

УДК 007.5; 004.85

А.С. Горностаев, С.Ф. Чалый

Харьковский национальный университет радиоэлектроники, Харьков

ИССЛЕДОВАНИЕ МЕТОДОВ ОТБОРА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКИХ ЗАПРОСОВ НА ПРИМЕРЕ ИНТЕРНЕТ-МАГАЗИНА

В статье рассмотрена проблема поиска товаров в интернет - магазинах и методы решения этой проблемы- использование алгебры нечеткой логики при работе с запросами. Предложен способ перехода от нечетких запросов к четким на основе применения трапецевидной функции принадлежности.

***Ключевые слова:** трапецевидная функция принадлежности, алгебра нечеткой логики, нечеткие запросы.*

Введение

Постановка проблемы. В настоящее время интернет-магазины становятся все более востребо-

ванными. При покупке разнообразных товаров клиенты таких магазинов выбирают необходимую им продукцию, устанавливая требуемые значения параметров искомого товара [1].

Для хранения данных о товарах традиционно используются реляционные базы данных, основанные на реляционной модели данных. В реляционных базах данных все данные, которые доступны пользователю, организованы в виде таблиц, а все операции над данными сводятся к операциям над этими таблицами.

При применении четких запросов к базе данных интернет-магазина, часто у пользователей возникает проблема при формировании набора параметров и определении их конкретных значений для искомого товара, другими словами, возникает проблема поиска товаров на основе нечетких критериев. В связи с этой проблемой, требуется применить соответствующие математические методы для перехода от нечетких запросов к четким, что позволит упростить задачу поиска.

Большая часть данных, обрабатываемых в современных информационных системах, носят четкий, числовой характер. Однако в запросах к базам данных, которые пытается сформулировать человек, часто присутствуют неточности и неопределенности, причинами которых является семантическая неоднозначность языка. Поэтому для корректного перехода нечетких запросов к четким может быть использован аппарат нечеткой логики.

Точные значения входных переменных преобразуются в значения лингвистических переменных посредством применения некоторых положений теории нечетких множеств, а именно – при помощи определенных функций принадлежности.

Постановка задачи. В рамках рассмотренной выше проблемы, задача отбора данных с использованием нечетких запросов на примере интернет-магазина формулируется следующим образом: адаптировать и применить метод отбора товаров на основе нечетких запросов для пользователей интернет-магазинов.

Исходными данными задачи являются атрибуты, характеризующие свойства товаров в интернет-магазине. Требуется получить и реализовать процедуру отбора товаров с использованием нечетких запросов.

Для решения этой задачи целесообразно использовать аппарат нечеткой логики [4], позволяющий упростить задачу поиска необходимого товара для пользователя, посредством перехода от нечетких запросов к четким и, в результате, отобрать необходимые товары. Такой подход к отбору товаров позволит пользователям формулировать нечеткие запросы к базе данных и получать ссылки на искомую информацию, упорядоченные по степени релевантности (соответствия) запросу.

Применение алгебры нечеткой логики при работе с запросами к БД

В нечеткой логике значения любой величины представляются не числами, а словами естественно-

го языка. Принадлежность каждого точного значения к одному из термов лингвистической переменной определяется посредством функции принадлежности. Так же требуется определить продукционные правила, связывающие лингвистические переменные. Совокупность таких правил описывает стратегию управления, применяемую в данной задаче.

Характеристикой нечеткого множества выступает функция принадлежности. Степень принадлежности к нечеткому множеству S обозначается как $MF_c(x)$, представляющей собой обобщение понятия характеристической функции обычного множества. Тогда нечетким множеством S называется множество упорядоченных пар вида $S = \{MF_c(x) / x\}$. Значение $MF_c(x) = 0$ означает отсутствие принадлежности к множеству, 1 – полную принадлежность.

Для решения поставленной в работе задачи будет использована трапециевидная функция принадлежности, поскольку она упрощает построение базы знаний и обеспечивает высокую точность вычислений.

