широко используются практикующими врачами при диагностике заболеваний; 3) задача оценки устаревания результатов исследований, т.е. оценки их диагностической ценности на заданный момент времени. Для решения данных задач нами были предложен ряд математических методов, которые описаны ниже.

Введем некоторые обозначения. Пусть всего задано N показателей результатов исследований. Пусть для каждого показателя имеется M_i его значений, полученных в разное время в результате проведения разных исследований. Пусть $v_{ij} \in V_i, i \in \{1..N\}, j \in \{1..M_i\}$ — j -ое значение i -го показателя, где V_i — множество всех возможных значений i -го показателя. Обозначим через $h_{ij}, i \in \{1..N\}, j \in \{1..M_i\}$ количество дней, прошедших с момента получения значения показателя v_{ij} и через H_i — количество дней, на протяжении которых значение показателя имеет какую-либо диагностическую ценность, т.е. не является полностью устаревшим. Для каждого числового показателя введем также диапазон $[n_i^l; n_i^r], i \in \{1..N\}$, который будет включать значения этого показателя, допустимые для здорового человека. Пусть также задано некоторое множество $D = (d_1, d_2, \dots, d_k)$, включающее заболевания, гипотезы о которых можно выдвинуть по заданной информации о показателях результатов исследований. Таким образом, конечная задача рассматриваемой диагностической подсистемы сводится к выдвигению гипотезы о некотором множестве заболеваний $\{d\} \in D$ на основе анализа заданной информации о показателях результатов обследований:

$$v_{ij}, h_{ij}, n_i^l, n_i^r, V_i, H_i, i \in \{1..N\}, j \in \{1..M_i\}.$$

Для решения задачи оценки динамики изменения значений некоторого числового показателя во времени необходимо разбить множество всех значений этого показателя, полученных в разное время, на кластеры хронологически близких значений, а затем оценить динамику изменения значений каждого кластера. В качестве меры хронологической близости будем использовать модуль разницы количества дней, прошедших с моментов получения значений показателей:

$$\rho(h_{ik}, h_{il}) = |h_{ik} - h_{il}|, i \in \{1..N\}, k, l \in \{1..M_i\}. \text{ Введем также } \rho_{\max} \text{ — некоторое такое}$$

число, что если $p(h_{ik}, h_{il}) \leq p_{max}$, то значения h_{ik} и h_{il} объединяются в один кластер, иначе — нет. В результате такого разбиения, мы получаем разбиение множества $\{h_{ij}\}, i \in \{1..N\}, j \in \{1..M_i\}$ и, соответственно, множества $\{v_{ij}\}, i \in \{1..N\}, j \in \{1..M_i\}$ на T_i кластеров. Обозначим через $Q_{ik}, i \in \{1..N\}, k \in \{1..T_i\}$ множество кластеров, полученных в результате кластерного анализа множества $\{v_{ij}\}, i \in \{1..N\}, j \in \{1..M_i\}$. Обозначим через $M_{ik}^q, i \in \{1..N\}, k \in \{1..T_i\}$ количество элементов в k -ом кластере для i -го показателя. Обозначим также через $v'_{im}, i \in \{1..N\}, m \in \{1..M_{ik}^q\}$ такие $v_{ij}, i \in \{1..N\}, j \in \{1..M_i\}$ что $v'_{im} = v_{ij} : v_{ij} \in Q_{ik}, i \in \{1..N\}, j \in \{1..M_i\}, k \in \{1..T_i\}, m \in \{1..M_{ik}^q\}$. Соответственно, количество дней, прошедшее с момента получения значений показателей v'_{im} обозначим через $h'_{im}, i \in \{1..N\}, m \in \{1..M_{ik}^q\}$. В качестве оценки изменения значений кластера возьмем линейный тренд, построенный по значениям показателя, входящих в данный кластер:

$$v'_{ij}{}^{расч} = a_{ik} h'_{ij} + b_{ik}, i \in \{1..N\}, j \in \{1..M_{ik}^q\} k = 1..T_i \quad (1)$$

Где:

$v'_{ij}{}^{расч}$ — расчетные значения показателя.

h'_{ik} — количество дней, прошедших соответственно с моментов получения значений показателей v'_{ik} .

a_{ik}, b_{ik} — некоторые величины, значения которых необходимо определить.

