

Ахметов Б.С., Горбаченко В.И., Кузнецова О.Ю.

НЕЧЕТКИЕ СИСТЕМЫ И СЕТИ

Рекомендовано Научно-методическим советом
Университета в качестве учебного пособия

Алматы 2014

УДК 004(075.8)
ББК 32.973-018.2я73
A95

Рецензенты:

д.т.н., профессор Тукаев У.А., д.т.н., профессор Малыгин А.Ю.,
д.т.н., профессор Джурунтаев Д.З.

A95 Ахметов Б.С., Горбаченко В.И., Кузнецова О.Ю.

Нечеткие системы и сети: учебное пособие. – Алматы: ТОО
«Издательство LEM», 2014. – 104 с.

ISBN 978-601-239-357-6

Пособие посвящено основам нечетких систем и сетей: теории нечетких множеств, нечетким отношениям, нечеткой логике, системам нечеткого вывода и нечетким нейронным сетям. Предназначено для подготовки бакалавров, магистров и докторантов специальностей «Вычислительная техника и программное обеспечение», «Системы информационной безопасности», «Информационные системы», а также всем специалистам, интересующимся вопросами искусственного интеллекта.

Учебное пособие подготовлено авторским коллективом Казахского национального технического университета им. К.И. Сатпаева и Пензенского государственного университета в соответствии с Меморандумом о сотрудничестве.

УДК 004(075.8)
ББК 32.973-018.2я73

*Печатается по решению научно-методического
совета КазНТУ имени К.И. Сатпаева*

ISBN 978-601-239-357-6

© Ахметов Б.С., Горбаченко В.И.,
Кузнецова О.Ю., 2014
© КазНТУ имени К.И. Сатпаева, 2014

СОДЕРЖАНИЕ

Введение	4
1 Основы нечетких множеств.....	6
1.1 Понятие нечетких множеств.....	6
1.2 Функция принадлежности	8
1.2.1 Кусочно-линейные функции принадлежности	10
1.2.2 Z-образные и S-образные функции принадлежности	12
1.2.3 П-образные функции принадлежности	15
1.3 Построение функций принадлежности	16
1.4 Основные характеристики нечетких множеств	18
1.5 Операции над нечеткими множествами	20
1.6 Нечеткие отношения	30
1.7 Лингвистическая переменная. Нечеткие величины, числа и интервалы.....	34
2 Основы нечеткой логики. системы нечеткого вывода.....	44
2.1 Нечеткое высказывание. Логические операции с нечеткими высказываниями.....	44
2.2 Вывод в нечеткой логике.....	49
2.3 Нечеткие лингвистические высказывания	52
2.4 Правила нечетких продукций	53
2.5 Этапы нечеткого вывода	55
2.6 Алгоритмы нечеткого вывода.....	62
3 Понятие нечетких нейронных (гибридных) сетей	65
3.1 Основные понятия и определения нечетких нейронных (гибридных) сетей	66
3.1.1 Архитектура нечеткой нейронной сети ANFIS (Adaptive Network based Fuzzy Inference System)	72
3.1.2 Архитектура NNFLC (Neurons network fuzzy logic controller) ...	74
3.1.3 Архитектура TSK (Takagi, Sugeno, Kang'a)	77
3.2 Обучение нечеткой нейронной сети	80
3.2.1 Обучение нечеткой нейронной сети ANFIS	81
3.2.2 Обучение нечеткой нейронной сети TSK.....	82
3.2.2 Алгоритм нечёткой самоорганизации C-means	86
3.2.2 Алгоритм обратного распространения ошибки для сети TSK	88
3.3 Алгоритмы оптимизации нечеткой нейронной сети	90
Заключение	99
Список использованных источников	101

ВВЕДЕНИЕ

Теория нечетких множеств (fuzzy sets theory) ведет свое начало с 1965 г., когда профессор Лотфи Заде (Lotfi Zadeh) из университета Беркли опубликовал основополагающую работу «Fuzzy Sets» в журнале «Information and Control» [6, 32]. Прилагательное «fuzzy», которое можно перевести на русский как нечеткий, размытый, ворсистый, пушистый, введено в название новой теории с целью дистанцирования от традиционной четкой математики, оперирующей с четкими понятиями: «принадлежит - не принадлежит», «истина-ложь».

Началом практического применения теории нечетких множеств можно считать 1975г., когда Мамдани и Ассилиан (Mamdani and Assilian) построили первый нечеткий контроллер для управления простым паровым двигателем. В 1982 Холмблад и Остергад (Holmblad and Osregaad) разработали первый промышленный нечеткий контроллер, который был внедрен в управление процессом обжига цемента на заводе в Дании. Успех первого промышленного контроллера, основанного на нечетких лингвистических правилах «Если – то» привел к всплеску интереса к теории нечетких множеств среди математиков и инженеров. Несколько позже Бартоломеем Коско (Bart Kosko) [28-29] была доказана теорема о нечеткой аппроксимации (Fuzzy Approximation Theorem) [5], согласно которой любая математическая система может быть аппроксимирована системой, основанной на нечеткой логике. Другими словами, с помощью естественно-языковых высказываний-правил «Если – то», с последующей их формализацией средствами теории нечетких множеств, можно сколько угодно точно отразить произвольную взаимосвязь «входы-выход» без использования сложного аппарата дифференциального и интегрального исчислений, традиционно применяемого в управлении и идентификации [24].

Основанные на этой теории методы построения информационных моделей существенно расширяют традиционные области применения компьютеров и образуют самостоятельное направление

ние научно-прикладных исследований, которое получило специальное название – нечеткое моделирование.

Системы, основанные на нечетких множествах разработаны и успешно внедрены в таких областях, как: управление технологическими процессами, управление транспортом, медицинская диагностика, техническая диагностика, финансовый менеджмент, биржевое прогнозирование, распознавание образов. Спектр приложений очень широкий – от видеокамер и бытовых стиральных машин до средств наведения ракет ПВО и управления боевыми вертолетами. Практический опыт разработки систем нечеткого логического вывода свидетельствует, что сроки и стоимость их проектирования значительно меньше, чем при использовании традиционного математического аппарата, при этом обеспечивается требуемый уровень робастности и прозрачности моделей.

1 ОСНОВЫ НЕЧЕТКИХ МНОЖЕСТВ

1.1 Понятие нечетких множеств

Рассмотрим основные понятия нечетких систем, привлекая только самые основные математические понятия. Среди многочисленных работ, посвященных нечетким системам [4-12, 26] (приведены только некоторые основные работы на русском языке) зачастую нет единой терминологии. За небольшими исключениями используем терминологию, принятую в [12].

В классической теории множеств некоторый элемент x или принадлежит некоторому множеству A или не принадлежит множеству A . Принадлежность множеству можно описать с помощью характеристической функции $\chi_A(x)$, которая принимает два значения

$$\chi_A(x) = \begin{cases} 1, & \text{если } x \in A, \\ 0, & \text{если } x \notin A. \end{cases}$$

Нечеткое множество (fuzzy set) представляет собой совокупность элементов произвольной природы, относительно которых нельзя с полной определенностью утверждать – принадлежит ли тот или иной элемент рассматриваемой совокупности данному множеству или нет [12]. Другими словами, нечеткое множество отличается от обычного множества тем, что для всех или части его элементов не существует однозначного ответа на вопрос: «Принадлежит или не принадлежит тот или иной элемент рассматриваемому нечеткому множеству?» Можно этот вопрос задать и по-другому: «Обладают или нет его элементы некоторым характеристическим свойством, которое может быть использовано для задания этого нечеткого множества?».

Для определения нечеткого множества используется обобщение понятия характеристической функции – функция принадлежности $\mu_A(x)$, показывающая степень принадлежности элемента x нечеткому множеству A . При этом возможны три случая:

$\mu_A(x) = 1$ означает полную принадлежность элемента x нечеткому множеству A ;

$\mu_A(x)=0$ означает, что элемент x не принадлежит нечеткому множеству A ;

$0 < \mu_A(x) < 1$ означает, астичную принадлежность элемента x нечеткому множеству A .

Тогда нечеткое множество может быть определено как множество упорядоченных пар $A = \{x, \mu_A(x)\}$. Нечеткие множества с конечным количеством элементов $\{x_1, x_1, \dots, x_n\}$ часто записываются в виде

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \dots + \frac{\mu_A(x_n)}{x_n} = \sum_{i=1}^n \frac{\mu_A(x_i)}{x_i}.$$

При этом горизонтальная черта не является знаком деления, а приписывает конкретным элементам определенных степеней принадлежности. Знак «+» означает теоретико-множественное объединение элементов.

Если элементы нечеткого множества принадлежат пространству X с бесконечным количеством элементов, то нечеткое множество иногда символически записывают в виде

$$A = \int_x \frac{\mu_A(x)}{x}.$$

Нечеткость часто путают с вероятностью. Среди неопределенности выделяют стохастическую и лингвистическую неопределенности. Стохастическая неопределенность существуют только для точно описанных событий, которые в будущем могут произойти, а могут не произойти. Когда событие произошло, понятие вероятности его совершения теряет смысл.

Лексическая неопределенность означает неопределенность в описании события независимо от времени их рассмотрения. Например, имеется нечеткое множество «Рост человека» со значениями «низкий», «средний», «высокий». Функция принадлежности значения «высокий», равная, например, 0,25 не означает, что с вероятностью 0,25 в большой выборке будут встречаться высокие люди. Функция принадлежности означает лишь субъективную оценку степени принадлежности понятия «высокий» множеству «Рост человека».

Введем некоторые понятия, связанные с нечеткими множествами.

Пустым нечетким множеством $\emptyset$ называется нечеткое множество, не содержащее ни одного элемента. Формально функция принадлежности пустого множества равна нулю для всех его элементов.

Универсум (универсальное множество) – обычное множество, содержащее в рамках некоторого контекста все возможные элементы. Формально функция принадлежности универсума равно единице для всех элементов множества. Будем обозначать универсум через X .

Носителем нечеткого множества A называется четкое подмножество области определения X , содержащее все элементы, для которых функции принадлежности нечеткого множества A отличны от нуля $S(A) = \text{supp}(A) = \{x: \mu_A(x) > 0, x \in X\}$.

Нечеткое множество называется конечным, если его носитель является конечным множеством. Нечеткое множество, носитель которого не является конечным множеством, называется бесконечным.

1.2 Функция принадлежности

Нечеткие множества могут быть заданы двумя основными способами:

- в форме списка с явным перечислением всех элементов и соответствующих им значений функции принадлежности. Элементы с нулевыми значениями функции принадлежности обычно не указываются в данном списке. Этот способ используется для задания нечетких множеств с конечным дискретным носителем и небольшим числом элементов;

- аналитически и/или графически в форме математического выражения для соответствующей функции принадлежности. Этот способ может быть использован для задания произвольных нечетких множеств как с конечным, так и с бесконечным носителем.

Формальное определение нечеткого множества не накладывает никаких ограничений на выбор конкретной функции принадлежности для его представления. Однако на практике удобно использовать те из них, которые допускают аналитическое представление в виде некоторой простой математической функции. Это упрощает не только соответствующие численные расчеты, но и сокращает вычислительные ресурсы, необходимые для хранения отдельных значений этих функций принадлежности. Необходимость типизации отдельных функций принадлежности также обусловлена наличием реализаций соответствующих функций в рассматриваемых далее инструментальных средствах.

В качестве примера рассмотрим нечеткое множество «Небольшое число» [14]. Табличное задание функции принадлежности для данного примера имеет вид табл. 1

Таблица 1 – Пример табличного задания функции

x	$x_1=a$	$x_2=-0.75a$	$x_3=-0.50a$	$x_4=-0.25a$	$x_5=0$	$x_6=0.25a$	$x_7=0.50a$	$x_8=0.75a$	$x_9=a$
$\mu_A(x)$	0.00	0.25	0.50	0.75	1.00	0.75	0.50	0.25	0.00

Графические формы задания функции принадлежности представлены на рис. 1. Формульное представление функций принадлежности будет рассмотрено ниже.

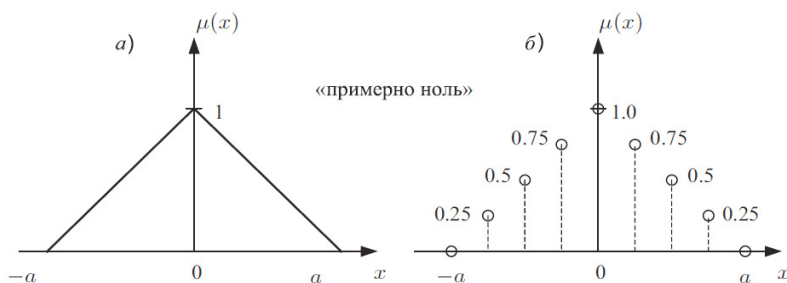


Рис. 1 – Графическая форма задания непрерывной (а) и дискретной функции принадлежности

На практике применяются различные типы функций принадлежности. Широко распространенными являются: кусочно-линейные функции принадлежности, Z-образные функции принадлежности, S-образные функции принадлежности, П-образные функции принадлежности. В [12, 14] можно найти примеры различных типов функций принадлежности и анализ достоинств и недостатков каждого типа.

1.2.1 Кусочно-линейные функции принадлежности

В качестве первого типа функций принадлежности рассмотрим функции, которые, как следует из их названия, состоят из отрезков прямых линий, образуя непрерывную или кусочно-непрерывную функцию. Наиболее характерным примером таких функций являются «треугольная» и «трапециевидная» функции принадлежности. Треугольная функция принадлежности, описываемая выражением [12, 14]:

$$\mu_A(x; a, b, c) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ \frac{c-x}{c-b}, & b < x \leq c, \\ 0, & c < x, \end{cases} \quad (1.1)$$

где a , b и c – некоторые числовые параметры, связанные отношением $a \leq b \leq c$.

График треугольной функции принадлежности представлен на рис. 2¹.

¹ Графики функций принадлежности построены с помощью инструментов, представленных на сайте <http://nrsu.bstu.ru/chap22.html>.

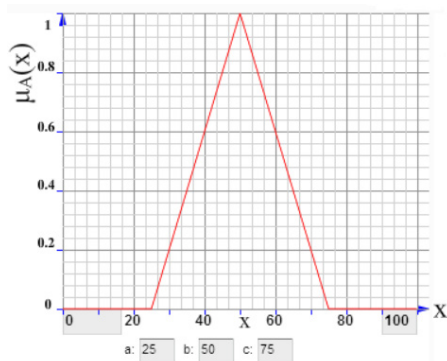


Рис. 2 – График треугольной функции принадлежности

Треугольная функция принадлежности используется, когда известно, что нечеткая переменная ограничивается некоторым диапазоном значений и известно предположение о среднем значении переменной. Тогда a – это минимальное значение переменной, b – среднее, c – максимальное значение.

Трапецевидная функция принадлежности задается выражением [12, 14]:

$$\mu_A(x; a, b, c, d) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ 1, & b < x \leq c, \\ \frac{d-x}{d-c}, & c < x \leq d, \\ 0, & d < x, \end{cases} \quad (1.2)$$

где a, b, c, d – некоторые числовые параметры, связанные отношением $a \leq b \leq c \leq d$.

График трапецевидной функции принадлежности представлен на рис. 3.

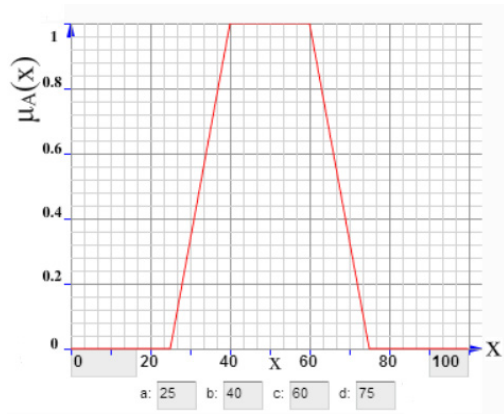


Рис. 3 – График несимметричной трапецевидной функции принадлежности

Трапецевидная функция принадлежности используется, когда известен диапазон изменения нечеткого параметра и диапазон возможного изменения среднего значения.

Эти функции используются для задания таких свойств множеств, которые характеризуют неопределенность типа: «приблизительно равно», «среднее значение», «расположен в интервале», «подобен объекту», «похож на предмет».

1.2.2 Z-образные и S-образные функции принадлежности

Эти функции принадлежности также получили свое название по виду кривых, которые представляют их графики. Первая из функций этой группы называется Z-образной кривой или сплайн-функцией и в общем случае может быть задана аналитически следующим выражением [12, 14]:

$$\mu_A(x; a, b) = \begin{cases} 1, & x < a \\ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{x-a}{b-a} \pi\right), & a \leq x \leq b \\ 0, & x > b \end{cases} \quad (1.3)$$

где a, b – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением: $a < b$.

График трапецевидной функции принадлежности представлен на рис. 4.

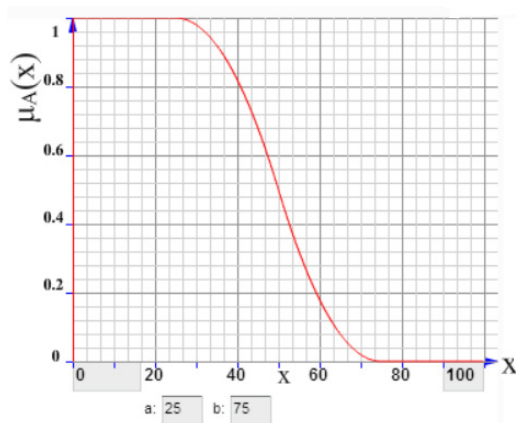


Рис. 4 – График Z-образной функции принадлежности

Эта функция используется для представления таких свойств нечетких множеств, которые характеризуются неопределенностью типа: «малое количество», «небольшое значение», «незначительная величина», «низкая себестоимость продукции», «низкий уровень цен или доходов», «низкая процентная ставка». Общим для всех таких ситуаций является слабая степень проявления того или иного качественного или количественного признака. Особенность нечеткого моделирования при этом заключается в представлении соответствующих нечетких множеств с помощью невозрастающих (монотонно убывающих) функций принадлежности.

Вторая из функций рассматриваемой группы называется S-образной кривой или сплайн функцией и в общем случае может быть задана аналитически следующим выражением [12, 14]:

$$\mu_A(x; a, b) = \begin{cases} 0, & x < a \\ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{x-a}{b-a}\pi\right), & a \leq x \leq b \\ 1, & x > b \end{cases} \quad (1.4)$$

где a, b – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением: $a < b$.

График трапецевидной функции принадлежности представлен на рис. 5.

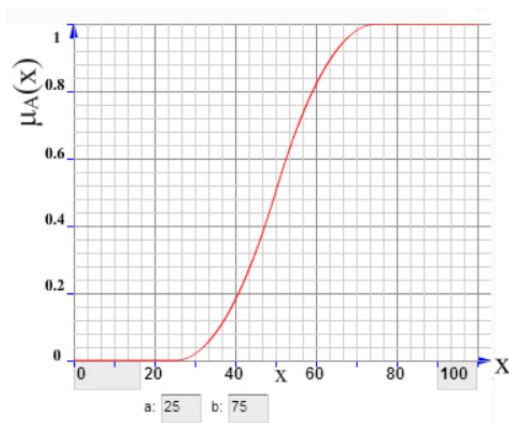


Рис. 5 – График S-образной функции принадлежности

Рассмотренные S-образные функции используются для представления таких нечетких множеств, которые характеризуются неопределенностью типа: «большое количество», «большое значение», «значительная величина», «высокий уровень доходов и цен», «высокая норма прибыли», «высокое качество услуг», «высокий сервис обслуживания». Общим для всех таких ситуаций является высокая степень про явления того или иного качественного или количественного признака. Особенность нечеткого моделирования при этом заключается в представлении соответствующих нечетких множеств с помощью неубывающих (монотонно возрастающих) функций принадлежности.

1.2.3 П-образные функции принадлежности

К данному типу функций принадлежности можно отнести целый класс кривых, которые по своей форме напоминают колокол, сглаженную трапецию или букву «П». П-образная функция, и в общем случае задается аналитически следующим выражением [12, 14]:

$$\mu_A(x; a, \sigma) = e^{-\frac{1}{2} \left(\frac{x-a}{\sigma} \right)^2} \quad (1.5)$$

где σ , a – числовые параметры, при этом квадрат первого из них σ^2 в теории

вероятностей называется дисперсией распределения, а второй параметр a – математическим ожиданием.

График трапециевидной функции принадлежности представлен на рис. 6.

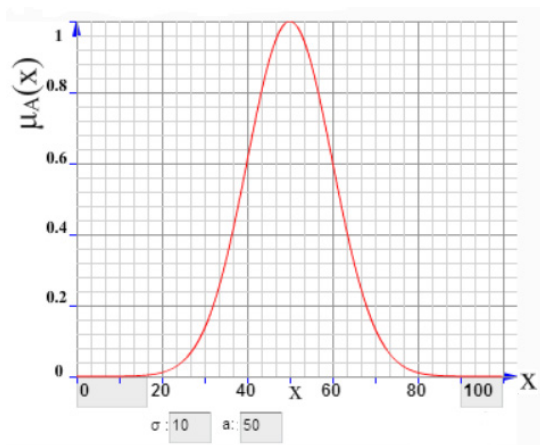


Рис. 6 – Гауссова функция принадлежности

К П-образным функциям относится также так называемая колоколообразная (bellshaped) функция, которая в общем случае задается аналитически следующим выражением:

$$\mu_A(x; a, b, c) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (1.6)$$

где a, b, c – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением: $a < b < c$, причем параметр $b > 0$. Здесь функция $|x|$ обозначает модуль действительного числа.

График трапециевидной функции принадлежности представлен на рис. 7.

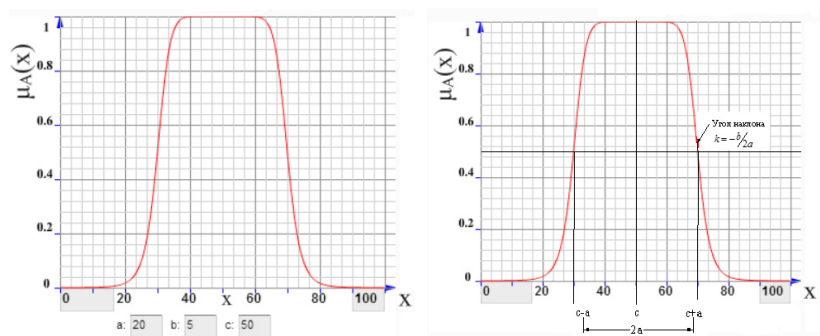


Рис. 7 – Колоколообразная функция принадлежности

П-образные функции принадлежности, использующиеся для задания неопределенностей типа: «приблизительно в пределах от и до», «примерно равно», «около».

1.3 Построение функций принадлежности

Рассмотрим построение функций принадлежности нечетких множеств. Простейшими являются прямые методы построения функции принадлежности. В прямых методах эксперт задает для каждого x значение функции принадлежности $\mu_A(x)$. Как правило, прямые методы построения функций принадлежности используются для таких

свойств, которые могут быть измерены в некоторой количественной шкале. При этом следует учитывать, что теория нечетких множеств не требует абсолютно точного задания функций принадлежности. Зачастую бывает достаточно зафиксировать лишь наиболее характерные значения и вид (тип) функции принадлежности [12].