Трапециевидная функция формируется с использованием кусочно-линейной аппроксимации и является обобщением треугольной, она позволяет задавать ядро нечеткого множества в виде интервала. В случае трапециевидной функции принадлежности возможна следующая интерпретация: ядро нечеткого множества – оптимистическая оценка; носитель нечеткого множества – пессимистическая оценка.

Аналитическая запись трапециевидной функции имеет вид:

$$\mu(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases},$$

где $[a, d]$ – носитель нечеткого множества – пессимистическая оценка значений переменной, $[b, c]$ – ядро нечеткого множества – оптимистическая оценка значений переменной; x – вектор, для координат которого необходимо рассчитать степени принадлежности.

Таким образом, интерпретация данной функции принадлежности применительно к задаче отбора данных с использованием нечетких запросов на примере интернет-магазина состоит в следующем: для каждого вида товаров должны быть установлены множества, которые обеспечат переход от привычных пользователям лингвистических переменных к числовым значениям параметров в базе данных.

Трапециевидная функция принадлежности имеет вид (рис. 1):

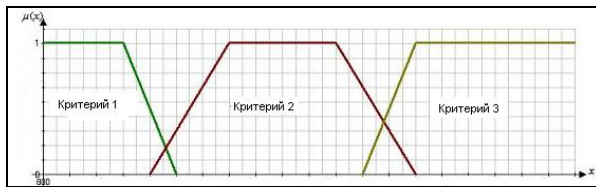


Рис. 1. Трапецевидная функция принадлежности

Для нечетких множеств, как и для обычных, определены основные логические операции. Самыми основными, необходимыми для расчетов, являются пересечение и объединение.

Пересечение двух нечетких множеств (нечеткое "И"): A B:

$$MF_{AB}(x) = \min(MF_A(x), MF_B(x)).$$

Объединение двух нечетких множеств (нечеткое "ИЛИ"): A B:

$$MF_{AB}(x) = \max(MF_A(x), MF_B(x)).$$

Основой для проведения операции нечеткого логического вывода, при решении задачи отбора данных с использованием нечетких запросов, является база правил, содержащая нечеткие высказывания в форме "Если-то" и функции принадлежности для соответствующих лингвистических термов. Результатом нечеткого вывода является четкое значение переменной y^* на основе заданных четких значений $x_k, k = 1..n$.

Метод логического вывода, обеспечивающий возможность проведения отбора данных с использованием нечетких запросов, включает в себя следующую последовательность действий:

1. Выполнение фазификации, при которой определяются значения функций принадлежности для левых частей каждого правила (предпосылок). Для базы правил с m правилами обозначим степени истинности как

$$A_{ik}(x_k), i = 1..m, k = 1..n.$$

2. Реализация нечеткого вывода. Сначала определяются уровни "отсечения" для левой части каждого из правил:

$$\alpha_i = \min_j (A_{ij}(x_k)).$$

Далее находятся "усеченные" функции принадлежности:

$$B_i^*(y) = \min_j (A_{ij}(x_k), B_i(y)).$$

3. Композиция, или объединение полученных усеченных функций, для чего используется максимальная композиция нечетких множеств:

$$MF(y) = \max_i (B_i^*(y)),$$

где $MF(y)$ – функция принадлежности итогового нечеткого множества.

4. Дефазификация. Существует несколько методов дефазификации. Например, метод среднего

центра, или центроидный метод

5. Формирование запроса к БД.

Пример применения аппарата нечеткой логики: Рассмотрим базу данных «Видеокарты».

Предположим, что вся информация находится в одной табл. 1.