Значения a_{ik} и b_{ik} находим по методу наименьших квадратов:

$$F_{ik} = \sum_{j=1}^{M_{ik}^q} [v'_{ij} - v'_{ij}{}^{расч}]^2 = \sum_{j=1}^{M_{ik}^q} [v'_{ij} - a_{ik} h'_{ij} - b_{ik}]^2 \rightarrow \min, i \in \{1..N\}, i \in \{1..N\}, \quad (2)$$

$$\begin{cases} a_{ik} \sum_{j=1}^{M_{ik}^q} h'_{ij} + M_{ik}^q b_{ik} = \sum_{j=1}^{M_{ik}^q} v'_{ij} \\ a_{ik} \sum_{j=1}^{M_{ik}^q} (h'_{ij})^2 + b_{ik} \sum_{j=1}^{M_{ik}^q} h'_{ij} = \sum_{j=1}^{M_{ik}^q} v'_{ij} h'_{ij} \end{cases}, i \in \{1..N\} i \in \{1..N\}. \quad (3)$$

Решая данную систему уравнений, например, методом Крамера, находим a_{ik} — тангенс угла наклона линейного тренда и b_{ik} — смещение линейного тренда по оси ординат. По тангенсу угла наклона полученного линейного тренда для кластера Q_{ik} определяем тенденцию изменения входящих в него значений показателя: возрастание или убывание. Соответственно, отрицательное значение тангенса угла наклона свидетельствует об убывании функции, а положительное значение — о возрастании. Причем чем больше по модулю тангенс угла наклона линейного тренда, тем явнее прослеживается соответствующая тенденция. Это в значительной степени совпадает с основным свойством нечеткого множества: чем больше значение функции принадлежности для того или иного элемента, тем «увереннее» он принадлежит нечеткому множеству [4]. Опираясь на данное сходство,

введем два нечетких множества: «Возрастает» и «Убывает» для описания «уверенности» в возрастании и убывании полученного линейного тренда:

$$\begin{aligned} \text{«Возрастает»} = \{ \text{«Возрастающий» } (a_{ik})/a_{ik} \}, \quad \text{«Убывает»} = \{ \text{«Убывающий» } (a_{ik})/a_{ik} \}, \\ i \in \{1..N\}, k = 1..T_i. \end{aligned} \quad (4)$$

Определим следующие функции принадлежности данным нечетким множествам.

$$\text{«Убывающий» } (a_{ik}) = \begin{cases} -\frac{2}{\pi} \arctg(a_{ik}), & \frac{180}{\pi} \arctg(a_{ik}) \in (-90;0) \\ 0, & \frac{180}{\pi} \arctg(a_{ik}) \in (0;90) \end{cases}, \quad i \in \{1..N\}, k = 1..T_i, \quad (5)$$

$$\text{«Возрастающий» } (a_{ik}) = \begin{cases} 0, & \frac{180}{\pi} \arctg(a_{ik}) \in (-90;0) \\ \frac{2}{\pi} \arctg(a_{ik}), & \frac{180}{\pi} \arctg(a_{ik}) \in (0;90) \end{cases}, \quad i \in \{1..N\}, k = 1..T_i. \quad (6)$$

Вид данных функций приведен на рис. 1.