Косвенные методы построения функции принадлежности используются тогда, когда нет элементарных измеримых свойств, через которые определяется нечеткое множество. Наиболее известным косвенным методом является метод попарных сравнений [19, 20, 23, 24]. Если бы значения искомой функций принадлежности были нам известны и равны $\mu_A(x_i) = w_i$, где $i = 1, 2, \dots, n$, то попарные сравнения можно представить матрицей отношений, где $a_{ij} = w_i / w_j$, а символ «/» означает операцию деления. Матрица A является положительной, так как ее элементы положительны, диагональные элементы матрицы равны единице, матрица является обратно симметричной: $a_{ij} = 1/a_{ji}$. Известно [20], что такая матрица имеет максимальное по модулю действительное собственное число, равное размерности матрицы n , которому соответствует действительный собственный вектор, а все остальные собственные числа равны нулю. Если через w обозначить вектор-столбец $[w_1 \ w_2 \ \dots \ w_n]^T$, то легко проверить, что $Aw = nw$. Учитывая упомянутые свойства матрицы, получаем, что вектор функций принадлежности может быть найден как собственный вектор, соответствующий максимальному собственному числу задачи на собственные значения $Aw = \lambda w$. Реально матрица A формируется на основе субъективных суждений, но известно [20], что при небольших изменениях положительной обратно-симметричной матрицы собственные значения также изменятся незначительно. Поэтому в случае субъективных суждений вектор функций принадлежности может быть найден как собственный вектор, соответствующий максимальному собственному числу матрицы A . Разница между найденным $\lambda_{\max}$ и n служит мерой несогласованности парных сравнений эксперта. И, соответственно, характеризует уровень доверия к полученным результатам. Чем больше это отличие, тем меньше доверие. Практически вектор w нормализуют, то есть делят на максимальный компонент.

Практически при построении матрицы парных сравнений эксперт рассматривает элементы u_i , $i = 1, 2, \dots, n$ нечеткого множества и оценивает преимущество одного элемента над другим по отношению к свойству нечеткого множества. Тогда элемент a_{ij} матрицы – это уровень преимущества элемента u_i над u_j , определяемый по девятибалльной шкале Саати [19]:

1 – если отсутствует преимущество элемента u_i над элементом u_j ;

3 – если имеется слабое преимущество u_i над u_j ;

5 – если имеется существенное преимущество u_i над u_j ;

7 – если имеется явное преимущество u_i над u_j ;

9 – если имеется абсолютное преимущество u_i над u_j ;

2, 4, 6, 8 – промежуточные сравнительные оценки: 2 – почти слабое преимущество, 4 – почти существенное преимущество, 6 – почти явное преимущество, 8 – почти абсолютное преимущество.

Причем в матрице парных сравнений соблюдается условие $a_{ij} = 1/a_{ji}$.

Практически функции принадлежности удобно задавать в параметрической форме: выбирается форма функции принадлежности [12, 14, 23, 24] и из особенностей нечеткого множества выбирают параметры функции.

1.4 Основные характеристики нечетких множеств

Рассмотрев понятие функции принадлежности, предложим знакомство с основными характеристиками нечетких множеств [12, 14].

Множество α – уровня – это обобщение понятия носителя нечеткого множества. Это обычное множество A_α , удовлетворяющее условию $A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\}$, где α – некоторое действительное число из интервала $[0, 1]$.

Высота нечеткого множества определяется с помощью точной верхней грани следующим образом $h_A = \sup_{x \in X} (\mu_A(x))$.

Нечеткое множество называется нормальным, если максимальное значение его функции принадлежности равно 1 и при

этом существуют элементы множества, функция принадлежности которых равна 1. Формально это означает, что для нормального нечеткого множества необходимо выполнение следующего условия: $\mu_A(x) = 1, (\exists x \in X)$.

Нечеткое множество называется субнормальным, максимальное значение его функции принадлежности меньше 1.

Унимодальное нечеткое множество. Нечеткое множество называется унимодальным (строго унимодальным), если его функция принадлежности $\mu_A(x)$ является унимодальной (строго унимодальной). Функция принадлежности $\mu_A(x)$ называется унимодальной (строго унимодальной), если она унимодальная (строго унимодальна) на носителе соответствующего нечеткого множества.

Ядром нечеткого множества A называется обычное, элементы которого удовлетворяют условию:

$$C(A) = \text{core}(A) = \{x \in X \mid \mu_A(x) = 1\}.$$

Мощность (кардинальное число) дискретного нечеткого множества определяется как сумма функций принадлежности всех его элементов

$$\|A\| = \text{card}(A) = \sum_{x \in S(A)} \mu_A(x),$$

где $S(A)$ – носитель нечеткого множества.

Мощность непрерывного нечеткого множества вычисляется интегрированием функции принадлежности

$$\|A\| = \text{card}(A) = \int_{x \in S(A)} \mu_A(x) dx.$$

Границы нечеткого множества называются такие элементы универсума, для которых значения функции принадлежности отличны от 0 и 1. Другими словами, границы нечеткого множества $A_\alpha = \{x, \mu_A(x)\}$ включают те и только те элементы универсума $x \in X$ для которых выполняется условие: $0 < \mu_A(x) < 1$.

Точки перехода нечеткого множества. Элементы нечеткого множества $y \in A$, для которых выполняется условие: $\mu_A(x) = 0.5$, называются точка перехода этого нечеткого множества.

Ближайшее четкое множество. Часто оказывается полезным понятие четкого множества A , ближайшего к нечеткому множеству A . Характеристическая функция такого множества может быть определена следующим выражением:

Для характеристики нечетких множеств используют также понятие выпуклости, которое ассоциируется с соответствующим графическим изображением функции принадлежности.

Выпуклое нечеткое множество. Нечеткое множество $A_a = \{x, \mu_A(x)\}$ с универсумом X называют выпуклым, если его функция принадлежности $\mu_A(x)$ удовлетворяет следующему неравенству: $\mu_A(x) \geq \min\{\mu_A(a), \mu_A(b)\}$ для любых значений $x, a, b \in X$, при которых $a < x < b$ и $a \neq b$.

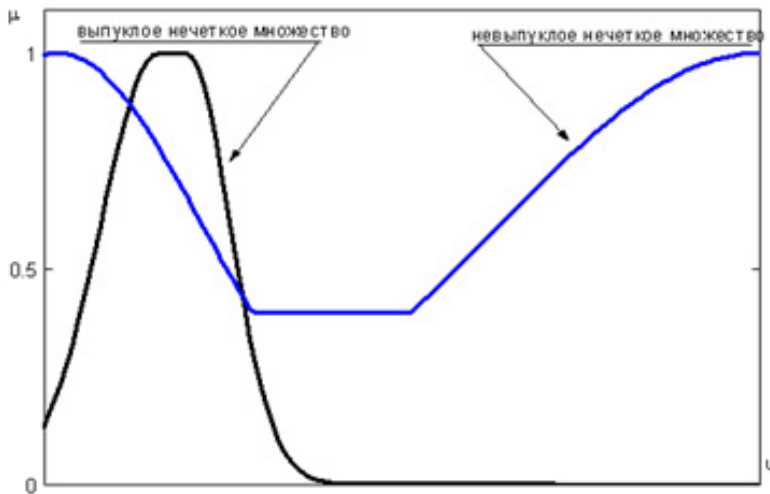


Рис. 8 – Графики функций принадлежности выпуклого и невыпуклого нечеткого множества

1.5 Операции над нечеткими множествами

Операции над нечеткими множествами имеют три особенности [12]: во-первых, все операции над нечеткими множествами являются обобщениями операций над классическими множествами

ми. Так как возможны разные варианты обобщения, это приводит к различным вариантам операций над нечеткими множествами. Во-вторых, выполнение операций над нечеткими множествами возможно, когда нечеткие множества определены на одном и том же универсуме. В-третьих, операции над нечеткими множествами сводятся к операциям над их функциями принадлежности. С учетом этих особенностей рассмотрим основные операции над нечеткими множествами. Подробно эти операции освещены в [4, 12, 14]. Следует отметить, что не все аксиомы теории множеств выполняются для операций над нечеткими множествами [4, 14].

Пусть A и B нечёткие множества, заданные на универсальном множестве X . Нечеткие множества A и B считаются равными, если их функции принадлежности принимают равные значения на всем универсуме $\mu_A(x) = \mu_B(x) \quad \forall x \in X$.

Равенство нечетких множеств записывается как $A = B$.

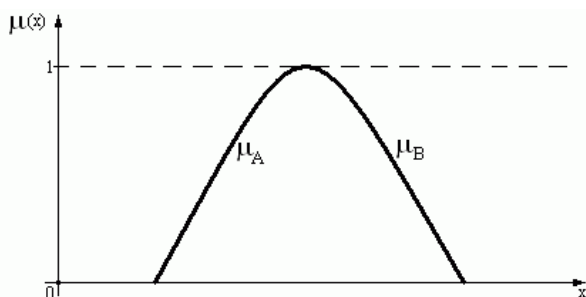


Рис. 9 – Равенство нечетких множеств

Нечеткое множество A называется несобственным подмножеством нечеткого множества B (записывается как $A \subseteq B$), если $\mu_A(x) \leq \mu_B(x) \quad \forall x \in X$.

Нечеткое множество A называется собственным подмножеством нечеткого множества B (записывается как $A \subset B$), если $\mu_A(x) < \mu_B(x) \quad \forall x \in X$.

Операции объединения и пересечения нечетких множеств могут быть определены различными способами в зависимости от особенностей конкретных задач. Эти операции удобно описывать, пользуясь понятиями треугольной нормы и треугольной конормы [4, 12, 14].

Треугольная норма (Т-норма, t-норма) – это функция, действующая на интервале $[0, 1]$ и удовлетворяющая следующим аксиомам:

- $T(x, 0) = 0; \quad T(x, 1) = 1$ ограниченность;
- $T(x, y) = T(y, x)$ коммутативность;
- $T(x, T(y, z)) = T(T(x, y), z)$ ассоциативность;
- $T(x, y) \leq T(x, z)$, если $y \leq z$ монотонность.

Треугольная конорма (Т-конорма, s-норма) – это функция, действующая на интервале $[0, 1]$ и удовлетворяющая следующим аксиомам:

- $S(x, 0) = x; \quad S(x, 1) = 1$ ограниченность;
- $S(x, y) = S(y, x)$ коммутативность;
- $S(x, S(y, z)) = S(S(x, y), z)$ ассоциативность;
- $S(x, y) \leq S(x, z)$, если $y \leq z$ монотонность.

Эти две треугольные нормы отличаются только аксиомами ограниченности.

Простейшими примерами t-норм и s-норм являются следующие [5, 25, 27]:

$T(x, y) = \min(x, y)$ – минимум Заде;

$S(x, y) = \max(x, y)$ – максимум Заде;

$T(x, y) = \max(x + y - 1, 0)$ – t-норма Лукасевича;

$S(x, y) = \min(x + y, 1)$ – s-норма Лукасевича (ограниченная сумма);

$T(x, y) = x \cdot y$ – произведение;

$S(x, y) = x + y - x \cdot y$ – вероятностная сумма

В нечетком выводе используются также взвешенные t-нормы и s-нормы:

$$T(x, y, w_1, w_2) = T(1 - w_1(1 - x), 1 - w_2(1 - y)),$$

$$S(x, y, w_1, w_2) = S(w_1x, w_2y).$$

Возможно обобщение треугольных норм на случай нескольких переменных.

В общем виде пересечением нечётких множеств A и B , заданных на одном универсуме X , называется третье нечеткое множество C , заданное на этом же универсуме, функция принадлежности которого определяется следующим образом

$$\mu_C(x) = \mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)),$$

где T – t-норма.

Различные реализации t-нормы приводят к различным реализациям операции пересечения нечетких множеств [22, 25].

Операция min-пересечения, или $\wedge$ -пересечения $A \cap B$ определяется функцией принадлежности (рис. 10):

$$\mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x) = \min(\mu_A(x), \mu_B(x)) \quad \forall x \in X \quad (1.7)$$

Операция алгебраического пересечения, или prod-пересечения $A \bullet B$ определяется функцией принадлежности

$$\mu_{A \bullet B}(x) = \text{prod}(\mu_A(x), \mu_B(x)) = \mu_A(x) \cdot \mu_B(x) \quad \forall x \in X.$$

Результат операции пересечения двух бесконечных нечетких множеств удобно изобразить графически (рис. 10).

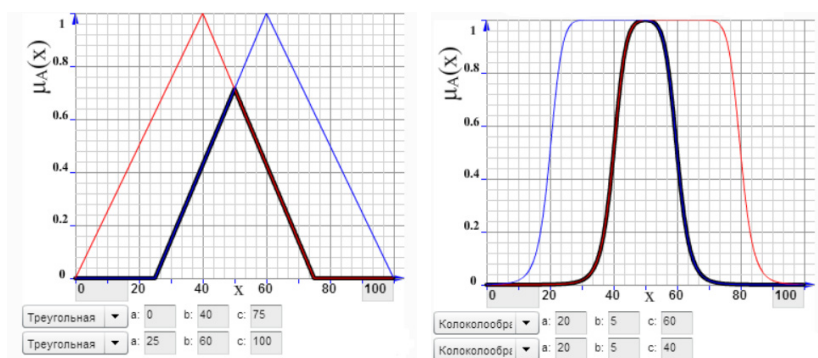


Рис. 10 – Графическое представление операции min-пересечения треугольных и П-образных функций принадлежности

Операция граничного пересечения $A^\circ B$ определяется функцией принадлежности $\mu_{A^\circ B}(x) = \max\{\mu_A(x) + \mu_B(x) - 1, 0\} \quad \forall x \in \mathbf{X}$.

Операция драстического пересечения (от англ. drastic – решительный, радикальный) $A\Delta B$ определяется функцией принадлежности

$$\mu_{A\Delta B}(x) = \begin{cases} \mu_A(x), & \text{если } \mu_B(x)=1, \\ \mu_B(x), & \text{если } \mu_A(x)=1, \\ 0, & \text{если } \mu_A(x)<1, \mu_B(x)<1, \end{cases} \quad \forall x \in \mathbf{X}.$$

Операция пересечения Ягера (Yager R.) [4] определяется функцией принадлежности

$$\mu_{A\cap B}(x) = 1 - \min\left\{1, \left[(1 - \mu_A(x))^p + (1 - \mu_B(x))^p\right]^{\frac{1}{p}}\right\} \quad p \geq 1, \quad \forall x \in \mathbf{X}.$$

В общем виде объединением нечётких множеств A и B , заданных на одном универсуме X , называется третье нечеткое множество C , заданное на этом же универсуме, функция принадлежности которого определяется следующим образом $\mu_C(x) = \mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x))$, где S – s-норма (t-конорма).

Различные реализации s-нормы приводят к различным реализациям операции пересечения нечетких множеств [12, 14].

Операция max-объединения, или $\vee$ -объединения $A \cup B$ определяется функцией принадлежности

$$\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \max(\mu_A(x), \mu_B(x)) \quad \forall x \in \mathbf{X} \quad (1.8)$$

Операция алгебраического объединения $A + B$ определяется функцией принадлежности

$$\mu_{A+B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x) \quad \forall x \in \mathbf{X}.$$

Результат операции объединения двух бесконечных нечетких множеств удобно изобразить графически (рис. 11).

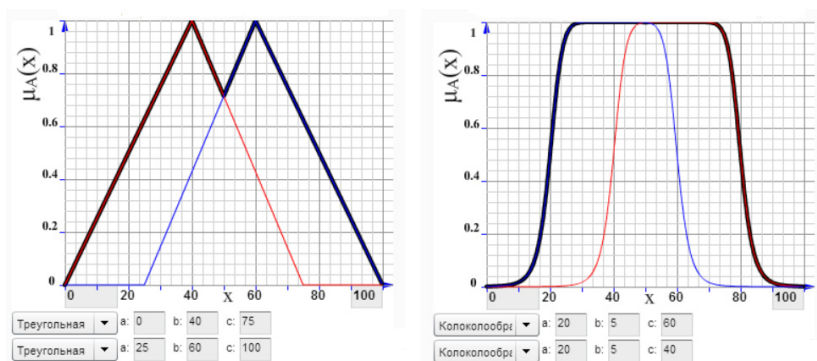


Рис. 11 – Графическое представление операции max-объединения для треугольных и П-образных функций принадлежности

Операция граничного объединения $A \oplus B$ определяется функцией принадлежности: $\mu_{A \oplus B}(x) = \min\{\mu_A(x) \cdot \mu_B(x), 1\} \quad \forall x \in X$.

Операция драстического объединения $A \nabla B$ определяется функцией принадлежности.

$$\mu_{A \nabla B}(x) = \begin{cases} \mu_A(x), & \text{если } \mu_B(x) = 0, \\ \mu_B(x), & \text{если } \mu_A(x) = 0, \\ 1, & \text{если } \mu_A(x) < 1, \mu_B(x) < 1, \end{cases} \quad \forall x \in X$$

Операция объединения Ягера определяется функцией принадлежности

$$\mu_{A \cup B}(x) = \min \left\{ 1, \left[(\mu_A(x))^p + (\mu_B(x))^p \right]^{\frac{1}{p}} \right\} \quad p \geq 1, \quad \forall x \in X.$$

Разностью нечётких множеств A и B , заданных на одном универсуме X , называется третье нечеткое множество $C = A \setminus B$, заданное на этом же универсуме, функция принадлежности которого определяется следующим образом

$$\mu_C(x) = \mu_{A \setminus B}(x) = \max\{\mu_A(x) - \mu_B(x), 0\} \quad \forall x \in X.$$

Операция разности не является коммутативной, то есть $A \setminus B \neq B \setminus A$. По аналогии с обычными множествами иногда применяют операцию симметричной разности $C = A - B$, функция принадлежности которого определяется следующим образом $\mu_C(x) = \mu_{A-B}(x) = |\mu_A(x) - \mu_B(x)| \quad \forall x \in X$.

Результат операции разности двух бесконечных нечетких множеств удобно изобразить графически (рис. 12).

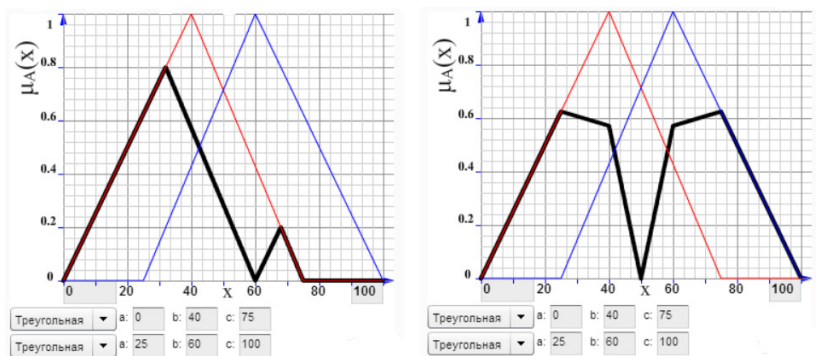


Рис. 12 – Графическое представление операции разности и симметричной разности для треугольных функций принадлежности

Операция дополнения(инверсии) нечеткого множества A (обозначается через $\bar{A}$) определяется функцией принадлежности $\mu_{\bar{A}}(x) = \mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad \forall x \in X$.

Приведенную операцию называют еще инверсией Заде.

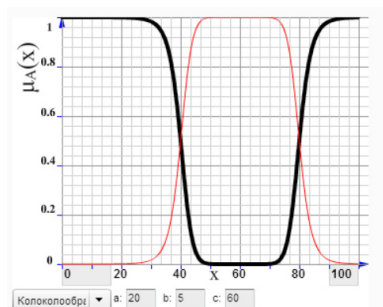


Рис. 13 – Графическое представление операции дополнения для П-образных функций принадлежности

Известны также инверсия Ягера $\mu_c(x) = (1 - x^p)^{\frac{1}{p}}$, $p > 0$ и инверсия Сугено $\mu_c(x) = \frac{1-x}{1+px}$, $p > -1$.

Если A – произвольное нечеткое множество, заданное на универсуме X , a – положительное число, причем $a \cdot h_A \leq 1$, где h_A – высота множества, то результат умножения нечеткого множества A на число a определяется как нечеткое множество $C = a \cdot A$, заданное на этом же универсуме X , функция принадлежности которого равна $\mu_C(x) = \mu_{a \cdot A}(x) = a \mu_A(x) \quad \forall x \in X$.

Результатом операции возведения нечеткого множества A , заданного на универсуме X , в степень k , где k – положительное действительное число, является нечеткое множество C , функция принадлежности которого определяется по формуле: $\mu_C(x) = (\mu_A(x))^k \quad \forall x \in X$. Операцию возведения в степень обозначают как A^k .

Для создания нечетких множеств, являющихся производными от некоторых ранее заданных, используются лингвистические модификаторы. Например, если имеется нечеткое множество «холодный», то на его основе с помощью лингвистических модификаторов можно получить множества «очень холодный» или «более или менее холодный».

Существуют три основных лингвистических модификатора (оператора):

- оператор концентрирования;
- оператор растяжения;
- оператор повышения/понижения контрастности.

Операция концентрирования, примененная к нечеткому множеству A , даст новое нечеткое множество C , функция принадлежности которого определяется формулой:

$$\mu_C(x) = \mu_{\text{con}(A)}(x) = (\mu_A(x))^2 \quad \forall x \in X.$$

Операция концентрирования обозначается через $C = \text{con}(A)$ – от англ. concentration – концентрирование.

Операция растяжения, обозначаемая через $\text{dil}(A)$ (от англ.

dilatation – растяжение), дает в результате нечеткое множество C , функция принадлежности которого определяется формулой $\mu_C(x) = \mu_{\text{dil}(A)}(x) = (\mu_A(x))^{0.5} \quad \forall x \in X$.

Операция концентрирования приводит к уменьшению степени принадлежности нечеткого множества, а операция растяжения – к увеличению степени принадлежности.

Результат действия операторов концентрирования и растяжения на треугольную функцию принадлежности представлен на рис. 14.

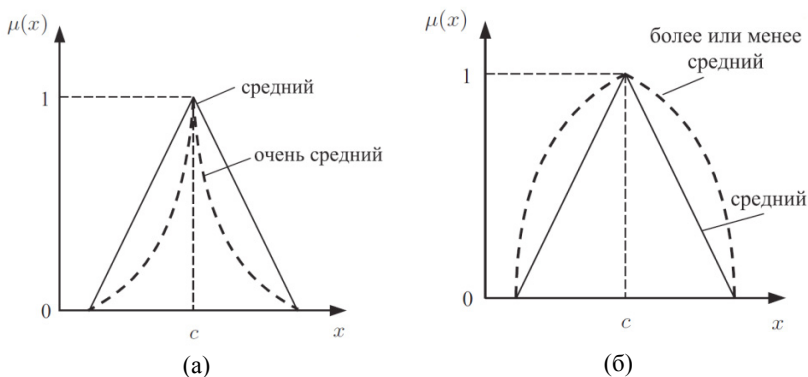


Рис. 14 – Графическое представление действия операции концентрирования (а) и растяжения (б) для треугольной функции принадлежности

Границы между нечеткими множествами являются размытыми. Оператор повышения контрастности (int – от англ. intensification – усиление, повышение) приводит нечеткие множества к более четкому виду. Изменяя углы наклона ветвей функции принадлежности, он позволяет более четко выделять границы перехода от одного нечеткого множества к другому. Операция повышения контрастности, примененная к нечеткому множеству A , дает в результате нечеткое множество C , функция принадлежности которого определяется формулой

$$\mu_C(x) = \mu_{\text{int}(A)}(x) = \begin{cases} 2^{n-1} (\mu_A(x))^n & \text{для } \mu_A(x) < 0,5, \forall x \in X \\ 1 - 2^{n-1} (1 - \mu_A(x))^n & \text{для } \mu_A(x) \geq 0,5, \forall x \in X \end{cases},$$

где n – целое число, обычно $n = 2$.