Для реализации нечетких запросов формализуем лингвистическую переменную: «цена»:

$$\langle \beta, \chi, T, G, M \rangle,$$

где β – цена; $T = \{\text{«дешевая», «средняя», «дорогая»}\}$;

$\chi = [800; 1800]$; G – процедура образования новых термов с помощью связей «и», «или» и модификаторов «не», «очень», «более-менее»;

M – процедура задания на χ нечетких множеств $A = \text{«дешевая»}$,

$B = \text{«средняя»}$ и $C = \text{«дорогая»}$, а также нечетких множеств для термов из $G(T)$ соответственно правилам трансляции нечетких связей и модификаторов и других операций над нечеткими множествами.

Таблица 1

Информационная таблица

| № | Наименование | Память | Цена |
|----|---------------------|--------|------|
| 1 | GeForce GTS450 | 512 | 1300 |
| 2 | GeForce GTS250 | 512 | 950 |
| 3 | GigaByte GF GTS 450 | 1 Гб | 1380 |
| 4 | ATI Radeon HD5750 | 1 Гб | 1350 |
| 5 | Radeon HD5770 | 1 Гб | 1550 |
| 6 | GeForce GTX460 | 1 Гб | 1520 |
| 7 | GeForce™ GTX 460 | 768 Мб | 1450 |
| 8 | RadeOn HD6850 | 1 Гб | 1750 |
| 9 | GF GTX460 | 1 Гб | 1630 |
| 10 | GigaByte GF GTX 460 | 1 Гб | 1300 |
| 11 | XFx 5830 | 1 Гб | 1480 |

Зададим Функции принадлежности для переменной «цена» трапецевидного вида:

$$\mu_A(x) = \begin{cases} 1, x \in [800, 950] \\ \frac{1050 - x}{150}, x \in (950, 1050]; \\ 0, x \geq 1050 \end{cases}$$

$$\mu_B(x) = \begin{cases} 0, x \leq 1000, x \geq 1500 \\ \frac{x - 1000}{150}, x \in [1000, 1150) \\ 1, x \in [1150, 1350] \\ \frac{1500 - x}{150}, x \in (1350, 1550) \end{cases};$$

$$\mu_C(x) = \begin{cases} 0, x \leq 1400 \\ \frac{x - 1400}{100}, x \in [1400, 1500) \\ 1, x \in [1500, 1800] \end{cases}.$$

График функции принадлежности выглядит показан на рис. 2.

Вычислив степени принадлежности для каждой видеокарты, таблицу можно преобразовать к сле-

дующому виду (табл. 2):

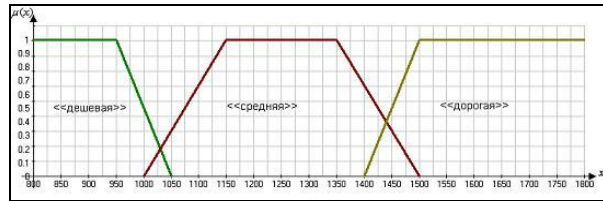


Рис. 2. График функции принадлежности

Таблица 2

Преобразованная таблица

| № | Наименование | Па-мять | Цена | | |
|----|---------------------|---------|----------------|----------------|----------------|
| | | | μ _A | μ _B | μ _C |
| 1 | GeForce GTS450 | 512 | 0 | 1 | 0 |
| 2 | GeForce GTS250 | 512 | 1 | 0 | 0 |
| 3 | GigaByte GF GTS 450 | 1 Гб | 0 | 0,8 | 0 |
| 4 | ATI Radeon HD5750 | 1 Гб | 0 | 1 | 0 |
| 5 | Radeon HD5770 | 1 Гб | 0 | 0 | 1 |
| 6 | GeForce GTX460 | 1 Гб | 0 | 0 | 1 |
| 7 | GeForce™ GTX 460 | 768 Мб | 0 | 0,33 | 0,5 |
| 8 | RadeOn HD6850 | 1 Гб | 0 | 0 | 1 |
| 9 | GF GTX460 | 1 Гб | 0 | 0 | 1 |
| 10 | GigaByte GF GTX 460 | 1 Гб | 0 | 1 | 0 |
| 11 | XFX 5830 | 1 Гб | 0 | 0,13 | 0,8 |