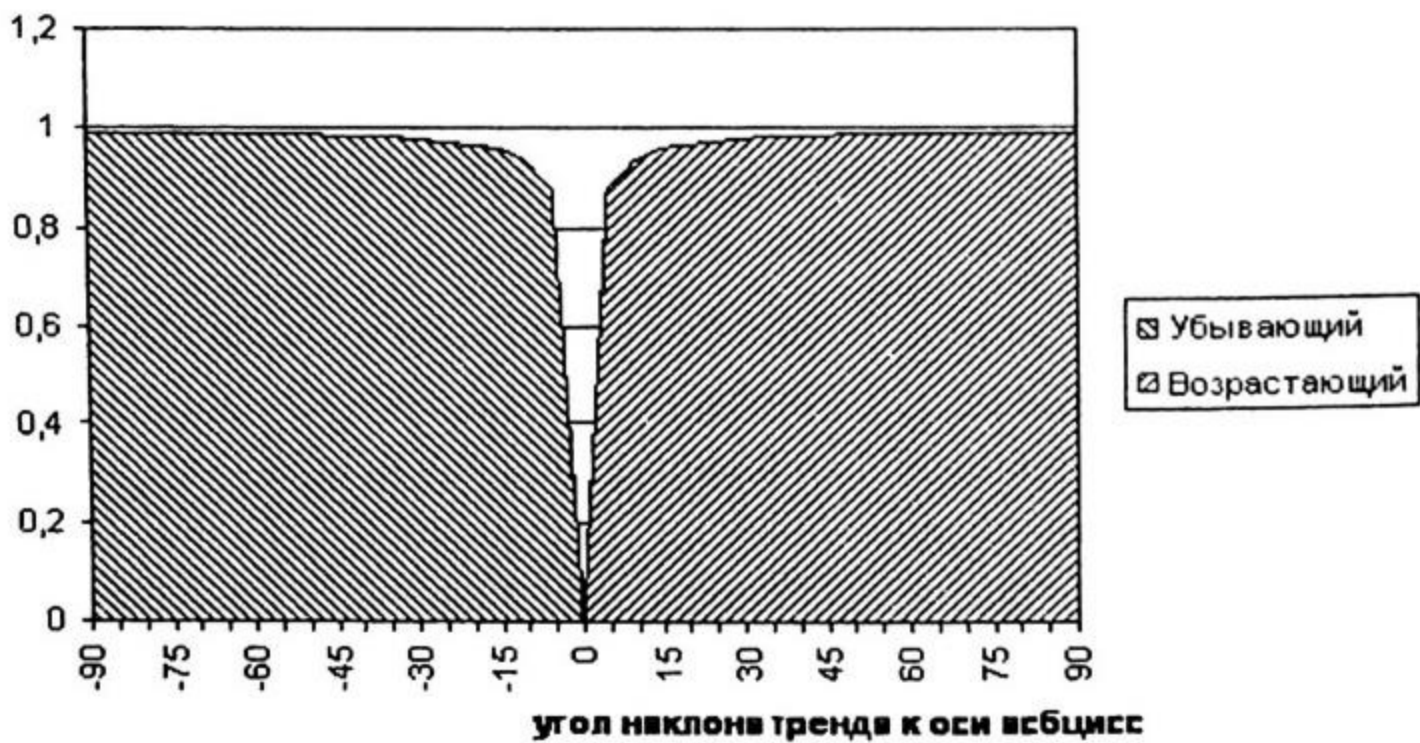


Рис. 1. Вид функций принадлежности «Убывающий» и «Возрастающий»

Таким образом, находя тангенс угла наклона линейного тренда каждого кластера и подставляя его значение в приведенные функции принадлежности нечетких множеств «Возрастает» и «Убывает», мы получаем оценку изменения динамики значений показателя, входящих в данный кластер.

Задача введения качественных оценок значений числовых показателей вытекает из сложности выделения четких числовых диапазонов, принадлежность значений показателей к которым позволяет делать определенные выводы. Рассмотрим некоторые качественные оценки значений показателей: «низкое», «норма» и «высокое». Каждую качественную оценку будем рассматривать как нечеткое множество с соответствующей функцией принадлежности (7):

$$\text{«Низкое»} = \{ \text{«Низкий» } (v_{ij})/v_{ij} \} \quad \text{«Норма»} = \{ \text{«Нормальный» } (v_{ij})/v_{ij} \}, \quad (7)$$

$$\text{«Высокое»} = \{ \text{«Высокий» } (v_{ij}) / v_{ij} \}, \quad i \in \{1..N\}, j \in \{1..M_i\}.$$

Нами были предложены трапециевидальные функции принадлежности (8), (9), (10). Они были получены из уравнения прямой по двум точкам. Например, прямая функции принадлежности «Низкий» была получена по точкам $(n_i^l(1-fb/2); 1)$ и $(n_i^l(1+fb/2); 0)$, $i \in \{1..N\}$. Абсциссы $(n_i^l(1-fb/2); 1)$ и $(n_i^l(1+fb/2); 0)$ были выбраны так, чтобы интервал размытости вокруг точки n_i был равен f_b и чтобы n_i была серединой этого интервала.