При стремлении n к бесконечности функция принадлежности принимает прямоугольную форму, а нечеткое множество превращается в обычное четкое множество с четкими границами.

Операция понижения контрастности (обозначается blr – от англ. *blurring* – размывание), примененная к нечеткому множеству A , дает в результате нечеткое множество C , функция принадлежности которого определяется формулой

$$\mu_C(x) = \mu_{\text{blr}(A)}(x) = \begin{cases} 1 - 2^{n-1} (1 - \mu_A(x))^n & \text{для } \mu_A(x) < 0,5, \forall x \in X \\ 2^{n-1} (\mu_A(x))^n & \text{для } \mu_A(x) \geq 0,5, \forall x \in X \end{cases},$$

где n – целое число, обычно $n = 2$.

При стремлении n к бесконечности нечеткое множество преобразуется в точку, которая совпадает со средним значением ядра нечеткого множества (с модальным значением нечеткого множества).

Графическое представление действия операторов повышения и понижения контрастности на функцию принадлежности представлен на рис. 15.

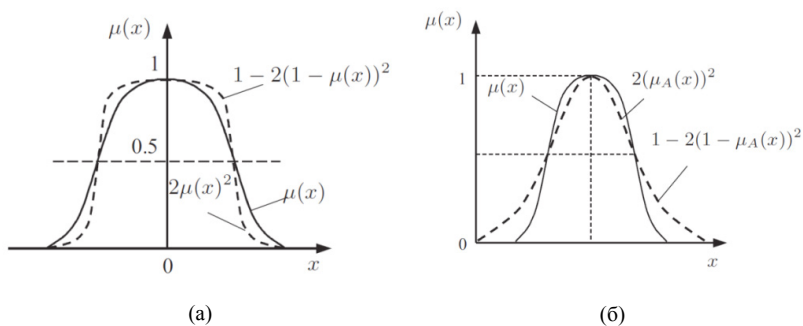


Рис. 15 – Графическое представление действия операции повышения (а) и понижения (б) контрастности для $m = 2$

1.6 Нечеткие отношения

Рассмотренные нечеткие множества имели в качестве области определения одномерное пространство. В нечетких системах зачастую областью определения нечетких множеств являются многомерные пространства. При этом особый интерес представляют множества с двумерной областью определения. Нечеткие множества на много мерных областях определения задаются как нечеткие отношения. Нечетким k -арным отношением, заданным на множествах (универсумах) $X_1, X_2, \dots, X_k$, называется нечеткое множество R , определенное на декартовом произведении $X_1 \times X_2 \times \dots \times X_k$. Другими словами, нечеткое отношение – это множество пар $R = \{ \langle x_1, x_2, \dots, x_k \rangle, \mu_R(\langle x_1, x_2, \dots, x_k \rangle) \}$, где $\langle x_1, x_2, \dots, x_k \rangle$ – кортеж из k элементов, каждый из которых выбирается из своего универсума: $x_1 \in X_1, x_2 \in X_2, \dots, x_k \in X_k$, $\mu_R(\langle x_1, x_2, \dots, x_k \rangle)$ – функция принадлежности нечеткого отношения, которая определяется как отображение $\mu_R : x_1 \in X_1, x_2 \in X_2, \dots, x_k \in X_k \rightarrow [0, 1]$. Напомним, что декартовым произведением конечного числа множеств $X_1 \times X_2 \times \dots \times X_k$ называется множество всех кортежей $\langle x_1, x_2, \dots, x_k \rangle$ длины k , составленных из элементов этих множеств: $x_1 \in X_1, x_2 \in X_2, \dots, x_k \in X_k$.

Отношение (четкое или нечеткое), построенное на основе двух множеств, называется бинарным, на основе трех множеств – тернарным, на основе k множеств – k -арным. Отношение может быть задано на одном универсуме. Например, бинарное нечеткое отношение, заданное на универсуме X , имеет вид $R = \{ \langle x_i, x_j \rangle, \mu(\langle x_i, x_j \rangle) \}$, при этом $x_i \in X$ и $x_j \in X$.

Нечеткие отношения могут быть заданы различными способами:

- в форме списка с явным перечислением всех кортежей отношения и соответствующих им функций принадлежности. Это способ применим для отношений с небольшим числом кортежей;
- аналитически в форме выражения для функции принадлежности;
- для бинарных отношений дополнительно могут быть использованы следующие способы:

графически в виде некоторой поверхности. При этом независимыми переменными являются значения универсумов X_1 и X_2 ,

а третья координата представляет собой соответствующее значение функции принадлежности;

в виде матрицы конечного бинарного отношения, строки которой соответствуют первым элементам кортежей, а столбцы – вторым элементам кортежей. Элементами матрицы являются соответствующие значения функции принадлежности;

в виде нечеткого графа, вершины которого соответствуют элементам универсумов X_1 и X_2 , а дуга, соединяющая вершины $x_i \in X$ и $x_j \in X$, показывает, что кортеж $\langle x_i, x_j \rangle$ входит в отношение. Около каждой дуги указывается функция принадлежности соответствующего кортежа.

Рассмотрим пример. Пусть даны два множества: $X_1 = \{a_1, a_2, \dots, a_6\}$ – множество пользователей Internet и $X_2 = \{b_1, b_2, \dots, b_4\}$ – множество Internet. Декартово произведение $X_1 \times X_2$ в матричной форме имеет вид

$$X_1 \times X_2 = \begin{bmatrix} a_1b_1 & a_1b_2 & a_1b_3 & a_1b_4 \\ a_2b_1 & a_2b_2 & a_2b_3 & a_2b_4 \\ a_3b_1 & a_3b_2 & a_3b_3 & a_3b_4 \\ a_4b_1 & a_4b_2 & a_4b_3 & a_4b_4 \\ a_5b_1 & a_5b_2 & a_5b_3 & a_5b_4 \\ a_6b_1 & a_6b_2 & a_6b_3 & a_6b_4 \end{bmatrix}.$$

Примером нечеткого отношения, заданного на $X_1 \times X_2$, является отношение «Пользоваться браузером». В матричной форме каждый элемент такого отношения функцию принадлежности, в нашем примере – степень популярности у каждого пользователя каждого браузера.

Так как нечеткое отношение является нечетким множеством, то сохраняются определения над отношениями. Например, [12], пересечением двух нечетких отношений

$$Q = \{ \langle x_1, x_2, \dots, x_k \rangle, \mu_Q(\langle x_1, x_2, \dots, x_k \rangle) \}$$

и $R = \{ \langle x_1, x_2, \dots, x_k \rangle, \mu_R(\langle x_1, x_2, \dots, x_k \rangle) \}$ называется нечеткое от-

ношение S , заданное на этом же декартовом произведении универсумов $X_1 \times X_2 \times \dots \times X_k$, функция принадлежности которого определяется по формуле

$$\mu_S(\langle x_1, x_2, \dots, x_k \rangle) = \min \{ \mu_Q(\langle x_1, x_2, \dots, x_k \rangle), \mu_R(\langle x_1, x_2, \dots, x_k \rangle) \},$$

$$\forall \langle x_1, x_2, \dots, x_k \rangle \in \mathbf{X}_1 \times \mathbf{X}_2 \times \dots \times \mathbf{X}_k.$$

Объединением двух нечетких отношений Q и R называется нечеткое отношение S , заданное на этом же декартовом произведении универсумов $X_1 \times X_2 \times \dots \times X_k$, функция принадлежности которого определяется по формуле

$$\mu_S(\langle x_1, x_2, \dots, x_k \rangle) = \max \{ \mu_Q(\langle x_1, x_2, \dots, x_k \rangle), \mu_R(\langle x_1, x_2, \dots, x_k \rangle) \},$$

$$\forall \langle x_1, x_2, \dots, x_k \rangle \in \mathbf{X}_1 \times \mathbf{X}_2 \times \dots \times \mathbf{X}_k.$$

Операции над нечеткими отношениями могут быть определены с помощью t-норм и s-норм.

Зачастую требуется выполнить операции над нечеткими множествами, заданными на разных универсумах X_1 и X_2 . Такие множества с помощью операции цилиндрического продолжения (расширения) предварительно приводят к нечетким отношениям, заданным на декартовом произведении X_1 и X_2 , а затем выполняют операцию над отношениями, заданными на одном универсуме. Цилиндрическое продолжение определяется следующим образом [14]. Пусть X_1 и X_2 – четкие множества (функции принадлежности элементов которых равны единице), а A – нечеткое множество, заданное на X_1 . Цилиндрическим продолжением A^* множества A на область определения $X_1 \times X_2$ называется отношение, представляющее собой декартово произведение $A \times \mathbf{X}_2$. Функция принадлежности A^* имеет вид

$$\mu_{A^*}(x_1, x_2) = \min \{ \mu_A(x_1), \mu_{\mathbf{X}}(x_2) \} = \min \{ \mu_A(x_1), 1 \} = \mu_A(x_1)$$

$$\forall \langle x_1, x_2 \rangle \in \mathbf{X}_1 \times \mathbf{X}_2$$

В матричном представлении отношения A^* столбцы матрицы функций принадлежности будут одинаковыми и равными значениям функции принадлежности нечеткого множества A^* .

Используется также операция проекции, противоположная цилиндрическому продолжению. Если A – нечеткое отношение с областью определения $X_1 \times X_2$, то проекцией этого отношения на множество X_1 называется нечеткое множество A^* , функция принадлежности которого имеет вид $\mu_{A^*}(x_1, x_2) = \max_{x_2 \in X_2} \mu_A(x_1, x_2)$.

Рассмотрим также операцию композиции (комбинации) бинарных отношений. Пусть бинарное нечеткое отношение $Q = \{\langle x_i, x_j \rangle, \mu_Q(\langle x_i, x_j \rangle)\}$ задано на декартовом произведении $X_1 \times X_2$, а нечеткое отношение $R = \{\langle x_j, x_k \rangle, \mu_R(\langle x_j, x_k \rangle)\}$ задано на декартовом произведении $X_2 \times X_3$. Композицией бинарных отношений Q и R называется нечеткое бинарное отношение, обозначаемое через $Q \otimes R$ (используется также обозначение $Q \circ R$ [18]), заданное на декартовом произведении $X_1 \times X_3$ функция принадлежности которого определяется формулой $\mu_{Q \otimes R}(\langle x_i, x_k \rangle) = \max_{x_j \in X_2} \left\{ \min \left\{ \mu_Q(\langle x_i, x_j \rangle), \mu_R(\langle x_j, x_k \rangle) \right\} \right\}$, $\forall \langle x_i, x_k \rangle \in X_1 \times X_3$.

Такую композицию называют еще (max-min)-композицией или максиминной сверткой нечетких отношений.

Вместо операции $\min \left\{ \mu_Q(\langle x_i, x_j \rangle), \mu_R(\langle x_j, x_k \rangle) \right\}$ в композиции можно использовать операцию

$$\text{prod} \left(\mu_Q(\langle x_i, x_j \rangle), \mu_R(\langle x_j, x_k \rangle) \right) = \mu_Q(\langle x_i, x_j \rangle) \cdot \mu_R(\langle x_j, x_k \rangle).$$

Важную роль в нечетком выводе играет композиция (комбинация) нечеткого множества $A \subseteq X$ и нечеткого отношения $R \subseteq X \times Y$, которая обозначается $A \otimes R$ (или $A \circ R$) и определяется как нечеткое множество $B \subseteq Y$ и функцией принадлежности [18]

$$\mu_{A \otimes R}(y) = \max_{x \in X} \left\{ \min \left\{ \mu_A(x), \mu_R(\langle x, y \rangle) \right\} \right\}. \quad (1.9)$$

Вместо операции $\min$ в композиции можно использовать операцию prod . Более подробно с операциями над нечеткими отношениями можно познакомиться в [4, 12, 14, 18].

1.7 Лингвистическая переменная.

Нечеткие величины, числа и интервалы

Нечеткая переменная [12] определяется как кортеж $\langle \alpha, X, A \rangle$, где α – имя нечеткой переменной, X – универсум, являющийся областью определения нечеткой переменной, $A = \{x, \mu_A(x)\}$ – нечеткое множество, определяющее возможные значения нечеткой переменной.

Например, нечеткая переменная «температура больного» определяется областью возможных значений и функцией принадлежности.

Лингвистическая переменная отличается от числовой переменной тем, что ее значениями являются не числа, а слова или словосочетания естественного языка, что более естественно для человека, чем числа.

Лингвистическая переменная определяется как кортеж $\langle \alpha, T, X, G, M \rangle$, где α – имя лингвистической переменной; T – базовое терм-множество лингвистической переменной, или множество её значений (термов), каждое которых представляет собой наименование отдельной нечеткой переменной; X – область определения (универсум) нечетких переменных, входящих в терм-множество; G – синтаксическое правило, порождающее из базового терм-множества T множество новых значений лингвистической переменной, имеющих смысл для конкретной ситуации; M – семантическое правило, которое ставит в соответствие каждому значению лингвистической переменной, получаемому с помощью правила G , некоторое нечеткое множество.

Например, для лингвистической переменной с именем «температура тела человека» терм-множество имеет вид:

$$T = \{\text{"пониженная"}, \text{"нормальная"}, \text{"повышенная"}, \text{"высокая"}\}.$$

Область определения X – это диапазон изменения температуры тела человека. G – правила формирования новых термов с помощью логических связок «И», «ИЛИ» и модификаторов типа «НЕ», «очень», «слегка». Например, «очень высокая температу-

ра». Модификатор «НЕ» задается с помощью операции дополнения (инверсии), модификатор «очень» задается с помощью операции концентрирования, модификатор «слегка» – операцией растяжения (см. раздел 1.5). Правила M задают на X нечеткие переменные – термы из множеств T и правил G . Правила сводятся к формированию функций принадлежности с учетом логических связей и модификаторов. Лингвистические переменные кроме словесных значений могут принимать и численные значения.

Нечеткой величиной называется произвольное нечеткое множество $A = \{x, \mu_A(x)\}$, заданное на множестве действительных чисел R . Если в качестве универсума взять подмножество неотрицательных действительных чисел R_+ , то получим определение не отрицательной нечеткой величины A_+ .

Нечетким интервалом называется нечеткая величина с выпуклой функцией принадлежности. Напомним, что функция $f(x)$, определенная на некотором интервале, выпукла, если для любых двух значений аргумента x_1, x_2 и любого числа $\alpha \in [0, 1]$ выполняется неравенство $f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)$

Если это неравенство является строгим для всех $\alpha \in [0, 1]$, функция называется строго выпуклой; если выполняется обратное неравенство, функция называется вогнутой, или выпуклой вверх. График выпуклой дифференцируемой функции одной переменной лежит не ниже касательной, проведенной к этому графику в любой точке промежутка выпуклости.

Нечетким числом называется такая нечеткая величина, функция принадлежности которой является выпуклой и унимодальной. Напомним, что функция $f(x)$ называется унимодальной, если на заданном отрезке она имеет единственный максимум (минимум) $f(x^*)$, причем слева от x^* функция монотонно возрастает (убывает в случае минимума), а справа – монотонно убывает (возрастает). График функции принадлежности нечеткого числа имеет единственный максимум, обычно равный единице. График функции принадлежности нечеткого интервала принимает максимальное значение на некотором интервале.

Так как нечеткие числа и интервалы представляют собой нечеткие множества, то для них справедливы все свойства и опе-

рации, определенные для нечетких множеств. Дополнительно выделяют понятия нечеткий нуль – нечеткое число с модальным значением (модой), равным нулю, и положительное (или отрицательное) нечеткое число, имеющее строго положительный (или строго отрицательный) носитель.

При выполнении операций над нечеткими числами и интервалами используется принцип обобщения Заде [6, 16, 24]. Пусть $y = f(x_1, x_2, \dots, x_n)$ – четкая функция n переменных, заданных нечеткими числами $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$. Тогда значением функции $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ называется нечеткое число $\tilde{y}$ с функцией принадлежности

$$\mu_{\tilde{y}}(y^*) = \sup_{\substack{y^* = f(x_1^*, x_2^*, \dots, x_n^*) \\ x_i^* \in \text{supp}(\tilde{x}_i), \quad i=1, n}} \left[\min_{i=1, n} (\mu_{\tilde{x}_i}(x_i^*)) \right], \quad (1.10)$$

где функция $\sup$ означает верхнюю грань, а supp – носитель нечеткого множества.

Вычисления по приведенной формуле достаточно громоздки и для нечетких чисел, заданных на дискретных носителях могут быть реализованы следующим образом [16, 24]:

Шаг 1. Зафиксировать значение функции $y = y^*$.

Шаг 2. Найти все множества аргументов $\{x_{1j}^*, x_{2j}^*, \dots, x_{nj}^*\}$, $j = 1, P$, удовлетворяющие условиям

$$y^* = f(x_{1j}^*, x_{2j}^*, \dots, x_{nj}^*) \text{ и } x_{ij}^* \in \text{supp}(\tilde{x}_i), \quad i = 1, n.$$

Шаг 3. Значение функции принадлежности элемента y^* нечеткому числу $\tilde{y}$ вычислить по формуле

$$\mu_{\tilde{y}}(y^*) = \max_{j=1, \dots, P} \min_{i=1, n} (\mu_{\tilde{x}_i}(x_{ij}^*)).$$

Шаг 4. Если рассмотрены все элементы y^* , то перейти к шагу 5, иначе выбрать новое значение y^* и перейти к шагу 2.

Шаг 5. Конец.

Если функции принадлежности нечетких аргументов $\tilde{x}_i$, $i = 1, n$ являются непрерывными функциями, то нечеткие аргу-

менты дискретизируют [24], а непрерывную функцию принадлежности числа $\tilde{y}$ получают в результате аппроксимации дискретной функции $\mu_{\tilde{y}}(y^*)$. Применение принципа обобщения Заде связано с большими вычислительными трудностями. Известны [16, 24] другие принципы обобщения, обладающие меньшей вычислительной сложностью.

Арифметические операции для нечетких чисел и интервалов могут быть определены с помощью принципа обобщения Заде. Пусть $A \in R$ и $B \in R$ – произвольные нечеткие числа или интервалы с функциями принадлежности $\mu_A(x)$ и $\mu_B(y)$, R – множество вещественных чисел. Причем множества A и B могут быть бесконечными.

Для операции получения противоположного нечеткого числа (интервала) $-A = C = \{z, \mu_C(z)\}$ функция принадлежности определяется по формуле

$$\mu_C(z) = \sup_{z=-x} \{ \min(\mu_A(x)) \}. \quad (1.11)$$

Для операции получения обратного нечеткого числа (интервала) $A^{-1} = C = \{z, \mu_C(z)\}$ функция принадлежности определяется по формуле

$$\mu_C(z) = \sup_{z=\frac{1}{x}} \{ \min(\mu_A(x)) \}. \quad (1.12)$$

Двуместные арифметические операции (сложение, вычитание, умножение и деление) определены на декартовом произведении $R \times R$.

Функция принадлежности операции сложения нечетких чисел (интервалов) $A + B = C = \{z, \mu_C(z)\}$ определяется по формуле

$$\mu_C(z) = \sup_{z=x+y} \{ \min(\mu_A(x), \mu_B(y)) \}. \quad (1.13)$$

Функция принадлежности операции вычитания нечетких чисел (интервалов) $A - B = C = \{z, \mu_C(z)\}$ определяется по формуле

$$\mu_C(z) = \sup_{z=x \cdot y} \left\{ \min(\mu_A(x), (\mu_B(y))) \right\}. \quad (1.14)$$

Функция принадлежности операции умножения нечетких чисел (интервалов) $A \bullet B = C = \{z, \mu_C(z)\}$ определяется по формуле

$$\mu_C(z) = \sup_{z=x \cdot y} \left\{ \min(\mu_A(x), (\mu_B(y))) \right\}. \quad (1.15)$$

Функция принадлежности операции деления нечетких чисел (интервалов) $A / B = C = \{z, \mu_C(z)\}$ определяется по формуле

$$\mu_C(z) = \sup_{z=x/y} \left\{ \min(\mu_A(x), (\mu_B(y))) \right\}. \quad (1.16)$$

В формулах — справа от знака равенства функция супремум (supp), как следует из берется по каждому из множеств значений элементов универсума, которые в свою очередь являются результатом соответствующей обычной арифметической операции над численными значениями элементов области определения операции. Для конечных множеств A и B вместо операции супремум можно использовать операцию максимум.

Выполнение арифметических операций на основе принципа обобщения Заде весьма трудоемко. Практически наибольшее распространение получили операции над нечеткими числами и интервалами, представленными в виде так называемых (L-R)-функций [12, 14, 17, 26]. Представление нечетких чисел и интервалов в форме (L-R)-функций возможно для нормальных чисел и интервалов. Практически используемые числа и интервалы обычно являются нормальными, поэтому это ограничение не является существенным.

(L-R)-функции — произвольные функции, заданные на множестве действительных чисел R и осуществляющие отображение $R \rightarrow [0, 1]$, не возрастающие на интервале $[0, +\infty)$ и удовлетворяющие дополнительным условиям

$$L(-x) = L(x), \quad R(-x) = R(x) \quad \text{условие четкости};$$

$L(0)=R(0)$ условие нормирования.

Примерами (L-R)-функций являются треугольные и трапециевидные функции принадлежности, а также функции $f(x)=e^{-|x|^p}$

(рис. 16), $f(x)=\frac{1}{1+|x|^p}$, где $p \geq 0$

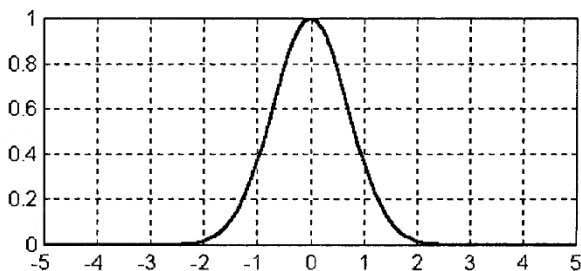


Рис. 16 – График функции $f(x)=e^{-|x|^p}$ при $p=2$

Нечетким числом (L-R)-типа называется нечеткая величина $A=\{x, \mu_A(x)\}$, функция принадлежности которой может быть представлена в виде композиции некоторой L-функции (от англ. left – левый) и некоторой R-функцией (от англ. right – правый):

$$\mu_A(x)=\begin{cases} L\left(\frac{a-x}{\alpha}\right), & \text{если } x \leq a, \\ R\left(\frac{x-a}{\beta}\right), & \text{если } x \geq a, \end{cases} \quad (1.17)$$

где a – модальное значение (мода) нечеткого числа, $\alpha > 0$ и $\beta > 0$ – левый и правый коэффициенты нечеткости.