Построим для нее четкие и нечеткие запросы:

Четкий запрос:

«Найти видеокарты стоимостью 1350 грн и»

На языке SQL:

select *from videocards where Cost=1350);

Результат такого запроса будет следующим:

| | | | |
|---|-------------------|------|------|
| 4 | ATI Radeon HD5750 | 1 Гб | 1350 |
|---|-------------------|------|------|

Аналогичный нечеткий запрос:

«Найти видеокарты среднеценового диапазона»

На SQL-подобном языке:

select *from videocards where Cost= «средняя»

Результат – в табл. 3.

ДОСЛІДЖЕННЯ МЕТОДІВ ВІДБОРУ ДАНИХ З ВИКОРИСТАННЯМ НЕЧІТКИХ ЗАПИТІВ НА ПРИКЛАДІ ІНТЕРНЕТ-МАГАЗИНА

С.Ф. Чалий, О.С. Горностаев

В статті розглянута проблема пошуку товарів в інтернет-магазинах і методи рішення цієї проблеми – використання алгебри нечіткої логіки при роботі із запитами. Запропонований спосіб переходу від нечітких запитів до чітких на основі вживання трапецієвидної функції приналежності.

Ключові слова: трапецієвидна функція приналежності, алгебра нечіткої логіки, нечіткі запити.

RESEARCH OF METHODS OF SELECTION OF INFORMATION IS WITH THE USE OF UNCLEAR REQUESTS FOR THE EXAMPLE OF THE INTERNET-SHOP

S.F. Chalyi, A.S. Gornostaev

A problem of search of commodities in the internet shops and methods of decision to this problem- using of unclear boolean algebra for work with queries is considered. The method of transition from unclear queries to clear on the basis of the use of trapezoid function of belonging is offered.

Keywords: to the trapezoid function of belonging, unclear boolean algebra, unclear queries.

Таблица 3

Результирующая таблица

| № | Наименование | Память | Цена | | |
|----|---------------------|--------|----------------|----------------|----------------|
| | | | μ _A | μ _B | μ _C |
| 1 | GeForce GTS450 | 512 | 0 | 1 | 0 |
| 4 | ATI Radeon HD5750 | 1 Гб | 0 | 1 | 0 |
| 10 | GigaByte GF GTX 460 | 1 Гб | 0 | 1 | 0 |
| 3 | GigaByte GF GTS 450 | 1 Гб | 0 | 0,8 | 0 |
| 7 | GeForce™ GTX 460 | 768 Мб | 0 | 0,33 | 0,5 |
| 11 | XFX 5830 | 1 Гб | 0 | 0,13 | 0,8 |

Выводы

В статье обосновано использование аппарата нечеткой логики для задачи поиска информации в базах данных, поскольку это дает следующие основные преимущества:

– упрощается последовательность принятия решений при тех же объемах входной и выходной информации;

– используется гибкая стратегия адаптивного приближения к искомому результату;

– описание необходимых результатов запроса выполняется в терминах лингвистических переменных, что обеспечивает привычный для человека интерфейс.

Список литературы

1. Новак В. Математические принципы нечёткой логики: пер. с англ. / В. Новак, И. Перфильева, И. Мочкрож. – М.: Физматлит, 2006. – 352 с.

2. Тэрано Т. Прикладные нечёткие системы / Т. Тэрано, К. Асаи, М. Сугэно. – М.: Мир, 1993. – 368 с.

3. Нечёткая логика и искусственные нейронные сети / В.В. Крулов, М.И. Дли, Р.Ю. Голунов. – М.: Физматлит, 2001. – 221 с.

4. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений / Л. Заде. – М.: Мир, 1976. – 166 с.

Поступила в редколлегию 17.02.2011

Рецензент: д-р техн. наук, проф. Е.И. Кучеренко, Харьковский национальный университет радиоэлектроники, Харьков.