$$\text{«Низкий» } (v_{ij}) = \begin{cases} 1, & v_{ij} \in (-\infty; n_i^l(1-fb/2)) \\ \frac{n_i^l(1-fb/2) - v_{ij}}{n_i^l(1+fb/2) - n_i^l(1-fb/2)} + 1, & v_{ij} \in (n_i^l(1-fb/2); n_i^l(1+fb/2)] \\ 0, & v_{ij} \in (n_i^l(1+fb/2); \infty) \end{cases} \quad (8)$$

$$\text{«Нормальный» } (v_{ij}) = \begin{cases} 0, & v_{ij} \in (-\infty; n_i^l(1-fb/2)] \\ \frac{v_{ij} - n_i^l(1-fb/2)}{n_i^l(1+fb/2) - n_i^l(1-fb/2)}, & v_{ij} \in (n_i^l(1-fb/2); n_i^l(1+fb/2)] \\ 1, & v_{ij} \in (n_i^l(1+fb/2); n_i^r(1-fb/2)] \\ \frac{n_i^r(1-fb/2) - v_{ij}}{n_i^r(1+fb/2) - n_i^r(1-fb/2)} + 1, & v_{ij} \in (n_i^r(1-fb/2); n_i^r(1+fb/2)] \\ 0, & v_{ij} \in (n_i^r(1+fb/2); \infty) \end{cases} \quad (9)$$

$$\text{«Высокий» } (v_{ij}) = \begin{cases} 0, & v_{ij} \in (-\infty; n_i^r(1-fb/2)] \\ \frac{v_{ij} - (n_i^r(1-fb/2))}{n_i^r(1+fb/2) - n_i^r(1-fb/2)}, & v_{ij} \in (n_i^r(1-fb/2); n_i^r(1+fb/2)] \\ 1, & v_{ij} \in (n_i^r(1+fb/2); \infty) \end{cases} \quad (10)$$

$$i \in \{1..N\}, j \in \{1..M_i\}$$

Где:

fb («Fuzzy Blur») — заданный коэффициент размытости функций принадлежности нечетких множеств на границах.

Схематический вид функций принадлежности (8), (9), (10) показан на рис. 2.

Таким образом, проверяя на принадлежность значение числового показателя некоторому нечеткому множеству — «Низкое», «Норма» или «Высокое», мы переходим от непосредственной количественной оценки этого значения к качественной, лингвистической его оценке.

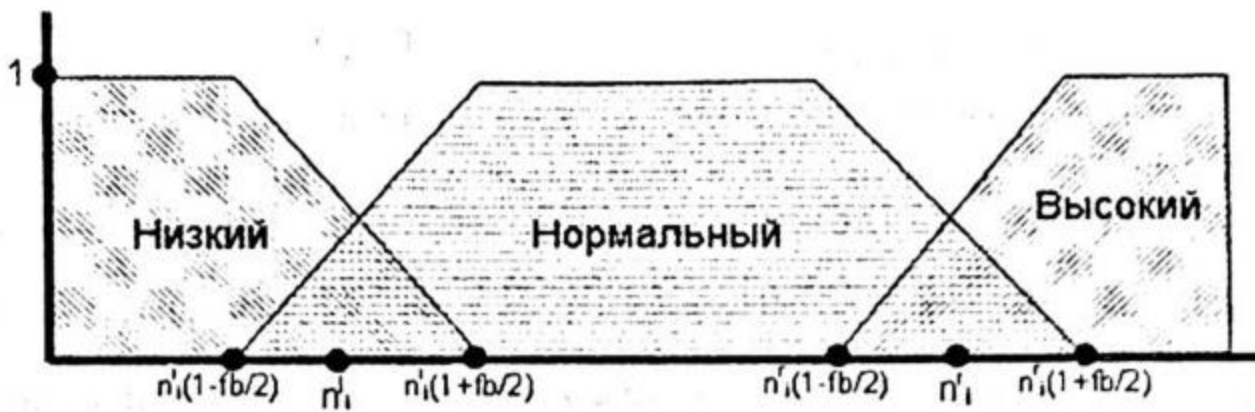


Рис. 2. Схематический вид функций принадлежности «Низкий», «Нормальный» и «Высокий»

Задача оценки устаревания результатов исследований состоит в разделении результатов исследований на «свежие» или актуальные на данный момент времени и «устаревшие» или неактуальные. Решение данной задачи требует введения функции оценки устаревания значений показателей результатов исследований $C(h)$ и расчета с помощью нее коэффициентов устаревания значений показателей $C'(h_{ij})$, $i \in \{1..N\}$, $j \in \{1..M_i\}$ и кластеров значений показателей $C'(h_{ij})$, $i \in \{1..N\}$, $j \in \{1..M_i\}$. Заранее условимся, что значения всех коэффициентов устаревания будут принадлежать диапазону $[0; 1]$. Коэффициент устаревания, равный нулю, будем присваивать только что полученным значениям показателей, а коэффициент устаревания, равный единице, — полностью устаревшим значениям.