Таким образом, нечеткое число (L-R)-типа при фиксированных функциях L и R определяется тройкой параметров $\langle a, \alpha, \beta \rangle$. Операции с нечеткими числами (L-R)-типа сводятся к операциям с этой тройкой параметров. Нечеткие числа (L-R)-типа часто обозначают как $A_{LR}=\langle a, \alpha, \beta \rangle$.

Нечетким интервалом (L-R)-типа называется нечеткая величина $A = \{x, \mu_A(x)\}$, функция принадлежности которой может быть представлена в форме композиции некоторой L-функции и некоторой R-функции:

$$\mu_A(x) = \begin{cases} L\left(\frac{a-x}{\alpha}\right), & \text{если } x \leq a, \\ R\left(\frac{x-a}{\beta}\right), & \text{если } x \geq a, \end{cases} \quad 1.18)$$

где a и b – нижнее и верхнее модальные значения ($a < b$), определяющие ядро нечеткого множества; $\alpha > 0$ и $\beta > 0$ – левый и правый коэффициенты нечеткости.

Нечеткий интервал (L-R)-типа при фиксированных функциях L и R определяется четверкой параметров $\langle a, b, \alpha, \beta \rangle$. Операции с нечеткими числами (L-R)-типа сводятся к операциям с этой четверкой параметров. Нечеткие интервалы (L-R)-типа часто обозначают как $A_{LR} = \langle a, b, \alpha, \beta \rangle$.

Результат арифметических операций над нечеткими числами и интервалами (L-R)-типа точно или приближенно равен нечеткому числу или интервалу (L-R)-типа с теми же функциями L-типа и R-типа, а параметры функций вычисляются на основе параметров исходных чисел и интервалов. Пусть $A_{LR} = \langle a_1, b_1, \alpha_1, \beta_1 \rangle$ и $B_{LR} = \langle a_2, b_2, \alpha_2, \beta_2 \rangle$ – произвольные нечеткие числа (L-R)-типа, $C_{LR} = \langle a, b, \alpha, \beta \rangle$ – нечеткое число, являющееся результатом операции. Приведем без доказательства формулы для расчета параметров результатов арифметических операций над нечеткими числами (L-R)-типа [14] (табл. 2). Некоторые формулы, приведенные в [12], отличаются.

Таблица 2 – Параметры результатов арифметических операций над нечеткими числами (L-R)-типа

Операция	a	a	β
Противоположное нечеткое число $C = -A$	$-a_1$	β_1	a_1
Обратное нечеткое число $C = A^{-1}$	$\frac{1}{a_1}, a_1 \neq 0$	$\frac{\beta_1}{a_1(a_1 + \beta_1)}, a_1 \neq 0$	$\frac{\alpha_1}{a_1(a_1 - \alpha_1)}, a_1 \neq \alpha_1$
Сложение нечетких чисел $C = A + B$	$a_1 + a_2$	$\alpha_1 + \alpha_2$	$\beta_1 + \beta_2$
Вычитание нечетких чисел $C = A - B$	$a_1 - a_2$	$\alpha_1 + \alpha_2$	$\beta_1 + \beta_2$
Умножение нечетких положительных чисел $A > 0, B > 0$ (носитель которых содержит только положительные числа) $A \bullet B$	$a_1 a_2$	$a_1 \alpha_2 + a_2 \alpha_1 - \alpha_1 \alpha_2$	$a_1 \beta_2 + a_2 \beta_1 + \beta_1 \beta_2$
Умножение положительного $A > 0$ и отрицательного $B < 0$ нечетких чисел $A \bullet B$	$a_1 a_2$	$a_1 \alpha_2 - a_2 \beta_1 + \alpha_2 \beta_1$	$a_1 \beta_2 - a_2 \alpha_1 - \alpha_1 \beta_2$
Умножение отрицательного $A < 0$ и положительного $B > 0$ нечетких чисел $A \bullet B$	$a_1 a_2$	$-a_1 \beta_2 + a_2 \alpha_1 + \alpha_1 \beta_2$	$-a_1 \alpha_2 + a_2 \beta_1 - \alpha_2 \beta_1$
Умножение отрицательных нечетких чисел $A < 0$ и $B < 0$ $A \bullet B$	$a_1 a_2$	$-a_1 \beta_2 - a_2 \beta_1 - \beta_1 \beta_2$	$-a_1 \alpha_2 - a_2 \alpha_1 + \alpha_1 \alpha_2$

Перемножение нечетких нулей (нечетких чисел «примерно»)	0	$\max(\alpha_1\beta_2, \alpha_2\beta_1)$	$\max(\alpha_1\alpha_2, \beta_1\beta_2)$
Умножение двух нечетких чисел неопределенного знака (произвольные числа – элементы носителя могут быть любыми) $A \cdot B$		$a_1a_2 - \min[(a_1 - \alpha_1)(a_2 - \alpha_2), (a_1 - \alpha_1)(a_2 + \beta_2), (a_1 + \beta_1)(a_2 - \alpha_2), (a_1 + \beta_1)(a_2 + \beta_2)]$	$-a_1a_2 + \max[(a_1 - \alpha_1)(a_2 - \alpha_2), (a_1 - \alpha_1)(a_2 + \beta_2), (a_1 + \beta_1)(a_2 - \alpha_2), (a_1 + \beta_1)(a_2 + \beta_2)]$
Деление положительных нечетких чисел $A \div B$	$\frac{a_1}{a_2}$	$\frac{a_1\beta_2 + a_2\alpha_1}{a_2(a_2 + \beta_2)}, a_2 \neq 0$	$\frac{a_1\alpha_2 + a_2\beta_1}{a_2(a_2 - \alpha_2)}, a_2 \neq 0, a_2 - \alpha_2 \neq 0$
Деление положительного нечеткого числа на отрицательное	$\frac{a_1}{a_2}$	$\frac{a_1\beta_2 - a_2\beta_1}{a_2(a_2 + \beta_2)}, a_2 \neq 0$	$\frac{a_1\alpha_2 - a_2\alpha_1}{a_2(a_2 - \alpha_2)}, a_2 \neq 0, a_2 - \alpha_2 \neq 0$
Деление отрицательного нечеткого числа на положительное	$\frac{a_1}{a_2}$	$\frac{-a_1\alpha_2 + a_2\alpha_1}{a_2(a_2 - \alpha_2)}, a_2 \neq 0, a_2 - \alpha_2 \neq 0$	$\frac{-a_1\beta_2 + a_2\beta_1}{a_2(a_2 + \beta_2)}, a_2 \neq 0$
Деление отрицательных нечетких чисел	$\frac{a_1}{a_2}$	$\frac{-a_1\alpha_2 - a_2\beta_1}{a_2(a_2 - \alpha_2)}, a_2 \neq 0, a_2 - \alpha_2 \neq 0$	$\frac{-a_1\beta_2 - a_2\alpha_1}{a_2(a_2 + \beta_2)}, a_2 \neq 0$
Деление двух нечетких чисел неопределенного знака	$\frac{a_1}{a_2}$	$\frac{a_1}{a_2} - \min\left(\frac{a_1 - \alpha_1}{a_2 - \alpha_2}, \frac{a_1 - \alpha_1}{a_2 + \beta_2}, \frac{a_1 + \beta_1}{a_2 - \alpha_2}, \frac{a_1 + \beta_1}{a_2 + \beta_2}\right), a_2 - \alpha_2 \neq 0$	$-\frac{a_1}{a_2} - \max\left(\frac{a_1 - \alpha_1}{a_2 - \alpha_2}, \frac{a_1 - \alpha_1}{a_2 + \beta_2}, \frac{a_1 + \beta_1}{a_2 - \alpha_2}, \frac{a_1 + \beta_1}{a_2 + \beta_2}\right), a_2 - \alpha_2 \neq 0$

Сравнивая результаты арифметических операций, выполненных с использованием принципа обобщения и L-R-представления [14], видно, что результаты имеют одинаковые модальные значения и величины разброса и похожие, но не совпадающие функции принадлежности. При выполнении операций над нечеткими числами всегда получаются нечеткие числа. Поэтому

разность двух одинаковых нечетких чисел $A - A$ равна нечеткому нулю.

На практике широко используются треугольные нечеткие числа (L-R)-типа и трапециевидные нечеткие интервалы (L-R)-типа. Сравнивая треугольную функцию принадлежности с функцией принадлежности нечеткого числа (L-R)-типа, получаем тройку, описывающую треугольное число (L-R)-типа $\langle b, b-a, c-b \rangle$, где a , b и c – параметры треугольной функции принадлежности. Сравнивая трапециевидную функцию принадлежности с функцией принадлежности нечеткого интервала (L-R)-типа, получаем четверку, описывающую нечеткий интервал (L-R)-типа $\langle b, c, b-a, d-c \rangle$, где a, b, c, d – параметры трапециевидной функции принадлежности. Операции с треугольными нечеткими числами (L-R)-типа и трапециевидными нечеткими интервалами (L-R)-типа производятся в соответствии с правилами операций с нечеткими числами (L-R)-типа, но вычисления существенно упрощаются.

2 ОСНОВЫ НЕЧЕТКОЙ ЛОГИКИ. СИСТЕМЫ НЕЧЕТКОГО ВЫВОДА

2.1 Нечеткое высказывание. Логические операции с нечеткими высказываниями

Элементарным нечетким высказыванием называется повествовательное предложение, выражающее законченную мысль, относительно истинности или ложности которой можно судить только с некоторой степенью уверенности. Подробнее понятие нечеткого высказывания рассматривается в разделе 2.3. Примером нечеткого высказывания является высказывание «высокая температура больного». В классической математической логике множество значений истинности элементарного высказывания состоит из двух элементов: {«истина», «ложь»} или {1,0}. В нечеткой логике значение (степень) истинности [1-2] элементарного нечеткого высказывания A принимает значения из замкнутого интервала $[0,1]$. Значения 0 и 1 являются предельными значениями степени истинности и совпадают со значениями «ложь» и «истина». В нечеткой логике истинность рассматривается как лингвистическая переменная с термами «истинно» и «ложно», имеющими свои функции принадлежности. Например, Заде [6] предложил следующие функции принадлежности для термов «истинно» и «ложно»

$$\mu_{\text{«истинно»}}(u) = \begin{cases} 0, & 0 \leq u \leq a, \\ 2\left(\frac{u-a}{1-a}\right)^2, & a < u \leq \frac{a+1}{2}, \\ 1 - 2\left(\frac{u-1}{1-a}\right)^2, & \frac{a+1}{2} < u \leq 1, \end{cases}$$

$$\mu_{\text{«ложно»}}(u) = \mu_{\text{«истинно»}}(1-u),$$

где $u \in [0, 1]$ – значение истинности; $a \in [0, 1]$ – параметр, задающий носители нечетких множеств «истинно» «ложно». Для нечеткого множества «истинно» носителем будет интервал $(a, 1]$, а для нечеткого множества «ложно» – $[0, a)$.

Лингвистические модификаторы «очень», «более-менее» позволяют получить термы «очень ложно», «более-менее ложно» и т. п.

В нечеткой логике нельзя для определения логических операций использовать таблицы истинности. Но нечеткие высказывания являются элементами множеств нечетких высказываний, поэтому для логических операций используются рассмотренные в разделе 1.5 операции с нечеткими множествами. Рассмотрим основные логические операции с нечеткими высказываниями. По аналогии с общим определением нечетких множеств обозначим нечеткие высказывания через A и B , а функции принадлежности, задающие значение истинности этих высказываний, через $\mu_A(u)$ и $\mu_B(u)$, $u \in U$, где U – множество элементарных нечетких высказываний.

Отрицанием нечеткого высказывания $\bar{A}$ (также используются обозначения $\neg A$, $\text{not } A$, $\text{не } A$) называется операция, результатом которой является нечеткое высказывание, функция принадлежности которого имеет вид $\mu_{\bar{A}}(u) = 1 - \mu_A(u)$.

Логическая конъюнкция нечетких высказываний A и B (обозначается через $A \wedge B$, используются также обозначения $A \& B$, $A \text{ and } B$, $A \text{ и } B$) – это логическая операция, результатом которой является нечеткое высказывание, функция принадлежности которого вычисляется с использованием t-нормы (разд. 1.5). Например, min-конъюнкция определяется формулой $\mu_{A \wedge B}(u) = \min(\mu_A(u), \mu_B(u))$.

Для определения конъюнкции также могут быть использованы алгебраическое произведение функций принадлежности $\mu_{A \wedge B}(u) = \mu_A(u) \cdot \mu_B(u)$, граничное произведение $\mu_{A \wedge B}(u) = \max\{\mu_A(u) + \mu_B(u) - 1, 0\}$, драстическое произведение

$$\mu_{A \wedge B}(x) = \begin{cases} \mu_A(u), & \text{если } \mu_B(u) = 1, \\ \mu_B(u), & \text{если } \mu_A(u) = 1, \\ 0, & \text{в остальных случаях.} \end{cases}$$

Дизъюнкцией нечетких высказываний A и B (обозначается как $A \vee B$, применяют также обозначения $A | B$, $A \text{ or } B$, A и B) называется логическая операция, результатом которой является нечеткое высказывание, функция принадлежности которого вычисляется с использованием s -нормы (разд. 1.5). Например, тах-конъюнкция определяется формулой

$$\mu_{A \vee B}(u) = \max(\mu_A(u), \mu_B(u)).$$

Для определения дизъюнкции также могут быть использованы алгебраическая сумма функций принадлежности $\mu_{A \vee B}(u) = \mu_A(u) + \mu_B(u) - \mu_A(u) \cdot \mu_B(u)$, граничная сумма $\mu_{A \vee B}(u) = \min\{\mu_A(u) + \mu_B(u), 1\}$, драстическое произведение

$$\mu_{A \vee B}(x) = \begin{cases} \mu_A(u), & \text{если } \mu_B(u) = 0, \\ \mu_B(u), & \text{если } \mu_A(u) = 0, \\ 0, & \text{в остальных случаях.} \end{cases}$$

Эквивалентностью нечетких высказываний A и B (обозначается как $A \equiv B$, используются также обозначения $A \approx B$, $A \Leftrightarrow B$, $A \leftrightarrow B$) называется логическая операция, результатом которой является нечеткое высказывание, функция принадлежности которого вычисляется по формуле

$$\mu_{A \equiv B}(u) = \min\{\max\{\mu_{\bar{A}}(u), \mu_B(u)\}, \max\{\mu_A(u), \mu_{\bar{B}}(u)\}\}.$$

Особую роль в нечеткой логике играет нечеткая импликация. Напомним, что в классической математической логике импликацией называется операция над двумя высказываниями A и B , ре-

зультат которой принимает значение «ложь» в случае истинности высказывания A и ложности высказывания B . В остальных случаях результатом импликации является «истина». Импликация обозначается через $A \rightarrow B$, используются также обозначения $A \Rightarrow B$, $A \supset B$, if A then B , если A то B . Высказывание $A \rightarrow B$ читается как « A включает B », «из A следует B », «если A то B », (последнюю форму прочтения не следует путать с правилом продукции).

Известно [4, 12, 14] несколько формул вычисления функции принадлежности нечеткой импликации. В модели Мамдани (E. Mamdani) функция принадлежности определяется по формуле $\mu_{A \rightarrow B}(u) = T(\mu_A(u), \mu_B(u))$, T – произвольная t -норма.

Чаще всего используется правило типа «min», называемое еще нечеткой импликацией Мамдани

$$\mu_{A \rightarrow B}(u) = \min(\mu_A(u), \mu_B(u)). \quad (1.19)$$

Применяется еще правило типа «произведение», или правило Ларсена (Larsen)

$$\mu_{A \rightarrow B}(u) = \mu_A(u) \cdot \mu_B(u). \quad (1.20)$$

Правила типа Мамдани не являются импликациями в логическом смысле, так как при предельных значениях $\mu_A(u)$ и $\mu_B(u)$, равных 0 и 1 правила не дают результат классической импликации четкой логики. Легко проверить, что при этих условиях правила Мамдани дают значение $\mu_{A \rightarrow B}(u) = 1$ только при $\mu_A(u) = 1$ и $\mu_B(u) = 1$. При всех остальных значениях $\mu_{A \rightarrow B}(u) = 0$, что противоречит классической математической логике.

Предложено несколько формул вычисления функции принадлежности импликации, дающих импликацию в логическом смысле [27]. Прежде всего, это классическая импликация Заде (Zadeh L.)

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

Используются также следующие формулы:

– бинарная импликация Клине-Динеса (Kleene-Dienes)

$$\mu_{A \rightarrow B}(u) = \max\{1 - \mu_A(u), \mu_B(u)\}$$

– импликация Лукасевича (Łukasiewicz)

$$\mu_{A \rightarrow B}(u) = \min\{1, 1 - \mu_A(u) + \mu_B(u)\}$$

– импликация Райхенбаха (Reichenbach)

$$\mu_{A \rightarrow B}(u) = -\mu_A(u) + \mu_A(u) \cdot \mu_B(u)$$

– импликация Фодора (Fodor)

$$\mu_{A \rightarrow B}(x) = \begin{cases} 1, & \text{если } \mu_A(u) \leq \mu_B(u), \\ \max\{1 - \mu_A(u), \mu_B(u)\}, & \text{если } \mu_A(u) > \mu_B(u); \end{cases}$$

– импликация Решера (Rescher)

$$\mu_{A \rightarrow B}(x) = \begin{cases} 1, & \text{если } \mu_A(u) \leq \mu_B(u), \\ 0, & \text{если } \mu_A(u) > \mu_B(u); \end{cases}$$

– импликация Гогуена (Goguen)

$$\mu_{A \rightarrow B}(x) = \begin{cases} \min\left\{1, \frac{\mu_B(u)}{\mu_A(u)}\right\}, & \text{если } \mu_A(u) > 0, \\ 1, & \text{если } \mu_A(u) = 0; \end{cases}$$

– импликация Гёделя (Gödel)

$$\mu_{A \rightarrow B}(x) = \begin{cases} 1, & \text{если } \mu_A(u) \leq \mu_B(u), \\ \mu_B(u), & \text{если } \mu_A(u) > \mu_B(u); \end{cases}$$

– импликация Ягера (Yager)

$$\mu_{A \rightarrow B}(x) = \begin{cases} 1, & \text{если } \mu_A(u) = 0, \\ \mu_B(u)^{\mu_A(u)}, & \text{если } \mu_A(u) > 0; \end{cases}$$

– импликация Вильмотта (Willmott)

$$\mu_{A \rightarrow B}(x) = \min \left\{ \max \{1 - \mu_A(u), \mu_B(u)\}, \max \{ \mu_A(u), 1 - \mu_B(u), \min \{1 - \mu_A(u), \mu_B(u)\} \} \right\};$$

– импликация Дюбуа-Прайда (Dubois-Prade)

$$\mu_{A \rightarrow B}(x) = \begin{cases} 1 - \mu_A(u), & \text{если } \mu_B(u) = 0, \\ \mu_B(u), & \text{если } \mu_A(u) = 1, \\ 1, & \text{в остальных случаях.} \end{cases}$$

Известны и другие формулы [12, 14].

2.2 Вывод в нечеткой логике

Рассмотрим правила вывода в классической математической логике. Логический вывод часто задается в виде схемы: над горизонтальной чертой записываются все суждения, на основании которых принимается решение, а под чертой — результат вывода. Если истинны все суждения над чертой, то истинно суждение под чертой (из истинных суждений может выводиться только истинный результат).

В математической логике применяются два основных правила вывода:

– modusponens (лат. утверждающий модус, читается «модус поненс»);

– modustollens (лат. отрицающий модус, читается «модус толленс»).

Схемы вывода по правилу modusponens имеет вид:

Суждение	A
Импликация	$A \rightarrow B$
Вывод	B

Схемы вывода по правилу modustollens имеет вид:

Суждение	$\bar{B}$
Импликация	$A \rightarrow B$
Вывод	$\bar{A}$

Здесь A и B – некоторые суждения.

В нечетком выводе суждения в правилах modusponens и modustollens характеризуются некоторыми нечеткими множествами. Таким образом получим обобщенные правила вывода в нечеткой логике. Обобщенное нечеткое правило modusponens определяется схемой вывода

Суждение	x есть A'
Импликация	Если x есть A то y есть B
Вывод	y есть B'

где $A, A' \subseteq X$ и $B, B' \subseteq Y$ – нечеткие множества; x и y – лингвистические переменные.

Рассмотрим пример [18]. Пусть имеется схема нечеткого вывода:

Суждение	Скорость автомобиля большая
Импликация	Если скорость автомобиля очень большая, то уровень шума высокий
Вывод	Уровень шума в автомобиле не очень высокий

В качестве лингвистических переменных выступают: x – «скорость автомобиля», y – «уровень шума». Терм-множеством лингвистической переменной x является множество

$$T_1 = \{ \text{"малая скорость"}, \text{"средняя скорость"}, \\ \text{"большая скорость"}, \text{"очень большая скорость"} \}.$$

Терм-множеством лингвистической переменной u является множество

$$T_2 = \{\text{"малый уровень шума"}, \text{"средний уровень шума"}, \\ \text{"не очень высокий уровень шума"}, \text{"высокий уровень шума"}\}.$$

Каждому элементу терм-множеств T_1 и T_2 можно поставить в соответствие нечеткое множество, имеющее некоторую функцию принадлежности. В примере можно выделить нечеткие множества A' – «большая скорость автомобиля» и B – «высокий уровень шума». С помощью модификатора концентрирования (раздел 1.5) можно построить функцию принадлежности множества A – «очень большая скорость автомобиля». С помощью модификатора растяжения можно построить множество B' – «не очень высокий уровень шума».

Нечеткий вывод отличается от четкого вывода. В четком выводе суждение A из импликации четкого правила присутствует в условии этого правила. В нечетком выводе суждение A' в общем случае лишь приблизительно равно суждению A из нечеткой импликации. В результате заключение нечеткого вывода B' отличается от заключения B нечеткой импликации.

Рассмотрим формирование функции принадлежности заключения B' нечеткого вывода. Нечеткая импликация $A \rightarrow B$ является нечетким отношением. Так как $A, A' \subseteq X$ и $B, B' \subseteq Y$, то нечеткая импликация $A \rightarrow B$ представляет собой отношение с областью определения $X \times Y$ и функцией принадлежности $\mu_{A \rightarrow B}(x, y)$. Функцию принадлежности нечеткой импликации можно вычислить (раздел 2.1). Заключение B' представляет собой композицию (комбинацию) нечеткого множества A' и нечеткого отношения – импликации $A \rightarrow B$ (разд. 1.6), то есть $B' = A' \otimes (A \rightarrow B)$, функция принадлежности которого может быть вычислена, например, по . В этом случае мы имеем дело с композиционным правилом нечеткого вывода Заде [23, 24].

Обобщенное нечеткое правило modustollens определяется схемой вывода

Суждение Импликация	y есть B' Если x есть A то y есть B
Вывод	x есть A'

где $A, A' \subseteq X$ и $B, B' \subseteq Y$ – нечеткие множества; x и y – лингвистические переменные.

В продолжение примера рассмотрим схему нечеткого вывода:

Суждение	Уровень шума в автомобиле не очень высокий
Импликация	Если скорость автомобиля очень большая, то уровень шума высокий
Вывод	Скорость автомобиля большая

Легко увидеть, что нечеткое заключение A' определяется композицией нечеткого отношения и нечеткого множества $A' = (A \rightarrow B) \otimes B'$.

2.3 Нечеткие лингвистические высказывания

Нечеткими лингвистическими высказываниями называются нечеткие высказывания трех видов [12]:

нечеткое высказывание « β есть α » (« β is α »), где β – имя лингвистической переменной, α – значение лингвистической переменной, которому соответствует отдельный лингвистический терм из базового терм-множества T лингвистической переменной β . Лингвистический терм задается своей функцией принадлежности;

высказывание « β есть $\nabla \alpha$ », где ∇ – лингвистический модификатор, соответствующий таким словам, как: «ОЧЕНЬ», «БОЛЕЕ ИЛИ МЕНЕЕ», «МНОГО БОЛЬШЕ» и другим. Значение $\nabla \alpha$ формируется с использованием процедур G и M данной лингвистической переменной;

составные высказывания, образованные из высказываний видов 1 и 2 и нечетких логических операций в форме связок: «И», «ИЛИ», «ЕСЛИ-ТО», «НЕ».

В дальнейшем изложении нечеткие лингвистические высказывания будем называть просто нечеткими высказываниями.

Например, высказывание «температура больного очень высокая» является высказыванием второго вида. Лингвистическая переменная «температура больного» принимает значение «очень высокая температура». Это значение получено применением модификатора «ОЧЕНЬ» к функции принадлежности терма «высокая температура».

2.4 Правила нечетких продукций

Правилом нечеткой продукции (или просто – нечеткой продукцией) называется выражение

$$(i): Q; P; A \rightarrow B; S; F; N,$$

где – имя нечеткой продукции; Q – сфера применения нечеткой продукции, то есть предметная область знания, которую представляет продукция; P – условие применимости ядра нечеткой продукции (логическое выражение, при истинности которого становится возможной активизация ядра продукции; этот элемент продукции часто опускается или вводится в ядро продукции); $A \rightarrow B$ – ядро нечеткой продукции; A – условие ядра (англ. antecedent – антецедент); B – заключение ядра (англ. consequent – консеквент); S – метод или способ определения количественного значения степени истинности заключения B ядра на основе известного значения степени истинности условия A (способ S определяет алгоритм или так называемую схему нечеткого вывода); F – коэффициент определенности или уверенности нечеткой продукции (весовой коэффициент нечеткого правила продукции), принимает значение из интервала $[0,1]$ (используется не во всех алгоритмах вывода); N – постусловие продукции – действия, ко-

которые должны быть выполнены в случае реализации ядра продукции (в случае истинности B).

Ядро продукции обычно записывается в форме: «ЕСЛИ A ТО B » (англ.: «IF A THEN B »). В системах нечеткого вывода условия A и заключения B ядра продукции формулируются в форме нечетких лингвистических высказываний (разд. 2.3). Тогда правило нечеткой продукции может быть записано, например, в виде:

ПРАВИЛО $\langle \# \rangle$: ЕСЛИ " β_1 есть α_1 " ТО " β_2 есть α_2 ", (1.21)

Где $\langle \# \rangle$ – номер правила, " β_1 есть α_1 " – условие правила, " β_2 есть α_2 " – заключение правила, причем $\beta_1 \neq \beta_2$.

Система нечетких правил продукции или продукционная нечеткая система – это согласованное множество отдельных нечетких продукций форме «ЕСЛИ A ТО B », где A и B – нечеткие лингвистические высказывания вида 1, 2 или 3.

В качестве условий или заключений правил нечеткой продукции зачастую выступают составные нечеткие высказывания, образованные из высказываний видов 1 и 2 (подусловий), соединенных нечеткими логическими операциями. Пусть нечеткими логическими операциями соединены нечеткие высказывания, относящиеся к одной и той же лингвистической переменной. Например, " β есть α_1 " ИЛИ " β есть α_2 ". Это высказывание эквивалентно нечеткому высказыванию " β есть α_3 ", где терм-множеству α_3 соответствует нечеткое множество равное пересечению нечетких множеств, соответствующих термам α_1 и α_2 . Например, нечеткое высказывание

«температура больного высокая» ИЛИ «температура больного очень высокая» эквивалентно высказыванию

«температура больного высокая или очень высокая».

Функция принадлежности терма «высокая или очень высокая» вычисляется по рассмотренным правилам. В примере использован модификатор «очень». Аналогично обрабатываются высказывания вида " β есть α_1 " И " β есть α_2 ".

Если подусловия содержат разные лингвистические переменные, например,

ПРАВИЛО $\langle \# \rangle$: ЕСЛИ " β_1 есть α_1 " И " β_2 есть α_2 " ТО " β_3 есть α_3 ", (1.22)

или

ПРАВИЛО $\langle \# \rangle$: ЕСЛИ " β_1 есть α_1 " ИЛИ " β_2 есть α_2 " ТО " β_3 есть α_3 ", (1.23)

где $\beta_1 \neq \beta_2 \neq \beta_3$, то используется агрегирование условий в левой части правил (подраздел 2.5).

Заключения правил нечеткой продукции также могут быть составными, то есть состоять из нескольких подзаключений. Если подзаключения содержат одну и ту же лингвистическую переменную, то используются нечеткие логические операции. Если подзаключения содержат разные лингвистические переменные, например,

ПРАВИЛО $\langle \# \rangle$: ЕСЛИ " β_1 есть α_1 " ТО " β_2 есть α_2 " И " β_3 есть α_3 ", (1.24)

где $\beta_1 \neq \beta_2 \neq \beta_3$, то используется аккумуляция заключений в правилах нечеткой продукции (подраздел 2.5).

2.5 Этапы нечеткого вывода

Типовая структура процесса нечеткого вывода показана на рис. 17.



Рис. 17 – Процесс нечеткого вывода

Прежде всего, должна быть сформирована база правил, представляющая собой конечное множество правил нечетких продукций. Формирование базы правил включает определение входных

и выходных лингвистических переменных, а также собственно правил. Входными лингвистическими переменными называются лингвистические переменные, используемые в подусловиях правил. Выходные переменные – переменные, используемые в подзаключениях правил. Определение лингвистических переменных означает определение базовых терм-множеств переменных и функций принадлежности терм-множеств. Правила формируются, как было рассмотрено в разделе 2.4. Каждому правилу может быть приписан вес, принимающий значение из интервала $[0,1]$. Если вес отсутствует, можно считать, что вес равен нулю.

На вход системы нечеткого вывода поступает вектор $\mathbf{x}^* = [x_1^*, x_2^*, \dots, x_i^*, \dots]^T$ четких значений лингвистических переменных β_i . Блок фаззификации (англ. fuzzification – приведении к нечеткости) вычисляет степени принадлежности этих значений нечетким множествам значений лингвистических переменных. Для этого должны быть известны функции каждого термина лингвистической переменной.

Фаззификация производится следующим образом. Пусть для каждой входной лингвистической переменной β_i известно ее числовое значение x_i^* . Рассматривается каждое высказывание подусловий, в котором фигурирует переменная β_i , например, " β_i есть α_i ", где α_i – терм с известной функцией принадлежности $\mu_i(x)$. Значение x_i^* используется в качестве аргумента $\mu_i(x)$, в результате чего находится $m_i = \mu_i(x_i^*)$. При этом могут использоваться модификаторы. Таким образом вычисляются значения истинности всех подусловий системы нечеткого вывода. Высказывания в подусловиях заменяются числами. На выходе блока фаззификации формируется вектор $\mathbf{m} = [\mu_1(x_1^*), \mu_2(x_2^*), \dots, \mu_i(x_i^*), \dots]^T$, который является входом блока вывода.

Блок нечеткого логического вывода получает на входе вектор степени истинности всех подусловий $\mathbf{m}$ и вычисляет результирующую функцию принадлежности $\mu_{рез}$ выходного значения (система вывода может иметь несколько выходов, тогда речь идет о выходном векторе). Вычисление результирующей функции принадлежности включает следующие процедуры [12] (в скобках указаны названия процедур в соответствии с международным

стандартом языков программирования контроллеров IEC 1131 – Programmable Controllers. Part 7 – Fuzzy Control Programming [12]):

- вычисление степени истинности условий (Aggregation – агрегирование);
- определение активизированных функций принадлежности заключений (Activftion – активизация);
- определение результирующих функций принадлежности выходных лингвистических переменных (Accumulation – аккумуляция).

В процедуре вычисления степени истинности условий по каждому из правил системы нечеткого вывода (агрегирование) рассматривается каждое условие правил системы нечеткого вывода и вычисляется степень истинности условий. Исходными данными являются степени истинности подусловий (вектор m), вычисленные в блоке фаззификации. Если условие содержит одно нечеткое высказывание вида , то степень истинности условия равна степени истинности высказывания условия. Если условие состоит из двух подусловий, связанных конъюнкцией , или дизъюнкцией , степень выполнения условия вычисляется с помощью треугольных норм (раздел 1.5). Например, для условия правила ЕСЛИ " β_1 есть α_1 " И " β_2 есть α_2 " получаем $\mu(x_1^*, x_2^*) = T(\mu_{\alpha_1}(x_1^*), \mu_{\alpha_2}(x_2^*))$,

x_1^* и x_2^* – значения входных переменных x_1 и x_2 ,

T – один из операторов t-нормы, $\mu_{\alpha_1}(x)$ и $\mu_{\alpha_2}(x)$ – функции принадлежности термов α_1 и α_2 .

Аналогично для условия правила :

$$\mu(x_1^*, x_2^*) = S(\mu_{\alpha_1}(x_1^*), \mu_{\alpha_2}(x_2^*)),$$

где S – один из операторов s-нормы. Если условие содержит множество подусловий, соединенных дизъюнкциями и конъюнкциями, то сначала вычисляются степени истинности подусловий, соединенных конъюнкциями, затем – дизъюнкциями. Как обычно, скобки нарушают порядок действий. Рекомендуется [12] исполь-

зовать согласованные правила расчета истинности. Например, если для вычисления нечеткой конъюнкции используется операция $\min$ -пересечения, то для вычисления нечеткой дизъюнкции следует применить операцию $\max$ -объединения.

Процедура определения активизированных функций принадлежности заключений (активизация) основана на операции нечеткой импликации (раздел 2.1). Входными данными для процедуры являются степени истинности условий правил и функции принадлежности выходных величин, выходными – функции принадлежности всех подзаключений. Рассмотрим пример [14]. Пусть правило имеет вид ЕСЛИ $(x = A)$ ТО $(y = B)$, функции принадлежности $\mu_A(x)$ и $\mu_B(y)$ – треугольные (рис. 18), входное значение $x^* = 6,5$, степень истинности условия $\mu_A(x^*) = 0,5$ (см. рис. 18).

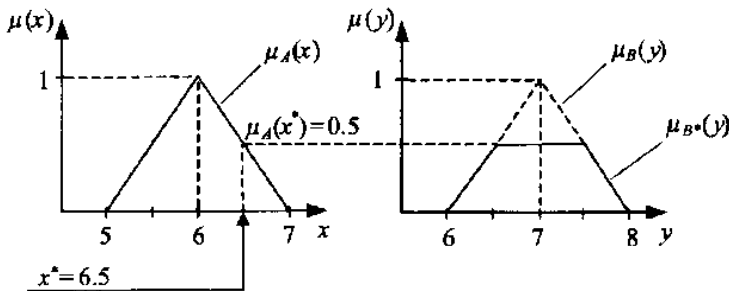


Рис. 18 – Активизация заключения с использованием импликации Мамдани

Используем импликацию Мамдани:

$$\mu_{A \rightarrow B}(x^*, y) = \min(\mu_A(x^*), \mu_B(y)).$$

Практически активизированная функция принадлежности заключения при использовании импликации Мамдани находится простым усечением функции принадлежности заключения $\mu_B(y)$ до уровня степень истинности условия $\mu_A(x^*)$ (рис. 18). Можно использовать другие операторы нечеткой импликации.

Например, результат активизации заключения с использованием правила «произведение» показан на рис. 19.

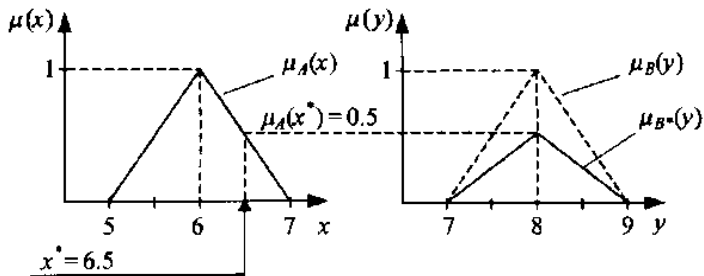


Рис. 19 – Активизация заключения с использованием правила «произведение»

На практике, особенно при наличии в правилах нескольких позаклучений, удобно использовать процедуру активизации, основанную на алгоритме вывода Мамдани [12] (алгоритм будет рассмотрен в разделе 2.6). В этом алгоритме для каждого правила задается весовой коэффициент $F_i \in [0, 1]$. Может быть $F_i = 1$, такое значение принимают, если весовой коэффициент не задан явно. Для отдельных подзаклучений одного правила могут быть заданы разные весовые коэффициенты. Степень истинности всех подзаклучений i -го правила рассчитывается по формуле $c_i = F_i \mu_{A_i}(x_i^*)$.

Активизированная функция принадлежности j -го подзаклучения i -го правила вычисляется по одной из формул, основанных на методе нечеткой композиции:

min-активизация $\mu_j^*(y) = \min\{c_i, \mu_j(y)\}$;

prod-активизация $\mu_j^*(y) = c_i \cdot \mu_j(y)$.

Рассмотренный алгоритм особенно удобен, когда правила содержат по несколько подзаклучений вида .

Так как подзакключения, относящиеся к одной и той же выходной лингвистической переменной, в общем виде принадлежат разным правилам, то необходимо построить единую результирующую функцию принадлежности для каждой выходной пере-

менной. Эта процедура называется аккумуляцией. Аккумуляция производится объединением с помощью одной из s-норм активизированных функций принадлежности каждой выходной лингвистической переменной. В результате для каждой выходной переменной получается одна функция принадлежности, возможно, весьма сложной формы.

Дефаззификация (приведение к четности) – нахождение для каждой выходной лингвистической переменной четкого значения в некотором смысле наилучшим образом, представляющим нечеткую переменную. Необходимость в дефаззификации объясняется тем, что на выходе системы нечеткого вывода нужны, как правило, четкие значения, которые поступают, например, на исполнительный механизм. Так как возможны разные критерии представления четкой переменной одним числом, то существуют различные методы дефаззификации [12, 14, 24]. В результате определения результирующих функций принадлежности выходных лингвистических переменных получаются результирующие функции принадлежности $\mu_{res}(y)$. Для унимодальной функции принадлежности простейшим методом дефаззификации является выбор четкого числа, соответствующего максимальной степени принадлежности. Обобщением этого метода на многомодальные функции являются методы левого и правого модального значения.

В методе левого модального значения (LM – LostMostMaximum), называемом еще метод первого максимума (FM – FirstofMaxima) [14], или наименьший из максимумов (SOM – Smallest Of Maximums) [24] в качестве четкого значения берется $y = \min\{x_m\}$, где x_m – модальное значение результирующей функции принадлежности. Другими словами, в качестве четкой выходной переменной берется наименьшая (самая левая) мода.

В методе правого модального значения (RM – RightMostMaximum), называемом еще метод последнего максимума (LM – LastofMaxima), или метод наибольшего максимума (LOM – Largest Of Maximums) в качестве четкого значения берется $y = \max\{x_m\}$, то есть наибольшая (самая правая) из мод. Примеры дефаззификации с использованием левого и правого модальных значений представлены на рис. 20а и 20б.

В методе среднего максимума (ММ – Middle of Maxima), или методе центра максимумов (МОМ – Mean Of Maxima) находится среднее арифметическое элементов универсального множества, имеющих максимальные степени принадлежности

$$y = \frac{\int_G u du}{\int_G du},$$

где G – множество всех элементов из интервала, имеющих максимальную степень принадлежности нечеткому множеству. Пример дефаззификации с использованием метода среднего максимума представлен на рис. 20в.

Дефаззификация по методу центра тяжести (СГ – Center of Gravity, Centroid) производится по формуле определения центра тяжести плоской фигуры, ограниченной осями координат и графиком функции принадлежности нечеткого множества

$$y = \frac{\int_{Min}^{Max} x \mu(x) dx}{\int_{Min}^{Max} \mu(x) dx},$$

где Min и Max – левая и правая точки интервала носителя выходной переменной.

Пример дефаззификации с использованием метода центра тяжести представлен на рис. 20г.

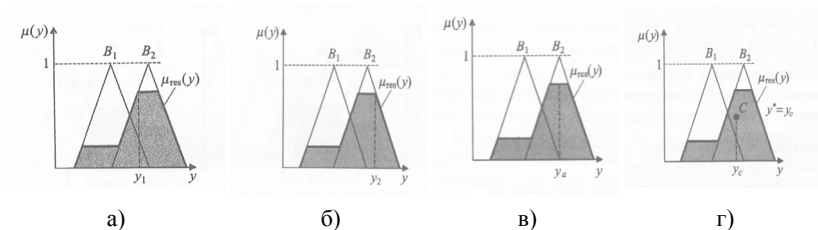


Рис. 20 – Примеры дефаззификации

а) результат дефаззификации по методу левого модального значения $y = y_1$;

б) результат дефаззификации по методу правого модального значения $y = y_2$;

в) результат дефаззификации по методу среднего максимума;

г) результат дефаззификации по методу центра тяжести.

Дефаззификация по методу центра площади (CA – Center of Area, Bisector of Area, Bisector) состоит в нахождении такого чис-

ла y , что $\int_{Min}^y \mu(x) dx = \int_y^{Max} \mu(x) dx$. Геометрический смысл метода

состоит в нахождении такой точки на оси абсцисс, что перпендикуляр, восстановленный в этой точке, делит площадь под кривой функции принадлежности на две равные части.

2.6 Алгоритмы нечеткого вывода

Нечеткий вывод может быть реализован по разным алгоритмам, так как каждый этап вывода позволяет использовать разные методы вычислений и параметры. Рассмотрим наиболее популярные алгоритмы.

Правила базы знаний алгоритма Мамдани (Mamdani E.H.) имеет вид:

(ЕСЛИ " x_1 есть a_{1j} " Θ " x_2 есть a_{2j} " Θ ... " x_n есть a_{nj} ") ТО y_k есть d_{kj} с весом w_j ,

где x_i – i -я ($i = 1, 2, \dots, n$) входная лингвистическая переменная, a_{ij} – терм i -й переменной в j -м правиле ($j = 1, 2, \dots, m$), y_k – k -я выходная переменная ($k = 1, 2, \dots, n$), d_{kj} – заключение k -ой переменной в j -м правиле (правило может содержать несколько подзаключений, соединенных логическими операциями), w_j – вес j -го правила, Θ – логическая операция И или ИЛИ.

В процедуре агрегирования (вычисления степени истинности условий) по каждому из правил системы нечеткого вывода для входного вектора $\mathbf{x}^* = [x_1^*, x_2^*, \dots, x_n^*]$ используется формула

$$\mu_j(\mathbf{x}^*) = \chi_j(\mu_1(x_1^*), \mu_2(x_2^*), \dots, \mu_n(x_n^*)),$$

где $\mu_j(\mathbf{x}^*)$ – степень истинности условия j -го правила, $\mu_i(x_i^*)$ – результат фаззификации входной переменной x_i , χ_j означает t-норму, если в j -ом правиле используется только логическая операция И, и означает s-норму, если в j -ом правиле используется только логическая операция ИЛИ. В алгоритме Мамдани t-норма обычно реализуется операцией минимума, а s-норма – операцией максимума. Если в условии присутствуют разные логические операции, то при отсутствии скобок сначала выполняются операции И (t-норма), а затем – ИЛИ (s-норма). Скобки нарушают порядок операций.

Истинность каждого подзаключения (процедура активизации) равна $\mu_{d_{kj}}^*(y) = \min\{w_j \mu_j(\mathbf{x}^*), \mu_{d_{kj}}(y)\}$, где w_j – вес j -го правила; $\mu_j(\mathbf{x}^*)$ – степень истинности условия j -го правила, найденная в процедуре агрегирования; $\mu_{d_{kj}}(y)$ – функция принадлежности k -ой выходной величины в j -ом правиле.

Единая результирующая функция принадлежности μ_{y_k} для каждой выходной переменной (процедура аккумуляции) в алгоритме Мамдани находится с помощью операции максимума над функциями принадлежности подусловий подзаключений, относящихся к одной выходной переменной.

Четкое значение выходной переменной y'_k (процедура дефаззификации) обычно производится по методу центра тяжести

$$y'_k = \frac{\int_{Min}^{Max} x \mu_{y_k}(x) dx}{\int_{Min}^{Max} \mu_{y_k}(x) dx},$$

где Min и Max – левая и правая точки интервала носителя выходной переменной y_k .

Особенностью алгоритма Сугено-Такаги (Takagi T., Sugeno M.) является представление правил базы знаний в виде

(ЕСЛИ " x_1 есть a_{1j} " Θ " x_2 есть a_{2j} " $\Theta \dots$ " x_n есть a_{nj} ") ТО $y_k = c_{j0} + c_{j1}x_1 + \dots + c_{jn}x_n$

где x_i – i -я ($i = 1, 2, \dots, n$) входная лингвистическая переменная, a_{ij} – терм i -й переменной в j -м правиле ($j = 1, 2, \dots, m$), y_k – k -я выходная переменная ($k = 1, 2, \dots, n$), $c_{j0}, c_{j1}, \dots, c_{jn}$ – действительные числа, Θ – логическая операция И или ИЛИ.

Заключения правил в базе знаний Сугено-Такаги являются линейными функциями входов.

Процедура фаззификации не отличается от фаззификации в алгоритме Мамдани. В процедуре агрегирования подусловий, как правило, используется операция $\min$ -конъюнкции. В процедуре активизации заключений рассчитываются четкие значения выходных переменных каждого j -го правила как линейные функции входных значений нефаззифицированных переменных $y_{kj} = c_{j0} + c_{j1}x_1^* + \dots + c_{jn}x_n^*$. Этап аккумуляции заключений в данном алгоритме отсутствует, так как выходные переменные являются четкими. Формально выходные множества являются одноточечными множествами. Поэтому для дефаззификации можно использовать метод центра тяжести для одноточечных множеств

$$y'_k = \frac{\sum_{j=1}^m \mu_j y_{kj}}{\sum_{j=1}^m \mu_j},$$

где y'_k – четкое значение выходной переменной y_k , μ_j – степень истинности j -го условия, y_{kj} – значение выходной переменной y_k , рассчитанное в j -ом условии.

С другими алгоритмами нечеткого вывода можно познакомиться в [5, 7, 12, 24].