Очевидно, что можно предложить различные виды функции устаревания.

Процесс устаревания можно в простейшем случае описать как резкий переход от актуальных значений показателя к устаревшим его значениям. Например, значение некоторого показателя можно считать актуальным в течение 30 дней, а после этого считать его устаревшим. В данном случае процесс устаревания можно описать как ступенчатую функцию:

$$C(h) = \begin{cases} 0, & h < H \\ 1, & h \geq H \end{cases} \quad (11)$$

Где:

h — количество дней, прошедших с момента определения некоторого значения показателя.

H — количество дней, в течение которых данное значение показателя можно считать актуальным.

Чтобы исключить резкий переход от актуальных значений показателей к его устаревшим значениям, можно предложить альтернативные виды функции устаревания. В качестве одного из таких видов можно рассмотреть экспоненциальный вид функции устаревания:

$$C(h) = \begin{cases} \frac{e^{h/H} - 1}{e - 1}, & h < H \\ 1 & h \geq H \end{cases} \quad (12)$$

Где:

h — количество дней, прошедших с момента определения некоторого значения показателя.

H — количество дней, в течение которых данное значение показателя можно считать актуальным.

В степени функции присутствует дробь h/H , введенная для нормирования аргумента функции. В числителе функции присутствует минус 1 для того, чтобы «снизить» функцию по оси ординат и получить пересечение с осью ординат в точке (0; 0). Знаменатель функции необходим для ее нормирования по оси абсцисс и, в частности, для того чтобы при $h=H$ функция принимала значение 1.

Вид данной функции приведен на рис. 3.

С помощью той или иной функции устаревания рассчитываем коэффициенты устаревания значений показателей:

$$C'(h_{ij}) = C(h_{ij}), \quad i \in \{1..N\}, \quad j \in \{1..M_i\}. \quad (13)$$

Далее по уже известным коэффициентам устаревания значений показателей необходимо рассчитать значения коэффициентов устаревания кластеров значений этих показателей $C'(h_{ij}), i \in \{1..N\}, j \in \{1..M_i\}$. Для оценки коэффициентов устаревания кластеров значений показателей можно предложить также различные варианты расчета.

1. Средняя оценка: среднее арифметическое входящих в заданный кластер коэффициентов устаревания значений показателей:

$$C_{ik}^* = \frac{\sum_{j=1}^{M_{ik}^q} C'_{ij} : h_{ij} \in Q_{ik}}{M_{ik}^q}, \quad i \in \{1..N\}, \quad k \in \{1..T_i\}. \quad (14)$$

2. Пессимистическая оценка: максимум входящих в заданный кластер коэффициентов устаревания значений показателей:

$$C_{ik}^* = \max_j C'_{ij} : h_{ij} \in Q_{ik}, \quad i \in \{1..N\}, \quad j = 1..M_{ik}^q, \quad k \in \{1..T_i\}. \quad (15)$$

3. Оптимистическая оценка: минимум входящих в заданный кластер коэффициентов устаревания значений показателей:

$$C_{ik}^* = \min_j C'_{ij} : h_{ij} \in Q_{ik}, \quad i \in \{1..N\}, \quad j = 1..M_{ik}^q, \quad k \in \{1..T_i\}. \quad (16)$$