3 ПОНЯТИЕ НЕЧЕТКИХ НЕЙРОННЫХ (ГИБРИДНЫХ) СЕТЕЙ

Основным компонентом в процедурах нечеткого вывода является база правил нечетких продукций. В то же время существуют целые классы прикладных задач, в которых выявление и построение правил нечетких продукций невозможно или связано с серьезными трудностями концептуального характера. К таким задачам относятся задачи распознавания образов, экстраполяции и интерполяции функциональных зависимостей, классификации и прогнозирования, нелинейного и ситуационного управления, а также интеллектуального анализа данных (Data Mining).

Общей особенностью подобных задач является существование некоторой зависимости или отношения, связывающего входные и выходные переменные модели системы, представляемой в форме так называемого «черного ящика». При этом выявление и определение данной зависимости в явном теоретико-множественном или аналитическом виде не представляется возможным либо по причине недостатка информации о моделируемой проблемной области, либо сложности учета многообразия факторов, оказывающих влияние на характер данной взаимосвязи.

Для конструктивного решения подобных задач разработан специальный математический аппарат, получивший название нейронных сетей. Достоинством моделей, построенных на основе нейронных сетей, является возможность получения новой информации о проблемной области в форме некоторого прогноза. При этом построение и настройка нейронных сетей осуществляется посредством их обучения на основе имеющейся и доступной информации.

Недостатком нейронных сетей является представление знаний о проблемной области в специальном виде, которое может существенно отличаться от возможной содержательной интерпретации существующих взаимосвязей и отношений.

Нечеткие нейронные сети или гибридные сети по замыслу их разработчиков призваны объединить в себе достоинства ней-

ронных сетей и систем нечеткого вывода. С одной стороны, они позволяют разрабатывать и представлять модели систем в форме правил нечетких продукций, которые обладают наглядностью и простотой содержательной интерпретации. С другой стороны, для построения правил нечетких продукций используются методы нейронных сетей, что является более удобным и менее трудоемким процессом для системных аналитиков. В последнее время аппарат гибридных сетей повсеместно признается специалистами как один из наиболее перспективных для решения слабо или плохо структурированных задач прикладного системного анализа.

3.1 Основные понятия и определения нечетких нейронных (гибридных) сетей

Нечеткая нейронная сеть (fuzzy-neural networks) функционирует стандартным образом на основе четких действительных чисел. Нечеткой является только интерпретация результатов.

Для пояснения сущности нечетких нейронных сетей, рассмотрим простую нейронную сеть, имеющую два входа и только один нейрон (рис. 21).

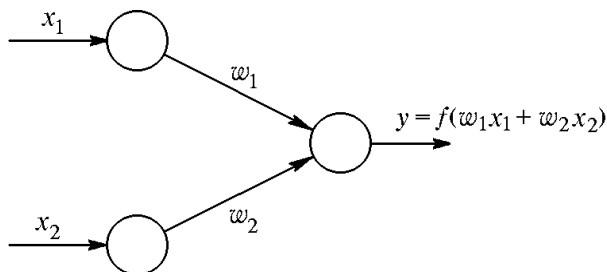


Рис. 21 – Элементарная нейронная сеть

Здесь входные сигналы x_i «взаимодействуют» с весами w_i образуя произведения

$$p_i = x_i w_i, \quad i = 1, 2$$

Такая частная информация (произведения) объединяются с использованием операции суммирования, образуя вход *net* нейрона:

$$net = p_1 + p_2 = x_1 w_1 + x_2 w_2$$

Выход нейрона образуется в результате преобразования входа *net* некоторой активационной функцией:

$$y = f(net) = f(x_1 w_1 + x_2 w_2),$$

например, сигмоидального типа

$$f(x) = \frac{1}{1 + e^{-x}}.$$

Приведенную однослойную нейронную сеть, в которой используются операции умножения, суммирования и сигмоидальная функция активации, будем называть стандартной нейронной сетью.

В случае применения других операций, таких как t-норма или t-конорма (подразд. 1.5), придем к нейронной сети, которая будет называться нечеткой нейронной сетью.

Нечеткая нейронная (гибридная) сеть – это нейронная сеть с четкими сигналами, весами и активационной функцией, но с объединением x_i и w_i , p_1 и p_2 с использованием t-нормы, t-конормы или некоторых других непрерывных операций. Входы, выходы и веса нечеткой нейронной сети вещественные числа, принадлежащие отрезку $[0, 1]$. Нечеткой нейронной сетью обычно называют четкую нейронную сеть, которая построена на основе многослойной архитектуры с использованием «И», «ИЛИ» нейронов.

Рассмотрим следующие примеры элементарных нечетких нейронных сетей.

Нечеткий нейрон «И». Сигналы x_i и w_i в данном случае объединяются с помощью треугольной конормы:

$$p_i = S(x_i w_i), \quad i = 1, 2$$

а выход образуется с применением треугольной нормы (рис. 22):

$$y = \text{AND}(p_1, p_2) = T(p_1, p_2) = T(S(w_1, x_1), S(w_2, x_2))$$

Если принять

$$T = \min, S = \max,$$

тогда нечеткий нейрон «И» реализует композицию min-max:

$$y = \min(w_1 \vee x_1, w_2 \vee x_2).$$

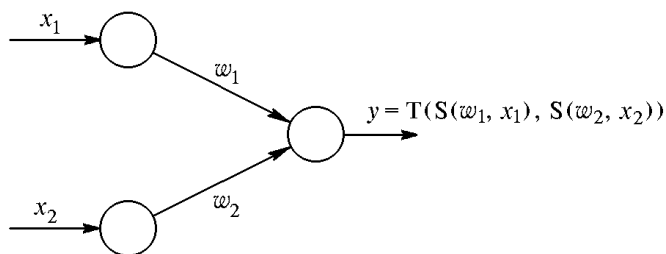


Рис. 22 – Структура гибридного нейрона «И»

Нечеткий нейрон «ИЛИ». Сигналы x_i и w_i здесь объединяются с помощью треугольной нормы: $p_i = T(x_i w_i)$, $i = 1, 2$, а выход образуется с применением треугольной конормы (рис. 23):

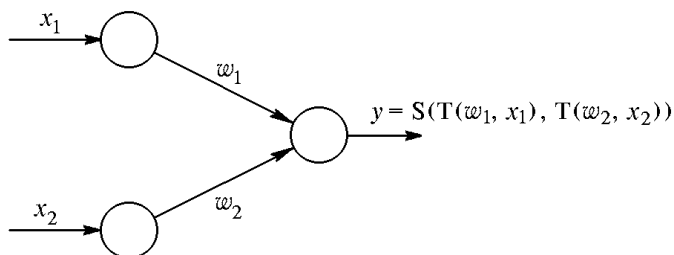


Рис. 23 – Структура гибридного нейрона «ИЛИ»

$$y = \text{OR}(p_1, p_2) = S(p_1, p_2) = S(T(w_1, x_1), T(w_2, x_2))$$

Если принять $T = \min$, $S = \max$, тогда нечеткий нейрон «ИЛИ» реализует композицию max-min: $y = \max(w_1 \wedge x_1, w_2 \wedge x_2)$.

Нечеткая нейронная сеть как правило состоит из четырех слоев: слоя фазификации входных переменных, слоя агрегирования значений активации условия, слоя агрегирования нечетких правил и выходного слоя.

Нечеткая нейронная сеть представляет собой набор нечетких правил, которые описывают классы в имеющемся наборе исходных данных, и нечеткую систему вывода для их переработки с целью получения результата диагностики [12]. Фрагмент нечеткой нейронной сети представлен на рисунке 24.

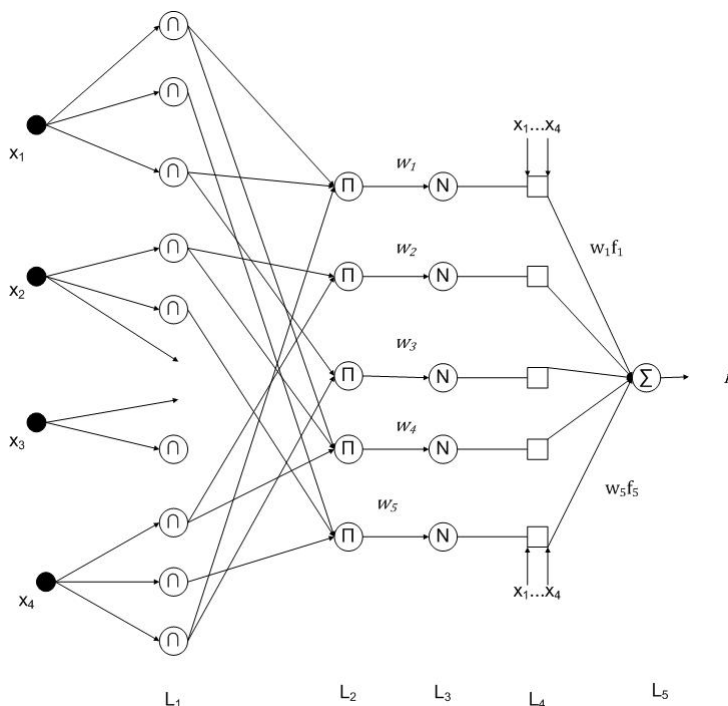


Рис. 24 – Схема нечеткой нейронной сети

На схеме показана нечеткая нейронная сеть с четырьмя входами ($n = 4$). Слои обозначены символами от L1 до L5. Элементы, обозначенные символом Π (мультипликаторы), перемножают все входные сигналы, элементы, обозначенные символом Σ (сумматоры) – суммируют их. Назначение слоев, следующее: первый слой – термы входных переменных; второй слой – antecedentes (посылки) нечетких правил; третий слой – нормализация степеней выполнения правил; четвертый слой – заключения правил; пятый слой – агрегирование результата, полученного по различным правилам. Входы сети в отдельный слой не выделяются.

Каждый элемент слоя 1 (L1) представляет один терм с функцией принадлежности. Входы сети соединены только со своими термами. Количество узлов первого слоя равно сумме мощностей терм-множеств входных переменных. Выходом узла является степень принадлежности значения входной переменной соответствующему нечеткому терму $\mu_{A_k}(\bar{x}_i)$. Фактически, в этом слое оценивается степень принадлежности входных данных $\bar{x}_i$ к соответствующим нечетким множествам A_k . Функциональная зависимость между входом и выходом в узлах этой сети определяется по формуле:

$$\mu_A(x) = \frac{1}{1 + \left(\frac{x - c}{\sigma} \right)^{2b}},$$

Ее параметры c , σ и b будут модифицироваться в процессе обучения, что позволит улучшить подбор нечетких множеств. Факт физической интерпретации этих параметров позволяет получить хорошее начальное размещение функции принадлежности нечетких множеств, а также анализировать ее в процессе обучения.

Количество узлов слоя 2 (L2) равно m . Каждый узел этого слоя соответствует одному нечеткому правилу. Узел второго слоя соединен с теми узлами первого слоя, которые формируют antecedentes соответствующего правила. Следовательно, каждый узел второго слоя может принимать от 1 до n входных сигналов. Выходом узла является степень выполнения правила (вес некоторого правила).

$$\omega_i = \mu_{A_i}(x) \times \mu_{B_i}(x), \quad i = 1, 2, 3, 4, 5.$$

Количество элементов этого слоя равно количеству правил N . Каждый узел связан с предыдущим слоем таким образом, что узел слоя L2, соответствующий k -му правилу, соединен со всеми узлами слоя L1, соответствующими нечетким множествам суждений этого правила.

Каждый i -й узел слоя 3 (L3) определяет отношение веса i -го правила к сумме весов всех правил:

$$\bar{\omega} = \frac{\omega_i}{\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5}, \quad i = 1, 2, 3, 4, 5$$

Выходные сигналы 3-го слоя называются нормализованными весами.

Каждый узел слоя 4 (L4) соединен с одним узлом третьего слоя, а также со всеми входами сети. Узел четвертого слоя рассчитывает вклад одного нечеткого правила в выход сети:

$$O_i^4 = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x + q_i y + r_i)$$

где $\bar{\omega}_i$ – выходной сигнал 3-го слоя, $\{p_i, q_i, r_i\}$ – набор параметров данного слоя (так называемые параметры вывода).

Слой 5 (L5) представляет собой реализацию блока дефазификации, реализующего зависимость.

$$O_i^5 = \sum_i \bar{\omega}_i f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i}$$

Веса связей, доходящих до верхнего узла слоя L4 и обозначенные u , интерпретируются как центры функций принадлежности нечетких множеств A_k . Эти веса, также как и значения параметров x_i , в слое L1, будут модифицироваться в процессе обучения. На выходе слоя L5 формируется «четкое» (дефазифицированное) выходное значение модуля управления u . Представленная на рисунке 24 структура имеет много общего с нейронными сетями –

она представляет собой многослойную сеть, основанную на идее нечеткого вывода. В отличие от «чистых» нейронных сетей, каждый слой в целом и отдельные составляющие его элементы, также, как и конфигурация связей, все параметры и веса имеют физическую интерпретацию. Это свойство оказывается необычайно важным, поскольку знания не распределяются по сети и могут быть легко локализованы и при необходимости откорректированы экспертом-наблюдателем.

Рассмотренные выше этапы нечеткого вывода могут быть реализованы неоднозначным образом, поскольку включают в себя отдельные параметры, которые должны быть фиксированы или специфицированы. Тем самым выбор конкретных вариантов параметров каждого из этапов определяет некоторый алгоритм, который реализует нечеткий вывод в системах правил нечетких продукций. Один из таких алгоритмов – алгоритм Сугено. Его основная особенность состоит в том, что заключения правил задаются линейной функцией от входов. Быстрые алгоритмы обучения и интерпретируемость накопленных знаний – эти факторы сделали сегодня нечеткие нейронные сети одним из самых перспективных и эффективных инструментов мягких вычислений. Наибольшее распространение в настоящее время получили архитектуры нечеткой нейронной сети вида ANFIS и TSK. Доказано, что такие сети являются универсальными аппроксиматорами.

3.1.1 Архитектура нечеткой нейронной сети ANFIS (Adaptive Network based Fuzzy Inference System)

ANFIS представляет адаптивную систему нейро-нечеткого вывода. С одной стороны, сеть ANFIS представляет собой нейронную сеть с единственным выходом и несколькими входами, которые представляют собой нечеткие лингвистические переменные. При этом термы входных лингвистических переменных описываются стандартными функциями принадлежности, которые приведены в главе 1.2, а термы выходной переменной представляются линейной или постоянной функцией принадлежности. С другой стороны, сеть ANFIS представляет собой систему нечеткого вы-

вода, в которой каждое из правил нечетких продукций имеет постоянный вес, равный 1. Данная сеть ANFIS может быть успешно использована для настройки функций принадлежности и настройки базы правил в нечеткой экспертной системе. Алгоритмы обучения адаптивных нечетких систем относительно трудоемки и сложны по сравнению с алгоритмами обучения нейронных сетей, и, как правило, состоят из двух стадий: 1. Генерация лингвистических правил; 2. Корректировка функций принадлежности. Первая задача относится к задаче переборного типа, вторая – к оптимизации в непрерывных пространствах.

Ниже представлена модель адаптивной сети ANFIS состоящая из двух правил:

ЕСЛИ $x_1 = A_1$ И $x_2 = B_1$, ТО $y_1 = c_{11}x_1 + c_{12}x_2$,

ЕСЛИ $x_1 = A_2$ И $x_2 = B_2$, ТО $y_2 = c_{21}x_1 + c_{22}x_2$

Выход сети формируется по формуле:

$$y = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2}$$

Слой сети ANFIS выполняют следующие функции.

Слой 1. представлен радиально базисными (РБФ)-нейронами и моделирует функции принадлежности.

Слой 2. это слой «И» нейронов, которые моделируют логическую связку И произведением $w_i = \mu_{A_i}(x_1) \cdot \mu_{B_i}(x_2) \cdot \dots$

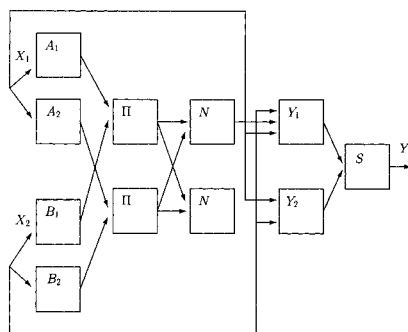


Рис. 25 – Структура сети ANFIS

Слой 3. вычисляет нормированную силу правила:

$$\langle w_i \rangle = \frac{w_i}{w_1 + w_2}$$

Слой 4. формирует значение выходной переменной:

$$y(x_1, x_2) = \langle w_i \rangle \cdot y_i = \langle w_i \rangle \cdot (c_{i1}x_1 + c_{i2}x_2)$$

Слой 5. выполняет дефаззификацию:

$$y = \langle w_i \rangle \cdot y_1 + \langle w_i \rangle \cdot y_2$$

гибридная сеть архитектуры ANFIS может быть обучена с помощью алгоритма обратного распространения ошибки.

3.1.2 Архитектура NNFLC (Neurons network fuzzy logic controller)

NNFLC – нечеткий контроллер на основе нейронной сети. Структура NNFLC приведена на рис. 26. Особенность структуры NNFLC – это различная функциональность слоев. Опишем кратко функции слоев.

Слой 1. $y_i^{(1)}(x) = \exp \left[\frac{-(x_i - c_i)^2}{2\sigma_i^2} \right]$.

Слой 2. $y_i^{(2)} = \min[y_1^{(1)}, \dots, y_n^{(1)}]$. Слой 2 моделирует И-условия правил.

Слой 3. $y_i^{(3)} = \max[y_1^{(2)}, \dots, y_n^{(2)}]$ – функция рабочего режима, которая представляет собой ИЛИ-комбинацию правил с одинаковыми термами в консеквентах. В режиме обучения слой настраивает параметры функций принадлежности выходных переменных.

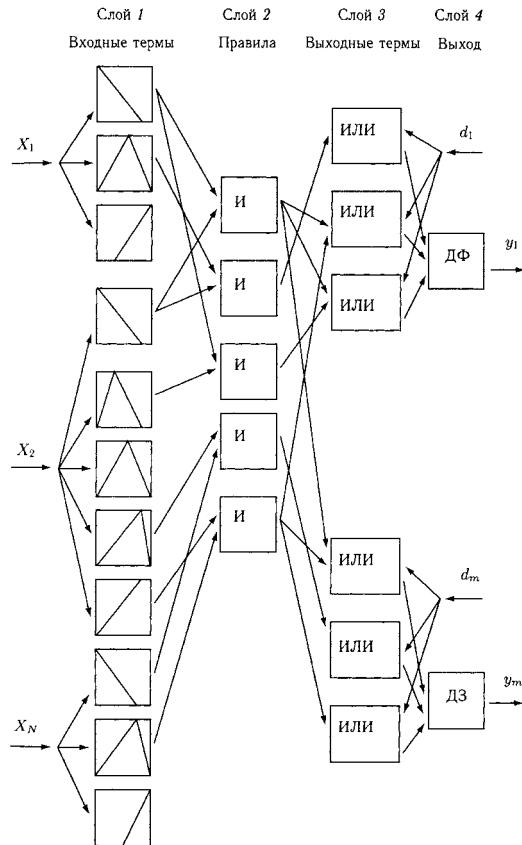


Рис. 26 – Структура сети NNFLC

Слой 4. В рабочем режиме нейроны выполняют дефазификацию, а в режиме обучения это дополнительный вход, выполняющий нормализацию:

$$z_i^{(4)} = \sum_j w_{ji} y_i^{(3)}, \quad y_i^{(4)}(z_i^{(4)}) = \sum_j y_i^{(3)} \left[\frac{w_{ji}}{\sum_j w_{ji}} \right]$$

Обучение нейронной сети сложной архитектуры (с различными функциональными слоями) обычно происходит многоэтапно,

причем на каждом этапе используются различные алгоритмы обучения: предобучение (offline), оперативное (online), без учителя, с учителем. Структура нечеткой нейронной сети NNFLC инициализируется по принципу формирования полной матрицы правил. Если x_i – входные переменные, $\tau(x_i)$ – количество нечетких меток (разбиений) x_i , то исходное количество правил: $T = \prod_{i=1}^n \tau(x_i)$

Общая схема обучения нечеткой не NNFLC содержит следующие этапы:

- формирование обучающих данных;
- самоорганизующаяся кластеризация (настройка функций принадлежности);
- соревновательное обучение (алгоритм победителя);
- удаление правил;
- комбинирование правил;
- окончательная настройка пара метров (тюнинг) функций принадлежности с помощью алгоритма обратного распространения ошибки.

Далее приведем содержательные характеристики этапов обучения. Настройка параметров функций принадлежности включает в себя обучение центров c_i и ширины σ_i для функции принадлежности, представленной функцией формы:

$$y_i^{(1)}(x) = \exp \left[-\frac{(x_i - c_i)^2}{2\sigma_i^2} \right]$$

Алгоритм победителя выявляет c_i : $\Delta c_{\omega}(t) = \eta(t)(x - c_{\omega}(t))$, где $\eta(t)$ – монотонно убывающий уровень обучения. Настройка ширины σ_i осуществляется эвристически, например по принципу «первого ближайшего соседа»:

$\sigma_i = \frac{(\sigma_i - \sigma_{\omega})}{\lambda}$, λ – параметр перекрытия.

Найти матрицу весов w_{ji} , которая оценивает качество связей левой и правой частей правил можно на основе алгоритма победителя

$$\square w = y_j^{(3)}(y_i^{(2)} - w_{ji}).$$

Комбинирование правил часто целесообразно выполнить с участием эксперта. Окончательная настройка функций принадлежности выполняется с помощью алгоритма обратного распространения ошибки для функции ошибки $e_i = (y_i^{(4)} - d_k)^2$. Цепочка правил распространяет ошибку до слоя 1 с обратным роутином. Таким образом, можно сделать вывод о том, что архитектура NNFLC может быть проинтерпретирована как система нечеткого вывода Такаги-Сугено.

3.1.3 Архитектура TSK (Takagi, Sugeno, Kang'a)

Структура нечёткой сети TSK основана на системе нечёткого вывода Такаги- Сугэно-Канга. При этом в качестве функции фаззификации для каждой переменной x_j используется обобщённая функция Гаусса:

$$\mu_A(x_j) = \frac{1}{1 + \left(\frac{x_j - c_j}{\sigma_j} \right)^{2b_j}} \quad (1.25)$$

Для агрегации условия i -го правила в системе вывода TSK используется операция алгебраического произведения:

$$\mu_A^{(i)}(x) = \prod_{j=1}^N \left[\frac{1}{1 + \left(\frac{x_j - c_j^{(i)}}{\sigma_j^{(i)}} \right)^{2b_j^{(i)}}} \right] \quad (1.26)$$

При M правилах вывода агрегирование выходного результата сети производится по формуле , которую можно представить в виде:

$$y(x) = \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i y_i(x), \quad (1.27)$$

где $y_i(x) = p_{i0} + \sum_{j=1}^N p_{ij}x_j$, агрегация импликации. Присутствующие

в этом выражении веса w_i интерпретируются как компоненты $\mu_A^{(i)}(x)$, определённые формулой. При этом формуле можно сопоставить многослойную структуру сети, изображённую на рисунке 27. В такой сети выделяется пять слоёв:

Слой 1. выполняет раздельную фаззификацию каждой переменной $x_j (j = 1, 2, \dots, N)$ определяя для каждого i -го правила вывода значение коэффициента принадлежности $\mu_A^{(i)}(x_j)$ в соответствии с применяемой функцией фаззификации. Это параметрический слой с параметрами $(c_j^{(i)}, \sigma_j^{(i)}, b_j^{(i)})$, подлежащими адаптации в процессе обучения.

Слой 2. выполняет агрегирование отдельных переменных x_j , определяя результирующее значение коэффициента принадлежности $w_i = \mu_A^{(i)}(x)$ для вектора x в соответствии с формулой. Этот слой непараметрический.

Слой 3. представляет собой генератор функции TSK, рассчитывающий значения $y_i(x) = p_{i0} + \sum_{j=1}^N p_{ij}x_j$. В этом слое также про-

исходит умножение сигналов $y_i(x)$ на значения w_i , сформированные в предыдущем слое. Это параметрический слой, в котором адаптации подлежат линейные веса p_{ij} для $i = 1, 2, \dots, M$ и $j = 1, 2, \dots, N$, определяющие функцию следствия модели TSK.

Слой 4. составляют два нейрона- сумматора, один из которых рассчитывает взвешенную сумму сигналов $y_i(x)$, а второй опреде-

ляет сумму весов $\sum_{i=1}^M w_i$. Это непараметрический слой.

Слой 5. состоит из одного выходного нейрона-это нормализующий слой, в котором веса подвергаются нормализации в соответствии с формулой. Выходной сигнал $y(x)$ определяется выражением, соответствующим зависимости:

$$y = \frac{\sum_{i=1}^M w_i}{\sum_{i=1}^M w_i} \left(p_{i0} + \sum_{j=1}^N p_{ij}x_j \right),$$

которая линейна относительно всех входных переменных системы x_j для $j = 1, 2, \dots, N$. Веса w_l являются нелинейными параметрами функции y , которые уточняются в процессе обучения. Коэффициенты $p_0, p_1, \dots, p_N$ – это веса, подбираемые в процессе обучения.

$$y(x) = f(x) = \frac{f_1}{f_2}. \quad (1.28)$$

Это также непараметрический слой.

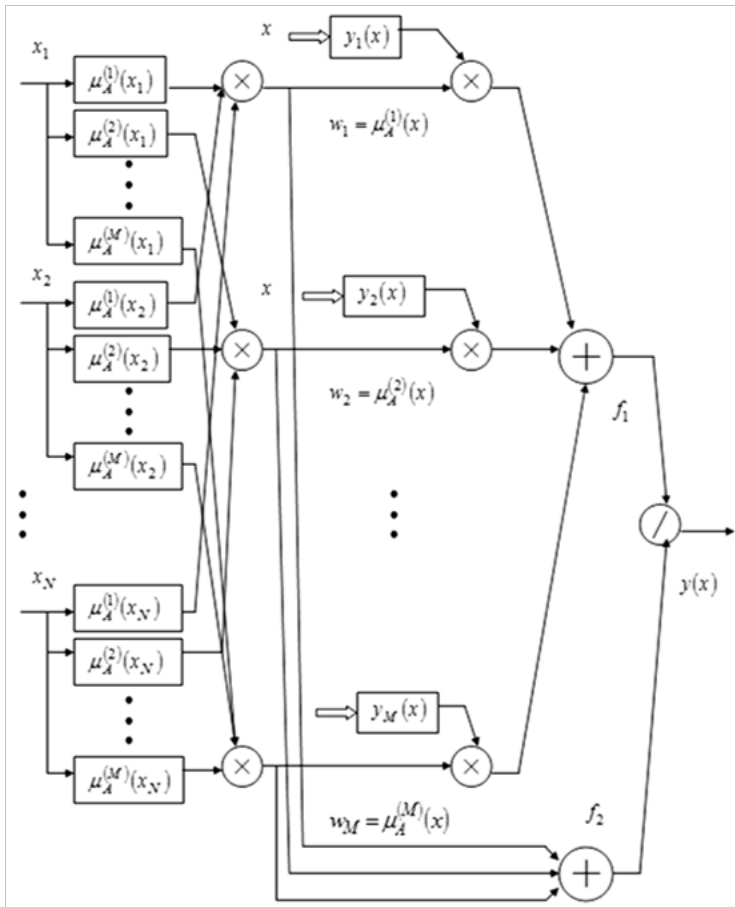


Рис. 27 – Структура нечёткой нейронной сети TSK

При уточнении функциональной зависимости для сети TSK получаем:

$$y(x) = \frac{1}{\sum_{i=1}^M \left[\prod_{j=1}^N \mu_A^{(i)}(x_j) \right]} \sum_{i=1}^M \left[\prod_{j=1}^N \mu_A^{(i)}(x_j) \right] \left[p_{i0} + \sum_{j=1}^N p_{ij} x_j \right]. \quad (1.29)$$

Если принять, что в конкретный момент времени параметры условия зафиксированы, то функция $y(x)$ является линейной относительно переменных $x_j (j = 1, 2, \dots, N)$.

При наличии N входных переменных каждое правило формирует $N + 1$ переменных p_{ij} линейной зависимости TSK. При M правилах вывода это даёт $M(N + 1)$ линейных параметров сети. В свою очередь каждая функция принадлежности использует три параметра, подлежащих адаптации. Так как каждая переменная x_j характеризуется собственной функцией принадлежности, то мы получим $3MN$ нелинейных параметров. В сумме это даёт $M(4N + 1)$ линейных и нелинейных параметров, значения которых должны подбираться в процессе обучения сети.

3.2 Обучение нечеткой нейронной сети

В существующих системах с нечеткими нейронными сетями одним из важнейших вопросов является разработка оптимального метода настройки нечеткой базы правил, исходя из обучающей выборки, для получения конструктивной и оптимальной модели нечеткой нейронной сети. В основном нечеткие правила описываются экспертами или операторами согласно их знаниям и опыту о соответствующих процессах. Но в случае разработки нечеткой нейронной сети довольно тяжело или почти невозможно сразу получить нечеткие правила или функции принадлежности вследствие неясности, неполноты или сложности описываемой системы. В таких случаях целесообразным считается уточнение нечетких правил и функций принадлежности, используя специальные алгоритмы обучения.

3.2.1 Обучение нечеткой нейронной сети ANFIS

Так как нечеткая сеть ANFIS, описанная ранее, представляется в виде многослойной структуры с прямым распространением сигнала, а значение выходной переменной можно изменять, корректируя параметры элементов слоев, то для обучения этой сети можно использовать градиентный алгоритм. Основной характеристикой данного подхода является то, что настройка параметров функций принадлежности осуществляется без модификации базы правил [7]. Пусть задана обучающая выборка, состоящая из множества примеров, $x_1^{(k)}, x_2^{(k)}, \dots, x_m^{(k)}$ – значения входных переменных $x_1, x_2, \dots, x_m$, $y^{(k)}$ – эталонное значение выходной переменной y в k -м примере, K – общее число примеров в обучающей выборке. Отметим, что для нечеткой нейронной сети ANFIS с использованием алгоритма Сугено настраиваемыми параметрами являются параметры функций принадлежности формула, где a_{ij} – центр функции принадлежности, b_{ij} – ширина данной функции.

На первом этапе обучения для каждого примера из обучающей выборки по значениям входных переменных $x_1^{(k)}, x_2^{(k)}, \dots, x_m^{(k)}$ нечеткая сеть рассчитывает значения выходной переменной $y^{(k)}$.

На втором этапе вычисляется функция ошибки для всех примеров обучающей выборки:

$$E^{(k)} = \sum \frac{(y^{(k)} - y^{(k)})^2}{2}, \quad k = 1, \dots, K \quad (1.30)$$

В данном случае функция ошибки может быть записана как функция, зависящая от следующих аргументов:

$$E^{(k)} = E^{(k)}(a_{ij}, b_{ij}) = \frac{1}{2} (y^{(k)}(a_{ij}, b_{ij}) - y^{(k)})^2, \quad k = 1, \dots, K \quad (1.31)$$

На третьем этапе корректируются значения (a_{ij}, b_{ij}) по каждому примеру обучающей выборки, исходя из соотношений

$$\begin{aligned} a_{ij}(t+1) &= a_{ij}(t) - \eta \frac{\partial E^{(k)}(t)}{\partial a_{ij}(t)}, \\ b_{ij}(t+1) &= b_{ij}(t) - \eta \frac{\partial E^{(k)}(t)}{\partial b_{ij}(t)}, \end{aligned} \quad (1.32)$$

где t – номер итерации обучения, $\eta \in [0,1]$ – коэффициент, характеризующий скорость обучения.

Этапы 1 – 3 итерационно повторяются, и процедура корректировки значений всех параметров считается завершенной в случае, если значение функции ошибки по каждому примеру обучающей выборки не превышает некоторого установленного порога:

$$E^{(k)} < \varepsilon, \quad k = 1, \dots, K,$$

либо оценка средней суммарной погрешности нечеткой сети с учетом всех примеров обучающей выборки не превышает некоторого установленного порога:

$$E = \frac{1}{K} \sum_{k=1}^K \left(y^{*(k)} - y^{(k)} \right)^2 < \varepsilon.$$

В этом случае считается, что нечеткая сеть успешно обучилась.

При обучении с помощью алгоритма обратного распространения ошибки обеспечивается свойство единственности представления лингвистических термов за счет связанной адаптации параметров функций принадлежности.

3.2.2 Обучение нечеткой нейронной сети TSK

Рассмотрим гибридный алгоритм обучения. Данный алгоритм служит для обучения нечеткой сети TSK. Нечеткие правила данной модели имеют вид, представленный формулой:

$$R_k : \text{Если } x^1 \in A_k^1 \text{ и } \dots \text{ и } x^n \in A_k^n, \quad \text{то класс } y = p_{k0} + \sum_{j=1}^N p_{kj} x^j$$

где R_k – метка k -го правила, $x = (x^1, \dots, x^n)$ – n -мерный вектор входных признаков, y – выходной признак, A_k^j – нечеткое множество предпосылки правила R_k , $p_0, \dots, P_N$ – веса.

В гибридном алгоритме параметры, подлежащие адаптации, делятся на 2 группы. Первая из них состоит из линейных параметров p_{kj} третьего слоя (весов), а вторая группа – из параметров нелинейной функции принадлежности первого слоя – это параметры p_{k0} . Уточнение параметров происходит в 2 этапа.

На первом этапе при фиксации отдельных значений параметров функции принадлежности (в первом цикле – это значения, которые получены путем инициализации), решая систему линейных уравнений, рассчитываются линейные параметры p_{kj} . При известных значениях функции принадлежности зависимость для выхода можно представить в виде линейной формы относительно параметра p_{kj} [7]:

$$y_k(x) = \sum_{k=1}^M \omega_k' \left(p_{k0} + \sum_{j=1}^N p_{kj} x_j \right),$$

$$\text{где, } \omega_k' = \frac{\prod_{j=1}^N \mu_A^{(k)}(x_j)}{\sum_{r=1}^M \prod_{j=1}^N \mu_A^{(r)}(x_j)}, \quad k = \overline{1, M}$$

При размерности обучающей выборки $L(x^{(k)}, y^{(k)})$, $k = 1, \dots, K$ и замене выходного сигнала сети ожидаемым значением $y^{(k)}$ получим систему из K линейных уравнений вида

$$\begin{bmatrix} \omega_{11}' & \omega_{11}' x_1^{(1)} & \dots & \omega_{11}' x_N^{(1)} & \dots & \omega_{1M}' & \omega_{1M}' x_1^{(1)} & \dots & \omega_{1M}' x_N^{(1)} \\ \omega_{21}' & \omega_{21}' x_1^{(2)} & \dots & \omega_{21}' x_N^{(2)} & \dots & \omega_{2M}' & \omega_{2M}' x_1^{(2)} & \dots & \omega_{2M}' x_N^{(2)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \omega_{K1}' & \omega_{K1}' x_1^{(K)} & \dots & \omega_{K1}' x_N^{(K)} & \dots & \omega_{KM}' & \omega_{KM}' x_1^{(K)} & \dots & \omega_{KM}' x_N^{(K)} \end{bmatrix} \times \begin{bmatrix} p_{10} \\ p_{11} \\ \dots \\ p_{1N} \\ \dots \\ p_{M0} \\ p_{M1} \\ \dots \\ p_{MN} \end{bmatrix} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(K)} \end{bmatrix}$$

где ω'_{ik} означает уровень активации (вес) условия k -го правила при предъявлении i -го входного вектора $x^{(i)}$. Это выражение можно записать в матричном виде:

$$Ap = y'$$

Размерность матрицы A равняется $K \times (N + 1)M$. При этом количество строк K обычно бывает значительно больше количества столбцов $(N + 1)M$.

При наличии N входных переменных каждое правило R_k формирует $(N + 1)$ переменную p_{kj} линейной зависимости $y^{(k)}$. При M правилах вывода имеем $(N + 1)M$ линейных параметров сети.

Решение этой системы уравнений можно получить используя псевдоинверсию матрицы A :

$$p = A^+ y',$$

где A^+ – псевдоинверсная матрица. Псевдоинверсия матрицы заключается в решении задачи минимизации:

$$\min \|A^+ A - E\|,$$

где E – единичная матрица.

На втором этапе после фиксации значения линейных параметров p_{kj} рассчитываются фактические выходные сигналы $y^{(k)}$, $k = 1, \dots, K$, для этого используется линейная зависимость:

$$y^{(k)} = Ap$$

После этого рассчитывается вектор ошибки $\varepsilon = y - y'$ и критерий

$$E = \frac{1}{2} \sum_{k=1}^K (y^{(k)} - y'^{(k)})^2$$

Сигналы ошибок направляются через сеть в обратном порядке согласно методу Back Propagation вплоть до первого слоя, где мо-

гут быть рассчитаны компоненты a , b . После вычисления вектора градиента делается шаг спуска градиентным методом.

После уточнения нелинейных параметров снова запускается процесс адаптации линейных параметров (первый этап) и нелинейных параметров (второй этап). Этот цикл продолжается до тех пор, пока не стабилизируются все параметры процесса.

При практической реализации гибридного метода обучения нечетких сетей доминирующим фактором их адаптации считается первый этап, на котором веса p_{kj} подбираются с использованием псевдоинверсии за один шаг. Для уравнивания его влияния второй этап много раз повторяется в каждом цикле.

Таким образом, из проведенного анализа видно, что с использованием алгоритма обратного распространения ошибки происходит оптимальное уточнение функций принадлежности, что дает возможность лучше подготовить нечеткую нейронную сеть для решения конкретных задач с меньшими затратами времени. С использованием метода оптимизации (метод наименьших квадратов) оцениваются коэффициенты заключений правил, так как они линейно связаны с выходом сети. Каждая итерация процедуры настройки выполняется в два этапа. На первом этапе на входы подается обучающая выборка, и по невязке между желаемым и действительным поведением сети итерационным методом наименьших квадратов находятся оптимальные параметры узлов четвертого слоя. На втором этапе остаточная невязка передается с выхода сети на входы, и методом обратного распространения ошибки модифицируются параметры узлов первого слоя. При этом найденные на первом этапе коэффициенты заключений правил не изменяются. Итерационная процедура настройки продолжается пока невязка превышает заранее установленное значение. Для настройки функций принадлежности кроме метода обратного распространения ошибки могут использоваться и другие алгоритмы оптимизации

3.2.2 Алгоритм нечёткой самоорганизации C-means

Алгоритм самоорганизации приписывает вектор x к соответствующей группе данных, представляемых центром c_i , с использованием обучения конкурентного типа подобно сетям Кохонена. При обучении этого типа процесс самоорганизации становится возможным при предъявлении вектора x .

Допустим, что в сети существует M нечётких нейронов с центрами в точках c ($i = 1, 2, \dots, M$). Начальные значения этих центров могут быть выбраны случайным образом из областей допустимых значений соответствующих компонентов векторов x ($t = 1, 2, \dots, p$), использованных для обучения. Пусть функция фузификации задана в форме обобщенной функции Гаусса, выраженной формулой .

Подаваемый на вход сети вектор x будет принадлежать к различным группам, представляемым центрами c_i , в степени u_{it} , причем $0 \leq u_{it} \leq 1$, а суммарная степень принадлежности ко всем группам, очевидно, равна единице. Поэтому

$$\sum_{i=1}^M u_{it} = 1 \quad (1.33)$$

для $t = 1, 2, \dots, p$. Функцию погрешности, соответствующую такому представлению, можно определить как сумму частных погрешностей принадлежности к центрам c_i с учетом степени принадлежности u_{it} . Следовательно

$$E = \sum_{i=1}^M \sum_{t=1}^P (u_{it})^m \|c_i - x_t\|^2, \quad (1.34)$$

где m – это весовой коэффициент, который принимает значения из интервала $[1, \infty)$, на практике часто принимают $m=2$. Цель обучения с самоорганизацией состоит в таком подборе центров c_i , чтобы для заданного множества обучающих векторов $x^{(i)}$ обеспечить достижение минимума функции при одновременном соблюдении условий ограничения . Таким образом возникает задача минимизации нелинейной функции с p ограничениями типа . Решение этой задачи можно свести к минимизации функции Лагранжа, определенной в виде [13].

$$LE = \sum_{i=1}^M \sum_{t=1}^P (u_{it})^m \|c_i - x_t\|^2 + \sum_{t=1}^P \lambda_t \left(\sum_{i=1}^M u_{it} - 1 \right), \quad (1.35)$$

где $\lambda_t (t = 1, 2, \dots, p)$ – это множители Лагранжа. Доказано, что решение задачи можно представить в виде

$$c_i = \frac{\sum_{t=1}^P (u_{it})^m x_t}{\sum_{t=1}^P (u_{it})^m}, \quad (1.36)$$

$$u_{it} = \frac{1}{\sum_{k=1}^M \left(\frac{d_{it}^2}{d_{kt}^2} \right)^{\frac{1}{m-1}}}, \quad (1.37)$$

где d_{it} – это евклидово расстояние между центром c_i и вектором x_t , $d_{it} = \|c_i - x_t\|$. Поскольку точные значения центров c_i в начале процесса не известны, алгоритм обучения должен быть итерационным:

выполнить случайную инициализацию коэффициентов u_{it} , выбирая их значения из интервала $[0, 1]$ таким образом, чтобы соблюдать условие ;

определить M центров c_i в соответствии с ;

рассчитать значение функции погрешности в соответствии с . Если её значение ниже установленного порога, либо если уменьшение этой погрешности относительно предыдущей итерации пренебрежимо мало, то закончить вычисления. Иначе, перейти к п.4;

рассчитать новые значения u_{it} по формуле и перейти к п.2.

Многократное повторение итерационной процедуры ведёт к достижению минимума функции E , который необязательно будет глобальным.

Значение функции E существенно зависит от начальной инициализации значений u_{it} и центров c_i . Поэтому, предварительно следует использовать алгоритмы инициализации, наиболее известными среди которых являются алгоритмы пикового и разностного группирования.

3.2.2 Алгоритм обратного распространения ошибки для сети TSK

Нечеткая нейронная сеть TSK имеет многослойную структуру с прямым распространением сигнала, значение выхода которой можно изменять, корректируя параметры элементов слоев, что позволяет для обучения этой сети использовать алгоритм обратного распространения ошибки [11]. Для этого потребуется обучающая выборка в виде пар (x, d) , где $x = [x_1, \dots, x_N]^T$ – это входной вектор, а d – эталонный сигнал. Задача заключается в такой коррекции параметров сети, описанной выражением чтобы мера погрешности, задаваемая выражением:

$$E = \frac{1}{2} (y(x) - d)^2, \quad (1.38)$$

была минимальной.

Если применяется простейший метод наискорейшего спуска, то соответствующие формулы адаптации принимают форму:

$$p_j^{(k)}(n+1) = p_j^{(k)}(n) - \eta_c \frac{\partial E(n)}{\partial p_j^{(k)}}, \quad (1.39)$$

$$c_j^{(k)}(n+1) = c_j^{(k)}(n) - \eta_c \frac{\partial E(n)}{\partial c_j^{(k)}}, \quad (1.40)$$

$$\sigma_j^{(k)}(n+1) = \sigma_j^{(k)}(n) - \eta_\sigma \frac{\partial E(n)}{\partial \sigma_j^{(k)}}, \quad (1.41)$$

где n обозначает номер очередной итерации.

Формулы (–) требуют расчета градиента целевой функции относительно параметров функции принадлежности. Окончательный вид этих формул зависит от используемого определения функции погрешности на выходе сети, так и от формы функции принадлежности. Например, при использовании функции Гаусса

$$\mu_A(x_i) = \exp \left[- \left(\frac{x_i - c_i}{\sigma_i} \right)^2 \right], \quad (1.42)$$

Соответствующие формулы градиента целевой функции для одной пары обучающих данных (x, d) принимают вид [21]:

$$\begin{aligned}\frac{\partial E}{\partial p_j^{(k)}} &= (y(k) - d)w_r x_j, \\ \frac{\partial E}{\partial c_j} &= (y - d)(y_k(x) - y)w_k \frac{2(x_j - c_{kj})}{(\sigma_{kj})^2}, \\ \frac{\partial E}{\partial \sigma_j^{(k)}} &= (y - d)(y_k(x) - y)w_k \frac{2(x_j - c_{kj})^2}{(\sigma_{kj})^3},\end{aligned}$$

где

$$w_k = \frac{\prod_{j=1}^N \mu_A^{(k)}(x_j)}{\sum_{r=1}^N \left[\prod_{j=1}^N \mu_A^{(r)}(x_j) \right]}, \quad y_k(x) = p_{k0} + \sum_{j=1}^N p_{kj} x_j \quad (1.43)$$

Несмотря на сложную структуру приведенных формул, выражающих компоненты вектора градиента, они позволяют аналитически определить величины, необходимые для уточнения параметров нечеткой сети.

Метод наискорейшего спуска имеет линейную сходимость, поскольку в нем используются только слагаемые первого порядка при разложении целевой функции в ряд Тейлора. Указанный недостаток, а также резкое замедление минимизации в ближайшей окрестности точки оптимального решения, когда градиент принимает очень малые значения, делают алгоритм наискорейшего спуска низкоэффективным. Повысить эффективность удастся путем эвристической модификации выражения, определяющего направление градиента.

Одна из модификаций получила название алгоритма обучения с моментом. При этом подходе уточнение весов сети производится по формуле:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \cdot \frac{\partial E(t)}{\partial w_{ij}(t)} + \alpha(w_{ij}(t) - w_{ij}(t-1)) \quad (1.44)$$

где α – это коэффициент момента, принимающий значения в интервале $[0, 1]$.