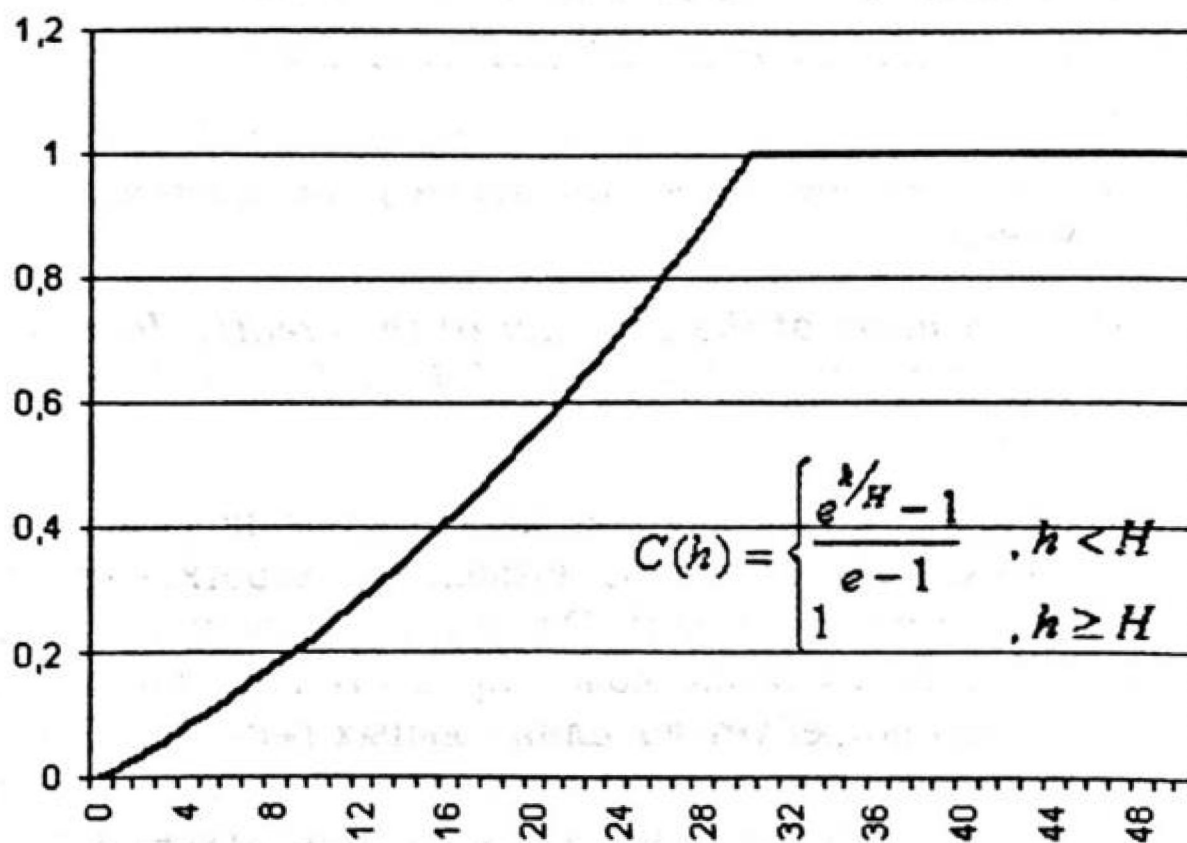


Рис. 3. Экспоненциальный вид функции устаревания

Таким образом, в результате работы были предложены математические методы для оценки динамики изменения значений числовых показателей, основанные на методе кластерного анализа и теории нечетких множеств. Также были разработаны методы для качественной оценки значений числовых показателей, опирающиеся на теорию нечетких множеств. Кроме того, предложены различные виды функций устаревания значений показателей и различные варианты расчета коэффициентов устаревания кластеров значений показателей.

СПИСОК ЛИТЕРАТУРЫ

1. Уэно Х., Исудзука М. Представление и использование знаний / М.: Мир, 1989. С. 50-54.
2. Джексон П. Введение в экспертные системы. М.: Вильямс, 2000. С. 213-217.
3. Кобринский Б.А., Фельдман А.Е. Анализ и учет ассоциативных знаний в медицинских экспертных системах //Новости искусств. интеллекта. 1995. № 3. С. 90-96.
4. Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Д.А. Поспелова. М.: Наука, 1986. 312 с.

*Александр Григорьевич ИВАШКО —
зав. кафедрой информационных систем
Института математики и компьютерных наук,
доктор технических наук, профессор*

*Михаил Викторович ГРИГОРЬЕВ —
старший преподаватель
кафедры информационных систем
Института математики и компьютерных наук
Тюменский государственный университет*

УДК 004.4'41

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЙ ЯЗЫК ОГРАНИЧЕНИЙ ДЛЯ ВЕРИФИКАЦИИ ПРОЦЕССА КОМАНДНОЙ РАЗРАБОТКИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

АННОТАЦИЯ. В статье описывается компилятор формального языка, реализующий правила верификации процесса разработки программного обеспечения в команде.

In article implementation of the compiler of the created formal description language of rules of verification of process of the software development in the command is described.

Современные методы программной инженерии предполагают итеративную командную разработку. Данный подход позволяет в короткие сроки создавать качественные программные продукты. Однако допущение ошибки одним из участников проекта на любом этапе может привести к созданию новой итерации, что в свою очередь потребует дополнительных ресурсов. В связи с этим задача верификации процессов выполнения программного проекта становится крайне важной. Многие из правил, известные в литературе, могут использоваться в настоящее время только при верификации в ручном режиме, что