Первое слагаемое в формуле соответствует алгоритму наискорейшего спуска, а второе слагаемое учитывает последнее изменение весов и не зависит от фактического значения градиента. Чем больше значение коэффициента α , тем большее значение оказывает показатель момента на подбор весов. При постоянном значении коэффициента обучения $\eta(t) = \eta$ приращение весов остается примерно одинаковым, то есть $\Delta w_{ij}(t) = \eta p(t) + \alpha \Delta w_{ij}(t)$, поэтому эффективное приращение весов можно писать формулой:

$$\Delta w_{ij}(t) = \frac{\eta}{1 - \alpha} p(t)$$

При значении $\alpha = 0,9$ это соответствует десятикратному увеличению значения коэффициента обучения и, следовательно, десятикратному ускорению процесса обучения. При малых значениях градиента показатель момента начинает доминировать, что приводит к такому приращению весов, которое соответствует увеличению значения целевой функции, позволяющему выйти из зоны локального минимума. Однако показатель момента, не должен доминировать на протяжении всего процесса обучения, поскольку это приводит к нестабильности алгоритма. На практике, увеличение целевой функции не допускается больше, чем на 4%. В противном случае, $\Delta w_{ij}(t) = 0$. При этом показатель градиента начинает доминировать над показателем момента и процесс развивается в направлении минимизации, заданном вектором градиента [17].

3.3 Алгоритмы оптимизации нечеткой нейронной сети

Адекватность модели, которая описывается нечеткими базами знаний, данным эксперимента определяется типом и качеством настройки функций принадлежности и весов правил, с помощью которых лингвистические оценки параметров модели превращаются в количественную форму.

Но поскольку функции принадлежности в свою очередь определяются экспертными методами [15], то адекватность нечеткой модели в целом также зависит от квалификации экспертов. Кроме того, далеко не всегда существует возможность привлечения эксперта, который бы промоделировал тот или иной объект (процесс), что может быть связано с его сложностью или определенной новизной и как следствие – недостаточным освоением. В связи с этим возникает вопрос о возможности автоматизации процесса построения нечетких баз знаний, которые моделируют указанные объекты, на основе имеющихся экспериментальных данных, полученных в результате их исследования. Так, преобразование экспериментальной информации в нечеткие базы знаний могло бы быть полезным в медицине, и других областях, где лица, принимающие решения, вместо строгих количественных соотношений (которых они зачастую не имеют по объективным причинам) отдают предпочтение прозрачным и легким для интерпретации словесным правилам. Чтобы улучшить классификационную способность базы правил рассмотрим различные алгоритмы оптимизации.

Для исследования были выбраны среднемасштабная оптимизация с использованием нелинейного программирования с ограничениями, и оптимизация с использованием алгоритма на основе фильтра Калмана [3, 8, 31].

Так как объем выборки недостаточно большой, то используется среднемасштабная оптимизация. В общем случае, эта задача относится к нелинейной оптимизации с ограничениями или к нелинейному программированию [3, 8, 31]. В данном методе на каждой итерации решается подзадача Квадратичного Программирования. В то же время квадратичная задача достаточно проста, и для нее существуют эффективные (в том числе конечные) методы решения. Таким образом, указанный подход имеет практический смысл, в отличие, например, от использования аппроксимаций более высокого порядка. С другой стороны, методы Последовательного Квадратичного Программирования (SQP) [24] могут рассматриваться как результат распространения фундаментальной идеи ньютоновских методов на задачи оптимизации

с функциональными ограничениями. Например, в случае задачи с чистыми ограничениями-равенствами методы SQP суть не более чем специальные реализации ньютоновских методов решения системы Лагранжа. На сегодняшний день методы SQP входят в число наиболее эффективных оптимизационных методов общего назначения.

Оптимизация с использованием нелинейного программирования с ограничениями относится к этапу алгоритма обучения адаптивных нечетких нейронных сетей. Оптимизация представляет собой нахождение таких параметров весовых коэффициентов правил $w = \{w_1, w_2, \dots, w_n\}$, которые являются оптимальными в смысле некоторого критерия. В этом случае такая задача сводится к минимизации целевой функции с ограничениями, в виде неравенств $0 \leq g_i(w) \leq 1$ ($i = 1, 2, \dots, m$) или параметрических границ w_L, w_U . Общая формулировка задачи параметрической оптимизации представляется следующим образом: требуется найти вектор w , обеспечивающий:

$$f(w) \rightarrow \min_{w \in R^n} \quad (1.45)$$

при ограничениях

$$0 \leq g_i(w) \leq 1 \quad (i = 1, 2, \dots, m), \quad (1.46)$$

где w – вектор оптимизируемых параметров ($w \in R^n$), $f(w)$ – скалярная целевая функция векторного аргумента ($f(w): R^n \rightarrow R$), $g_i(w)$ – некоторые скалярные функции векторного аргумента. Данная задача может быть решена с применением метода SQP. В настоящее время, более эффективным считается применение так называемых уравнений Куна-Таккера с использованием леммы Фаркаша, которая утверждает о существовании решения одной и только одной из двух систем линейных уравнений и неравенств [24]. То, что из уравнения следует $\nabla f(w^*) \geq 0$, означает, что существует множество таких неотрицательных скалярных постоянных, что

$$\begin{aligned}\nabla f(w^*) &= -\sum_{i=1}^m \lambda_i \nabla g_i(w^*), \\ \sum_{i=1}^m \lambda_i \nabla g_i(w^*) &= 0, \quad \lambda_i \geq 0, \quad 1 \leq i \leq m.\end{aligned}\tag{1.47}$$

где ∇ – векторный дифференциальный оператор, w^* – является точкой минимума функции $f(w)$ при ограничениях, удовлетворяющих условию регулярности в виде линейной независимости, λ_i – неотрицательный множитель Лагранжа, l – множество индексов активных ограничений.

Поскольку при $j \in l$ функция $\nabla g_i(w^*) = 0$, а при $j \notin l$ $\lambda_i = 0$, поскольку $g_i(w^*) < 0$ для всех $j \notin l$. Поскольку градиенты подлежат выходу на нулевые значения, то множители Лагранжа $(\lambda_i, i = 1, \dots, m)$ будут необходимы для того, чтобы уравновесить отклонения по величине данной целевой функции и градиентов ограничений. Поскольку только активные ограничения вовлечены в данную процедуру обнуления, то ограничения не активны и не должны подвергаться данной процедуре и поэтому соответствующие множители Лагранжа будут равны нулю. Решение уравнений Куна-Таккера служит основой для последовательного квадратичного программирования. В этом алгоритме используется прямое вычисление множителей Лагранжа. Пусть $j, g_i, i \leq m$, обладают непрерывными частными производными на некотором открытом множестве пространства R^n , содержащем w^* . Если w^* является точкой минимума функции $f(w)$ при ограничениях $0 \leq g_i(w) \leq 1$ ($i = 1, 2, \dots, m$), удовлетворяющих условию регулярности в виде линейной независимости, то существуют такие неотрицательные множители Лагранжа $(\lambda_i, i = 1, \dots, m)$, что

$$L(w, \lambda) = f(w) + \lambda^T g(w)\tag{1.48}$$

Используя последовательное квадратичное программирование на каждом шаге которого функции цели и ограничений заменяются на их квадратичные приближения, и решается следующая подзадача квадратичного программирования: нахождение вектора d , с помощью которого корректируется вектор весов

$w_{i+1} = w_i + d$, такого, что d является решением задачи квадратичного программирования:

$$\min_d \left(\nabla L(w, \lambda)^T d + \frac{1}{2} d^T H(w) d, \nabla g(w) + A(w) d \right) = 0 \quad (1.49)$$

где ∇ – векторный дифференциальный оператор, H – симметричная и положительно определенная матрица вторых частных и смешанных производных (матрица Гессе), A – якобиан ограничений.

В качестве матрицы H (гессиан Лагранжа) может быть использован как полный гессиан:

$$\nabla_w^2 L(w, \lambda) = \nabla_w^2 f(w) + \nabla_w^2 (\lambda^T g(w)) = H_1(w) + H_2(w, \lambda), \quad (1.50)$$

так и неполный:

$$H^{GN}(w, \lambda) = H_1(w), \quad (1.51)$$

где

$$H_1(w) = \nabla_w^2 f(w) = \nabla_w^2 \sum_{i=1}^m (\hat{w}_i - w_i)^2 = - \begin{bmatrix} 2 & & & \\ & \ddots & & \\ & & 2 & \\ & & & 0_m \end{bmatrix}, \quad (1.52)$$

$$H_2(w, \lambda) = \sum_{i=1}^m \nabla_w^2 (\lambda_i g_i(w)) \quad (1.53)$$

0_{n_m} – нулевая квадратная матрица размерности n_m .

Положим, что якобиан ограничений $g(w)$ равен $A(w)$, т.е. что

$$A(w) = \nabla_w c(w) \quad (1.54)$$

Отметим, далее, что ограничения $g(w)$ входят только в выражения якобиана и гессиана ограничений, а, следовательно, только эти матрицы зависят от схемы численного интегрирования целевой функции.

Неполный гессиан используется для возмещения отсутствия положительной определённости $\nabla_w^2 L(w, \lambda)$ и для уменьшения

сложности вычисления. Также использование неполного гессе-ана приводит к упрощению программного кода и ускорению его разработки, при этом, как будет показано, скорость сходимости алгоритма к решению будет выше при выполнении некоторых условий. Вид матрицы Гессе для Лагранжевой функции обновляется на каждой итерации с помощью метода BFGS (Broyden, Fletcher, Goldfarb, Shanno) [28].

Метод BFGS оптимизации, использующий аппроксимацию матрицы вторых производных H_i представлен ниже.

На первом шаге рассчитывается матрица H_0 . В качестве H_0 выбирается единичная матрица. На втором шаге рассчитывается градиент в текущей точке. На третьем шаге рассчитывается направление $d_i = -\nabla^T f(w_i) H_i^{-1}$. На четвертом шаге происходит расчет шага. На пятом шаге рассчитывается новая точка $d_i = w_{i+1} - w_i$. На шестом шаге алгоритма происходит пересчет матрицы H_{i+1} согласно методу BFGS.

$$H_{i+1} = H_i + \frac{q_i q_i^T}{q_i^T d_i} - \frac{H_i^T d_i d_i^T H_i}{d_i^T H_i d_i}, \quad (1.55)$$

$$q_i = \nabla f(w_{k+1}) - \nabla f(w_k)$$

На седьмом шаге происходит проверка критерия останова $q_i^T d_i > 0$. Если не выполняется, то положить $i = i + 1$ и перейти к шагу два.

Матрица H должна обладать двумя свойствами:

- симметричность;
- положительная определенность.

Матрицу H_i можно искать из условия линейной интерполяции для квадратичной модели.

$$\nabla^T f(w_i) + H_i(w_{i+1} - w_i) = \nabla^T f(w_{i+1}) \quad (1.56)$$

Если $d_i = w_{i+1} - w_i$, $q_i = \nabla^T f(w_{i+1}) - \nabla^T f(w_i)$, тогда соотношение текущих для оптимизации будет иметь вид:

$$H_{i+1} d_i = q_i \quad (1.57)$$

Пересчет H_{i+1} происходит по формуле .

Суть метода Бройдена [27] состоит в том, чтобы вычислить матрицу Гессе аналитически или с помощью метода конечных разностей на первой итерации, а после этого на каждой итерации обновлять матрицу Гессе, не вычисляя значения функций и их производных.

Рассмотрим следующий алгоритм оптимизации применительно к базе знаний нечеткой нейронной сети ANFIS типа Сугено [24]. Предполагается, что задана нечеткая база знаний Сугено и существует обучающая выборка из M пар экспериментальных данных, связывающих входы $X_r = (x_{r1}, x_{r2}, \dots, x_{rn})$ с выходом y исследуемой зависимости (X_r, y) , $r = 1, M$. Параметрическая идентификация нечеткой базы знаний Сугено сводится к следующей задаче нахождения вектора (P, W) по формуле:

$$\sqrt{\frac{1}{M} \sum_{i=1, M} (y_i - F(P, W, X_i))^2} \rightarrow \min \quad (1.58)$$

В этой задаче оптимизации на управляемые переменные P налагаются ограничения, обеспечивающие линейную упорядоченность элементов терм-множеств. Такие ограничения не позволяют алгоритму оптимизации сделать, нечеткое множество «низкий» больше «высокого». Что касается вектора W , то его координаты должны находиться в диапазоне $[0, 1]$. Уверенность эксперта в каждом правиле ЕСЛИ-ТО, входящем в нечеткую базу знаний, может быть различной. Для отражения значимости правил введены их веса. Вес правила которое характеризует субъективную меру уверенности эксперта в этом правиле. На практике такую задачу решают алгоритмами на основе фильтра Калмана [24], которые оптимизируют только линейные параметры нечеткой базы знаний – коэффициенты в заключениях правил. Задача сводится к нахождению таких коэффициентов в заключениях нечетких правил W , которые обеспечивают минимум квадратичной невязки:

$$\sum_{r=1, M} (y - y^r)^2 \rightarrow \min, \quad (1.59)$$

где y^f – результат вывода по нечеткой базе знаний с параметром W при значении входов из строки выборки X_r . Входному вектору $X_r = (x_{r1}, x_{r2}, \dots, x_{rm})$ соответствует такой результат нечеткого вывода:

$$y^f = \frac{\sum_{j=1,m} \mu_d(X_r) d}{\sum_{j=1,m} \mu_d(X_r)}, \quad (1.60)$$

где $d = b_j + \sum_{r=1,n} b_j x_r$ – заключение j -го, $\mu_d(X_r)$ – степень выполнения j -го правила.

Относительную степень выполнения j -го правила для входного X_r обозначим через:

$$\beta_e = \frac{\mu_d(X_r)}{\sum_{k=1,m} \mu_d(X_r)}, \quad (1.61)$$

Тогда формула можно переписать в виде:

$$y^f = \sum_{j=1,m} \beta_r d = \sum_{j=1,m} \beta_r w + \beta_r b x_{r1} + \beta_r w x_{r2} + \dots + \beta_r w x_{rm}$$

Введем следующие обозначения:

$$Y^f = (y_1^f, y_2^f, \dots, y_M^f);$$

$$Y = (y_1, y_2, \dots, y_M);$$

$$A = \begin{bmatrix} \beta_{11}, \dots, \beta_{1m}, x_{11}\beta_{11}, \dots, x_{11}\beta_{1m}, \dots, x_{1n}\beta_{11}, \dots, x_{1n}\beta_{1m} \\ \beta_{M1}, \dots, \beta_{Mm}, x_{M1}\beta_{1m}, \dots, x_{M1}\beta_{1m}, \dots, x_{Mn}\beta_{M1}, \dots, x_{Mn}\beta_{Mm} \end{bmatrix}.$$

Тогда задача настройки в матричной форме ставится следующим образом нахождения вектора W :

$$E = (Y - Y^f)^T (Y - Y^f) \rightarrow \min,$$

где $Y^f = AW$.

Минимальное значение квадратичной невязки E достигается $Y^f = Y$, что соответствует решению уравнения

$$Y = AW \quad (1.62)$$

Недостатком данного алгоритма является то, что при решении реальных задач количество настраиваемых параметров меньше объема выборки данных, следовательно, уравнение не имеет точного решения. Трудности при решении также связаны с сингулярностью матрицы A . Количество итераций данного алгоритма равно объему выборки экспериментальных данных. На первом шаге алгоритма координаты вектора W настраиваются по первой строчке выборки данных. На второй итерации координаты вектора W подстраивают под вторую пару «входы–выход» и т.д. Другой недостаток данного алгоритма оптимизации при больших обучающих выборках наступает эффект насыщения – точность оптимизации практически не улучшается с увеличением числа наблюдений.

Подход нелинейной оптимизации обеспечивает существенный вклад от ограничения с небольшими параметрами в значения штрафных параметров, что особенно актуально для активных ограничений вблизи точки решения. Преимущества данного метода оптимизации заключаются в его простоте и ясной содержательной интерпретации. Процент ошибок широко используется как критерий обучения различных систем распознавания образов. Целевая функция задачи оптимизации принимает дискретные значения.

Таким образом, из проведенного анализа видно, что с использованием алгоритма нелинейной оптимизации с ограничениями происходит оптимальная настройка весов заключений правил, что дает возможность лучше подготовить нечеткую нейронную сеть для решения конкретных задач с меньшими вычислениями. Это предполагает выбор данного алгоритма для дальнейших экспериментов.

ЗАКЛЮЧЕНИЕ

В современной интеллектуальной информатике соотношение фундаментальных и технологических задач складывается в пользу технологических. Отметим, что фундаментальные задачи моделирование ощущений, восприятий, распознавание образов, обучение и запоминание паттернов являются основными задачами когнитивной науки.

Технологические задачи связаны с повышением эффективности использования нечетких нейронных сетей в решении задач кластеризации, классификации образов и сцен, аппроксимации функций для процессов технической, экономической, биологической природы. Достижения нейроинформатики в решении технологических задач привели к созданию целого спектра разнообразных нейронных сетей: сетей с прямым распространением сигнала, рекуррентных сетей, радиально-базисных сетей. Созданы развитые архитектуры гибридных систем: нечетких нейронных сетей, нечетких нейронных сетей с генетической настройкой параметров.

Нечеткие нейронные сети с генетической оптимизацией параметров обладают комплексной структурой и сложными алгоритмами обучения, интегрирующими структурные части различных нейронных сетей.

Следовательно, именно гибридные системы, как можно предположить, должны обладать сильным когнитивным потенциалом.

Исследование известных архитектур гибридных систем и алгоритмов их обучения позволяет, считать их моделями, адекватно отражающими соотношение восприятия и логического умозаключения при когнитивной деятельности человека. При реальном решении задачи человек комбинирует процессы мышления (компьютерный аналог – логический вывод), воспоминания (сознательное «внутреннее» возбуждение образа), восприятия окружающих предметов (распознавание) и, возможно, движения (управление телом). Обученная нечеткая нейронная сеть и хранит паттерны, и выполняет логические операции, так как содержит И

ИЛИ-нейроны, и переключается от распознавания на логический вывод, так как обычно включает в себя несколько отдельно обученных сетей с разными функциями. Таким образом, несмотря на то что нечеткие нейронные сети созданы как нейронечеткие контроллеры для управления процессами, они могут служить когнитивной моделью для изучения взаимодействия процессов восприятия и мышления при когнитивной деятельности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Айвазян, С.А. Прикладная статистика. Классификация и снижение размерности / В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с.
- 2 Аюев, В.В. Кластерный метод восстановления пропусков в данных для обучения ИНС / З.Е. Аунг, Чжо Мин Тейн, М.Б. Логинова // Нейрокомпьютеры: разработка, применение. – 2009. – № 7. – С. 23–34.
- 3 Банди, Б. Методы оптимизации. Вводный курс / Под ред. В.А. Волынского – Москва.: Радио и связь, 1988. – 128 с.
- 4 Батыршин, И.З. Основные операции нечеткой логики и их обобщения. – Казань: Отечество, 2001. – 100 с.
- 5 Борисов, В.В. Нечеткие модели и сети / В.В. Борисов, В.В. Круглов, А.С. Федулов. – М.: Горячая линия – Телеком, 2007. – 284 с.
- 6 Заде, Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976. – 166 с.
- 7 Зайченко, Ю.П. Нечеткие модели и методы в интеллектуальных системах. К.: Слово, 2008. – 344 с.
- 8 Измаилов, А.Ф. Численные методы оптимизации. 2-е изд., перераб. и доп. – М.: Физматлит, 2008.
- 9 Комарцова, Л.Г. Нейрокомпьютеры. – М: Изд-во МГТУ им. Н. Э. Баумана, 2004. – 400 с.
- 10 Коньшева, Л.К. Основы теории нечетких множеств. – СПб.: Питер, 2011. – 192 с.
- 11 Круглов, В.В. Нечеткая логика и искусственные нейронные сети / В.В. Круглов, М.И. Дли, Р.Ю. Голунова. – М.: Физматлит, 2001. – 224 с.
- 12 Леоненков, А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб.: БХВ–Петербург, 2005. – 736 с.
- 13 Осовский, С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002. – 344 с.
- 14 Пегат, А. Нечеткое моделирование и управление. – М.: БИНОМ. Лаборатория знаний, 2009. – 796 с.
- 15 Ротштейн, А.П. Медицинская диагностика на нечеткой логике. – Винница: Континент-ПРИМ, 1996. – 132 с.
- 16 Ротштейн, А.П. Интеллектуальные технологии идентификации // <http://matlab.exponenta.ru/fuzzylogic/book5/index.php/>
- 17 Рудковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский. – М.: Горячая линия-Телеком, 2007. – 452 с.

- 18 Рутковский, Л. Методы и технологии искусственного интеллекта. – М.: Горячая линия-Телеком, 2010. – 520 с.
- 19 Саати, Т. Принятие решений при зависимостях и обратных связях: Аналитические сети. – М.: Издательство ЛКИ, 2008. – 360 с.
- 20 Саати, Т. Принятие решений. Метод анализа иерархий. – М.: Радио и связь, 1993. – 278 с.
- 21 Хайкин, С. Нейронные сети: полный курс. – М.: Вильямс, 2006. – 1104 с.
- 22 Шамис, А.Л. Пути моделирования мышления: Активные синергические нейронные сети, мышление и творчество, формальные модели поведения и «распознавания с пониманием». – М.: КомКнига, 2006. – 336 с.
- 23 Штовба, С.Д. Введение в теорию нечетких множеств и нечеткую логику. – [Электронный ресурс]. – <http://matlab.exponenta.ru/fuzzylogic/book1/>
- 24 Штовба, С.Д. Проектирование нечетких систем средствами MATLAB. – М.: Горячая линия-Телеком, 2007. – 288 с.
- 25 Ярушкина, Н.Г. Нечеткие нейронные сети в когнитивном моделировании и традиционных задачах искусственного интеллекта // VII Всероссийская научно-техническая конференция «Нейроинформатика-2005»: Лекции по нейроинформатике. – М.: МИФИ, 2005. – С. 166-213.
- 26 Яхьяева, Г.Э. Нечеткие множества и нейронные сети. – М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2006. – 316 с.
- 27 Broyden, C.G. The Convergence of a Class of Double-rank Minimization Algorithms, Journal of the Institute of Mathematics and Its Applications, 1970. – pp. 90.
- 28 Kelly, C.T. Iterative methods for Optimization // North Caroline State University, Raleigh, North Carolina, 1999. – pp. 177.
- 29 Kosko, B. Fuzzy Systems as Universal Approximators // IEEE Transactions On Computers. – 1994, Vol. 43. – No. 11. – P. 1329–1333.
- 30 Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D: A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 130:461-470, 1999
- 31 Qi L. Superlinearly convergent approximate newton methods for LC1 optimization problems // Math. Program. – 1994. – V. 64, № 3. – P. 277-294.
- 32 Zadeh L.A. Discussion: Probability theory and fuzzy logic are complementary rather than competitive // Technometrics. – 1995. – Vol. 37, N 3. – P.271-276.

Ахметов Б.С., Горбаченко В.И., Кузнецова О.Ю.

НЕЧЕТКИЕ СИСТЕМЫ И СЕТИ

Подписано в печать 26.12.2014.

Формат 60x84^{1/16}. Бумага офсетная. Гарнитура «Times».

Объем 6,5 печ. л. Тираж 500 экз. Заказ № 2588.

Отпечатано в ТОО «Издательство ЛЕМ»

050008, г. Алматы, ул. Ауэзова, 82, уг. ул. Мынбаева, тел./факс 375-51-33.